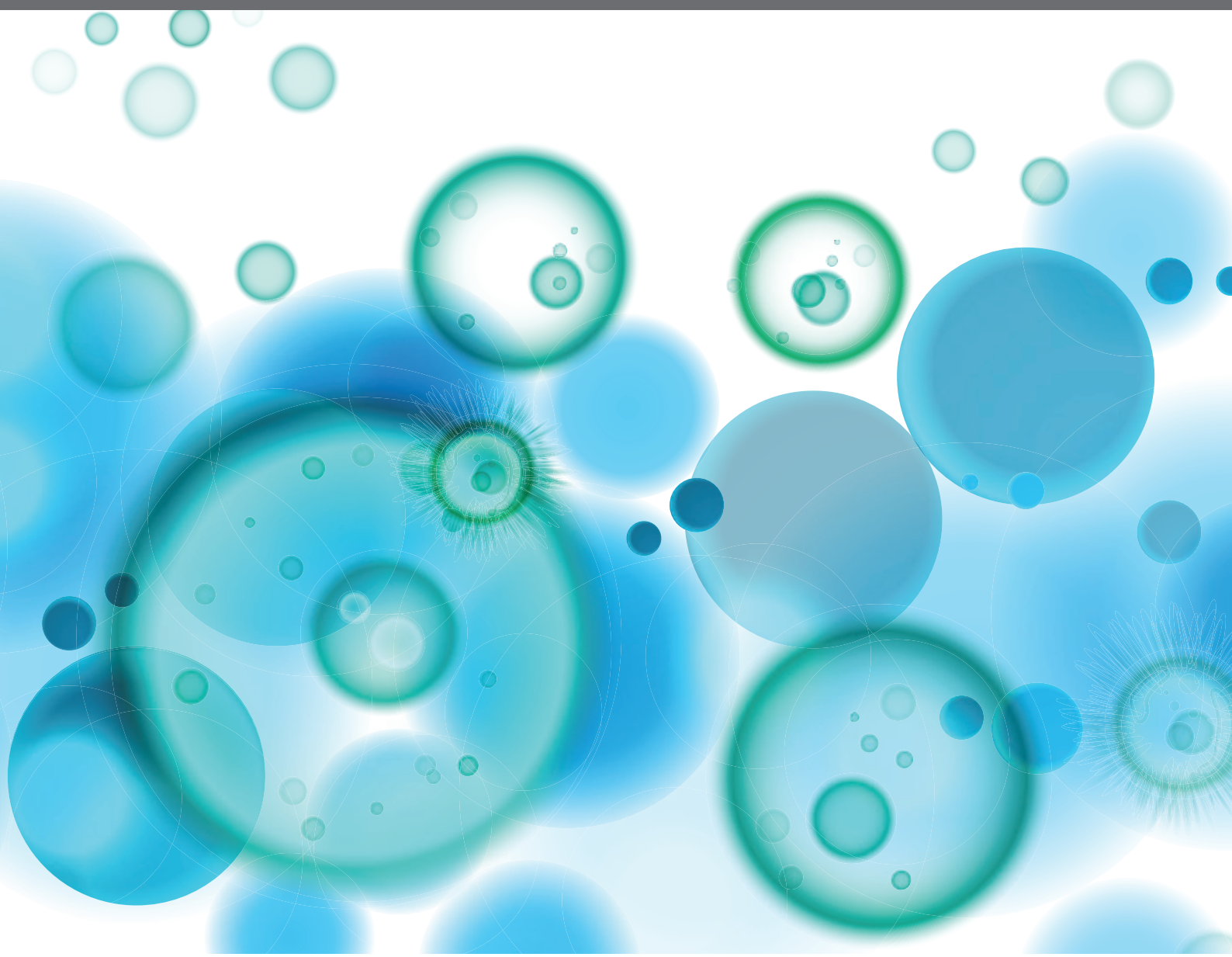
EVOLUTION OF INNATE IMMUNITY IN EUKARYA: ADVANCES AND IMPLICATIONS

EDITED BY: Jean L. Scholz, Ioannis Eleftherianos and Bostjan Kobe
PUBLISHED IN: Frontiers in Immunology and Frontiers in Plant Science



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-133-3
DOI 10.3389/978-2-88976-133-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

EVOLUTION OF INNATE IMMUNITY IN EUKARYA: ADVANCES AND IMPLICATIONS

Topic Editors:
Jean L. Scholz, University of Pennsylvania, United States
Ioannis Eleftherianos, George Washington University, United States
Bostjan Kobe, The University of Queensland, Australia

Citation: Scholz, J. L., Eleftherianos, I., Kobe, B., eds. (2022). Evolution of Innate Immunity in Eukarya: Advances and Implications. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-133-3

Table of Contents

frontiers | Frontiers in Immunology

EDITORIAL
published: 20 April 2022
doi: 10.3389/fimmu.2022.879429

Editorial: Evolution of Innate Immunity in Eukarya: Advances and Implications

Jean L. Scholz[1]*, *Ioannis Eleftherianos*[2]* and *Bostjan Kobe*[3]*

[1] Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States, [2] Infection and Innate Immunity Laboratory, Department of Biological Sciences, The George Washington University, Washington DC, United States, [3] School of Chemistry and Molecular Biosciences, Australian Infectious Diseases Research Centre and Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

Keywords: innate immunity, host-pathogen interactions, comparative immunology, immune signaling, cellular immunity

Editorial on the Research Topic

Evolution of Innate Immunity in Eukarya: Advances and Implications

OPEN ACCESS

Edited and reviewed by:
Stephanie DeWitte-Orr, Wilfrid Laurier University, Canada

***Correspondence:**
Jean L. Scholz
jeanl@pennmedicine.upenn.edu
Ioannis Eleftherianos
ioannise@gwu.edu
Bostjan Kobe
b.kobe@uq.edu.au

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *19 February 2022*
Accepted: *31 March 2022*
Published: *20 April 2022*

Citation:
Scholz JL, Eleftherianos I and Kobe B (2022) Editorial: Evolution of Innate Immunity in Eukarya: Advances and Implications. Front. Immunol. 13:879429. doi: 10.3389/fimmu.2022.879429

Immune networks are nearly as old as life itself. A rapidly-expanding body of evidence is revealing remarkable similarities in the recognition of non-self and in the attendant downstream signaling, across eukaryotic kingdoms. Studies are also revealing the likely drivers of the evolution of innate immune networks, both within eukaryotic cells and in the cellular 'dialogue' between the host and the corresponding commensal or pathogenic organisms. This Research Topic focuses on recent advances in our understanding of the origins of innate immunity, conditions that may have shaped its evolution in various species occupying different environments, and co-evolution between microbes and their host. It includes both reviews and primary research articles.

A large cluster of articles in this Research Topic comprises articles on the evolution of innate immune signaling mechanisms. Nimma et al. performed a comprehensive analysis of the structural features of assemblies formed by TIR (Toll/interleukin-1 receptor/resistance) domain-containing proteins, featuring in innate immunity pathways in animals and plants. They report that TIR-domain complexes can be classified into "scaffold" and "enzyme" assemblies, which play distinct roles in innate-immunity and cell-death signaling, respectively. The key enzymatic function featured by enzyme TIR domains is the cleavage of NAD^+ (nicotinamide adenine dinucleotide). An important protein possessing such activity is SARM1, initially characterized as a Toll-like receptor (TLR) adaptor protein. In neurons, the catalytic activity of SARM1 TIR domain is induced upon axon injury and leads to axon destruction. DiAntonio et al. highlight the evolutionarily conserved enzymatic function of TIR-domain proteins as a new class of metabolic regulatory enzymes and discuss the regulatory and functional resemblance between SARM1 activation and bacterial toxin-antitoxin systems.

Ozakman et al. used *Drosophila* and its interaction with the insect parasitic nematodes *Heterorhabditis bacteriophora* as a model to present novel evidence on the function of activin and bone morphogenetic protein (BMP) signaling in the regulation of host immune responses to parasitic nematode infection. They show that BMP signaling activity modulates the DUOX/ROS response and that activin signaling activity modulates the antimicrobial peptide and melanization

responses to nematode infection. These findings contribute to a better understanding of the evolution of the immune role of transforming growth factor β (TGF-β) signaling in invertebrates. Humoral and cellular immune signaling and function are well conserved in insects, but the timing of and duration of the response to bacterial infection is not well understood. Bruno et al. explored the immune reaction times in the black soldier fly *Hermetia illucens*, which comes into contact with a wide range of bacteria that live in decaying substrates. They find that cellular immune processes such as phagocytosis and encapsulation against Gram-negative and Gram-positive bacteria are activated faster than antimicrobial peptide activity in the hemolymph. This information can be used to improve mass-rearing practices, through enhancing insect immune capacity against microbial pathogens.

Comparing cellular immune processes between vertebrates and invertebrates can lead to the discovery of conserved mechanisms. Kokhanyuk et al. compared immune processes in coelomocytes in the earthworm *Eisenia andrei* and human macrophage-like cells. Using endocytosis inhibitors, to examine the engulfment mechanisms of *Escherichia coli* and *Staphylococcus aureus* bacteria, the authors show that although there were differences in intracellular signaling, bacterial internalization through actin-dependent phagocytosis in earthworm and human immune cells is well conserved. Comparative genomics analysis of immune molecules has the potential to reveal their potential conserved roles throughout evolution. Saco et al. performed a comparative genomics study of the key inflammatory cytokines in the interleukin-17 (IL-17) family across different phyla. Comparing multiple genomes from the Mediterranean mussel *Mytilus galloprovincialis*, they find 379 unique IL-17 sequences, which show similar patterns of expansion and variability to IL-17 sequences in other marine invertebrates. Functional validation of certain isoforms through experimental infection with *Vibrio* bacteria confirmed the conserved role of IL-17 in epithelial immune signaling. Li et al. look into the involvement of the newly described amphioxus (cephalochordate) apextrin C-terminal (ApeC)-containing proteins ACP3 and ACP5 in antimicrobial immune responses. They performed expression and microbial binding studies with recombinant proteins, to show that both ACPs bind and aggregate microbes. They exhibit binding specificity to microbial cell wall components, but lack microbial inhibitory activity, and ACP3 regulates the intracellular pathway involving TRAF6 and NF-κB. These findings have the potential to expand interest in understanding the immune role of ACPs in animals.

Recent work has led to an improved understanding of immune signaling and the evolution of immunity to pathogens in plants. Fang and Gu review the regulation of plant immunity by nuclear membrane proteins and highlight functional counterparts in animals. In particular, they discuss recent information on the significant participation of nucleoporins, nuclear transport receptors, and the nuclear lamina in the regulation of nuclear membrane-associated plant immunity. Also, on the topic of immunity in plants, another research article highlights the involvement of miRNAs in modulating the transcriptomic basis of the anti-fungal immune response in the poplar tree, *Marssonina brunnea*. Liao et al. use high-throughput sequencing, together with qPCR validation and functional approaches to identify key fungal-pathogen-responsive genes and miRNAs that take part in the poplar immune response to fungal infection, through regulating pathways that control diverse functions, including plant hormone signaling, antioxidant systems, and lignin biosynthesis.

The Research Topic also includes several articles analyzing the mutual evolutionary pressures between the host and the pathogen and the corresponding arms race. Tsu et al. highlight the evolutionary significance of interactions between viral proteases and the host. The authors summarize recent information suggesting that viral proteases from positive-sense single-stranded RNA genome have evolved to adapt to novel hosts and that this process continues to evolve. They posit that targeting essential host processes by viral proteases occurs in a virus-specific manner and this is indicated by the fact that viral proteases are typically directed against innate antiviral immune proteins. They further describe evolutionary conserved host signaling pathways that are known to detect viral protease activity and the consequences from this dynamic process. An important mechanism for the protection of host cells against viral infection is the necroptotic cell-death pathway. Águeda-Pinto et al. examine the molecular evolution of necroptosis in mammals and show that certain mammalian orders lost the necroptotic pathway during evolution. Their work suggests that disruption of necroptosis in mammalian lineages does not necessarily pose a disadvantage for their development, and that the corresponding naturally infecting poxviruses lost their ability to suppress this pathway. These findings demonstrate a co-evolutionary link between viral infection components and the immune signaling machinery in their hosts. Furthermore, genomic instability in the host may form an evolutionary mechanism to facilitate diversification of immune gene families. Barella Hudgell and Smith perform a detailed bioinformatic and phylogenetic examination of the immune-related *SpTransformer* (*SpTrf*) gene family in the genome of the purple sea urchin, *Strongylocentrotus purpuratus*, which reveals an additional cluster of genes. Analysis of the structural properties of this new gene cluster provides information on its evolutionary origin and offers clues for the type of genomic changes that took place through evolutionary history to expand the number of genes and their sequence variations in this gene family.

The manuscripts in this Research Topic present exciting findings and insights into the evolutionary basis of innate immune signaling and function in a wide range of eukaryotes. This information helps us clarify the fundamental processes that fine-tune the interplay between pathogenic microbes, their effector molecules, and the activity of host defense mechanisms. This is critical knowledge for designing novel means for the genetic manipulation of innate immune cells to combat the emergence of infectious diseases. We thank all the authors, reviewers and editors for their contributions to this collection.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

The research in the authors' laboratories was funded by the National Science Foundation (Award Number 2019869) to IE, the National Health and Medical Research Council (NHMRC) (Project Grant 1160570) and the Australian Research Council (ARC) (Discovery Projects DP190102526, DP220102832 and Laureate Fellowship FL180100109) to BK.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 27 July 2021
doi: 10.3389/fimmu.2021.692997

The Evolution and Diversity of Interleukin-17 Highlight an Expansion in Marine Invertebrates and Its Conserved Role in Mucosal Immunity

Amaro Saco[1], *Magalí Rey-Campos*[1], *Umberto Rosani*[2], *Beatriz Novoa*[1] *and Antonio Figueras*[1]*

[1] Institute of Marine Research (IIM), National Research Council (CSIC), Vigo, Spain, [2] Department of Biology, University of Padova, Padova, Italy

The interleukin-17 (IL-17) family consists of proinflammatory cytokines conserved during evolution. A comparative genomics approach was applied to examine IL-17 throughout evolution from poriferans to higher vertebrates. Cnidaria was highlighted as the most ancient diverged phylum, and several evolutionary patterns were revealed. Large expansions of the IL-17 repertoire were observed in marine molluscs and echinoderm species. We further studied this expansion in filter-fed *Mytilus galloprovincialis*, which is a bivalve with a highly effective innate immune system supported by a variable pangenome. We recovered 379 unique IL-17 sequences and 96 receptors from individual genomes that were classified into 23 and 6 isoforms after phylogenetic analyses. Mussel IL-17 isoforms were conserved among individuals and shared between closely related Mytilidae species. Certain isoforms were specifically implicated in the response to a waterborne infection with *Vibrio splendidus* in mussel gills. The involvement of IL-17 in mucosal immune responses could be conserved in higher vertebrates from these ancestral lineages.

Keywords: interleukin, IL-17, comparative genomics, mucosal immunity, evolution, cytokines, comparative immunology, mussel

OPEN ACCESS

Edited by:
Bostjan Kobe, The University of Queensland, Australia

Reviewed by:
Nguyen T. K. Vo, University of Waterloo, Canada
Jonathan P. Rast, Emory University, United States

***Correspondence:**
Antonio Figueras
antoniofigueras@iim.csic.es

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *09 April 2021*
Accepted: *13 July 2021*
Published: *27 July 2021*

Citation:
Saco A, Rey-Campos M, Rosani U, Novoa B and Figueras A (2021) The Evolution and Diversity of Interleukin-17 Highlight an Expansion in Marine Invertebrates and Its Conserved Role in Mucosal Immunity. Front. Immunol. 12:692997. doi: 10.3389/fimmu.2021.692997

INTRODUCTION

Conserved innate immunity mechanisms based on nonself recognition, signal transduction and immune gene activation are essential for effective protection against bacterial and viral invasions. Invertebrate species present a wide range of recognition mechanisms and molecules that specifically bind pathogens (molecules such as lectins, C1q-containing proteins, fibrinogen-like receptors or FREPs, Toll-like receptors, peptidoglycan recognition proteins, *etc.*), as well as effector agents against infection (antimicrobial peptides, lysozymes, proteases and protease inhibitors) (1). Because of the complexity of adaptive systems in higher vertebrates, certain invertebrate species represent a strong model to study the innate immune mechanisms acquired and conserved during evolution, of which great efficiency and importance have been demonstrated.

The signaling pathways induced by cytokines link recognition molecules with immune effectors, so they are essential for understanding how immune mechanisms are finely regulated. The existence of cytokines, which are classified as interleukins, chemokines and interferons in vertebrates, has

been slowly revealed in several invertebrate lineages (2, 3). Cytokines are soluble immune-regulatory molecules that can modulate the expression of immune genes in defensive cells and lead to proinflammatory and innate immune signaling functions from mucosal or epithelial cells under infection situations (4, 5).

Interleukin 17 (IL-17) is one of the most conserved cytokines among animal phyla; therefore, it is thought to be a key player in innate immunity. IL-17 was originally identified in humans by Rouvier et al. (6) as a cytolytic T lymphocyte (CTL)-associated antigen, and it was cloned and studied for the first time as an interleukin by Yao et al. (7). In humans, IL-17 is a family of proinflammatory cytokines comprising six members (named IL-17A-F) and five receptors (IL-17RA-E) (8). IL-17 is known to be produced by activated T lymphocytes and other cell types implicated in innate immunity, such as mucosal epithelial cells (4, 9). Their mode of action is based on union into dimers (homodimers or heterodimers) whose activity is dependent on attachment to their receptors (IL-17Rs). Receptors possess a cytoplasmic conserved domain (SEFIR) that interacts with adaptor proteins to induce downstream signal transduction pathways to activate transcription factors such as NF-kB and the expression of immune and proinflammatory target genes such as cytokines and antimicrobial peptides (10–13). Human IL-17 family members display functional diversification. IL-17A and IL17-F induce IL-6, IL-8 and CXC chemokines (8, 14), while other human IL-17 forms are characterized by the induction of type-2 immunity (8, 15, 16).

Human IL-17 family members share low amino acid identity (35% average identity between human forms). Other mammals present the same IL-17 repertoire as humans, but because of the low sequence identities in these cytokines, it is rare to find sequence homologs among phyla. The identification of IL-17 sequences is based instead on the detection of their functional domain, which is characterized by unique structural features among the cysteine knot fold motifs and which implies different receptor interactions than other cytokines (8). This 4-cysteine conserved motif forms intrachain disulphide bonds and facilitates dimerization (17).

Other vertebrates, such as teleost fishes, present similar IL-17 families including 4-7 members (18). *Danio rerio* presents a family with homologous proteins for the A, F, C, and D forms but lacks the B and E forms (19). *D. rerio* instead possesses a novel ligand (IL-17N) characteristic of teleost fishes (20). This ligand was first identified in the fish *Takifugu rubripes* and encodes 7 isoforms (21). Equivalent repertoires have been found in other fishes (22–24), and functional studies have demonstrated their proinflammatory implications (25–27).

Invertebrate phyla possess IL-17 genes as well, although the knot fold motif is characterized by 2 extra cysteines located in positions characterized by two serine residues in chordates (28). Close to vertebrate species is the invertebrate chordate *Ciona intestinalis*, with three IL-17 genes activated by LPS inoculation (29). Several IL-17 and IL-17R genes have been subsequently identified in other invertebrates. Some show particularly large repertoires, as in the purple sea urchin *Strongylocentrotus purpuratus*, with 35 IL-17 genes and 2 receptors (30). IL-17 genes from the sea urchin were classified into 10 different subfamilies, showing functional diversification of their immune implications in the gut epithelium or in immune cells (31). In the mollusc *Octopus bimaculoides*, an expansion of IL-17 genes was also reported when 31 genes were identified in its genome (32).

Il-17 genes have also been identified in bivalves such as *Pinctada fucata martenssi* (33), *Crassostrea gigas* (34, 35), and mussels (*Mytilus galloprovincialis*), with 6 IL-17s and 3 IL-17Rs (35). IL-17 pathway components such as CIKS and TRAF6 have been traced in bivalves, and IL-17 bivalve genes are responsive to LPS or bacterial infections, demonstrating inflammatory functions through NF-kB signaling pathways (34–37).

Bivalves, specifically, *M. galloprovincialis*, represent an interesting species in terms of comparative immunology and genomics. These animals are intertidal filter feeders capable of dealing with the wide range of different pathogens that they are persistently exposed to, achieving extraordinary survival success (38–40). This resilience to pathogens and their invasive behavior make mussels an interesting model to investigate genomic adaptations that could explain these phenomena. The recently published *M. galloprovincialis* pangenome was characterized by a core set of 45,000 genes and a set of dispensable genes (20,000) which are affected by presence/absence phenomena and therefore could be absent in individual genomes. This phenomena, in addition to the great variability of specific gene families, could endow the species with a more diverse and enriched response capacity (41).

In bivalves, only a few cytokines have been identified in addition to IL-17, including TNF homologs, allograft inflammatory factor 1, macrophage migration inhibitory factor, and astakine, and the presence and modulation of an interferon-related pathway after viral infection could indicate the presence of an interferon-homologue cytokine not yet discovered (42–45). Additionally, some molecules initially known as antimicrobial peptides, such as myticin C in *M. galloprovincialis*, were revealed to have chemotactic functions as well as modulatory activity over some immune genes, acting as cytokines (46). Myticin C genes are subjected to massive variation and presence/absence phenomena in their variants in the mussel pangenome (47).

IL-17 inflammatory triggering functions seemed to be of importance in the mussel immune response in gills after the recognition of a bath bacterial infection (48). IL-17 genes were modulated in gills against the incoming infection, while there was no modulation in the defensive cells (the hemocytes), against a systemic infection with the same bacteria (48, 49). This might be another example of the role of IL-17 in the immune response of epithelial cells in different species (4, 31).

In the present work, we carried out a comparative genomics analysis of the IL-17 gene family across different phyla of interest. We were able to reveal potential ancestor species and conserved patterns in the evolution of IL-17 repertoires in certain animal classes. The genomic resources originating from the sequencing of 16 individuals to construct the mussel pangenome (41) allowed us to confirm that Il-17 did not show the presence/absence phenomena characteristic of numerous immune gene families previously described. Instead, there was a similar

repertoire, in terms of expansion and variability, to those found in other marine invertebrates. Functional experiments and transcriptomic data for studying the implications of these cytokines in the mussel immune response against a waterborne infection in gills suggested functional specialization of the IL-17 family in mucosal immunity. The large IL-17 repertoires of certain marine invertebrates may be associated with different functions, but epithelial immune defenses appear to be conserved throughout evolution to humans.

MATERIAL AND METHODS

Identification and Analysis of IL-17 and IL-17R Sequences in the Mussel Genome

The *M. galloprovincialis* reference genome LOLA (41) was analyzed to identify the IL17-related sequences. Moreover, 15 resequenced mussel genomes (41) were also analyzed to identify putative IL-17 genes not included in the reference. These 16 genomes referred to five Galician female mussels (GALF1, GALF2, GALF3, and PURA apart from the mussel reference genome LOLA), five Galician males (GALM1, GALM2, GALM3, GALM6 and GALM11), three Italian females (ITAF1, ITAF2 and ITAF3) and three Italian males (ITAM1, ITAM2 and ITAM3). Analysis was mainly performed inside the CLC Genomics Workbench v.20 (Qiagen, Hilden, Germany).

First, the previously identified 6 IL-17 and 3 IL17-R mussel sequences (35) were used as queries and blasted (tblastn) with an e-value threshold of $1x10^{-3}$ against a blast database generated from the mussel reference genome. The resulting contigs were manually checked for the presence of complete open reading frames (ORFs), and the resulting genes were recovered. Then, a Pfam domain scan (hmmsearch algorithm) was performed on the predicted proteins from coding sequences of the mussel reference genome. Sequences containing IL17 domains (PF06083) or SEFIR domains (PF08357) were retrieved and compared with those obtained from the BLAST approach to remove redundancies. A final list of IL17 and IL17-R sequences obtained from the mussel reference genome with both methods was generated.

This list was blasted (blastn) against the 15 resequenced mussel genomes (41). All genomic contigs corresponding to IL17 and IL17-R sequences were retrieved from each genome. Blast hits were analyzed as described above, and the obtained ORFs were also translated to their encoded proteins. The final list of IL-17 and IL-17R genes obtained from the 16 mussel genomes was submitted to CD-HIT (50–52), and only unique sequences were retained. The unique nucleotide sequences cleared in the exon coding for their protein domain (IL17 or SEFIR) were aligned and submitted to model testing using CLC Genomics Workbench to test the best-fitting molecular model of evolution for these sequence sets. Jukes–Cantor (53) was obtained as the best-fitting model for IL-17 sequences, while GTR+G with a gamma-distributed rate of variation across sites (54) was the model for the receptors. IL-17 and IL-17R unique sequences were submitted to independent Markov chain Monte Carlo analyses run in Mr. Bayes v3.2.7a (55). The substitution model obtained for each case was considered, and analyses were run with a sampling frequency of 1,000 and a burn-in of 25% sampled trees until the average standard deviations of the split frequencies were ≤ 0.05. The analyses were run for 4 million generations with IL-17 and 350,000 generations with the receptors. The obtained consensus trees were graphically rendered using FigTree (56) and iTOL (57).

Clustering of mussel IL17 and IL17-R sequences was performed based on phylogenetic analyses. CD-HIT analyses with a sequence identity cut-off of 0.8 (80% homology) resulted in the same clusters. Primary sequence analysis was performed with a representative sequence of each cluster or isoform. Conserved domains were analyzed with HMMER 3 (58), transmembrane domains were analyzed using the TMHMM Server v. 2.0 (59, 60) and signal peptides were analyzed using the SignalP-5.0 Server (61, 62).

The genomic location in the reference genome scaffolds, the gene structure and the neighboring genes were also analyzed for the IL-17 and IL-17R genes present in the reference genome.

Phylogenetic Distribution of IL-17 Repertoires Among Species

Genome assemblies and their encoded protein databases were downloaded from the NCBI for several species of interest across different phyla (IDs are displayed). The analysis strategy was equivalent to that performed in mussels.

For each species, previous IL-17 sequences were downloaded from the Pfam database and used as a query seed list that was blasted in a tblastn search against the corresponding genome assembly (e-value threshold of $1x10^{-3}$). For species with no previous sequences with IL17 domains, sequences of the phylogenetically closest species were downloaded from Pfam and used as queries. BLAST hits were analyzed, and the potential ORFs and protein sequences were retrieved. In a complementary way, a Pfam domain scan (hmmersearch) was conducted for the encoded proteins of each downloaded genome. CD-HIT was used to remove redundant sequences obtained from both methods for each species (identity cut-off of 1).

The exceptions to this pipeline were *Homo sapiens* and *Mus musculus*, since their IL17 gene repertoires were retrieved directly from the UniProt database (Q16552, Q9UHF5, Q9P0M4, Q8TAD2, Q9H293, Q96PD4, Q62386, Q9QXT6, Q8K4C5, A0A0B4J1G4, Q8VHH8, and Q7TNI7).

The IL-17 families from each species were clustered with the same 80% homology criteria. The evolutionary cladogram of the studied species was built using the TimeTree resource of MEGA-X Software (63).

A subset of selected species was submitted to MEME Suite 5.1.1 (64) to identify conserved motifs among their IL17 sequences.

Several Mytilidae species were also analyzed by searching for repertoires homologous to those discovered in *Mytilus galloprovincialis*. Genomic or transcriptomic assemblies were downloaded from the NCBI-SRA database, depending on the material available (IDs are displayed). A list containing

representative sequences from each IL-17 isoform found in *M. galloprovincialis* was used as a query in a BLASTn search against these Mytilidae assemblies (e-value threshold of $1x10^{-3}$). BLAST hits were analyzed, and the potential ORFs were retrieved and translated to the protein sequences, confirming the presence of the IL-17 domain. The nucleotide coding sequences were submitted to phylogenetic neighbor joining analysis using the Jukes–Cantor substitution model (the best fitting molecular model of evolution for these sequences).

Expression of IL-17 and IL17-R in Transcriptomic Studies

Transcriptomic data were recovered from two different experiments performed with individual *Mytilus galloprovincialis* mussels, including a gill transcriptome obtained after 24 h of waterborne exposure to an infection with *Vibrio splendidus* (48) and a transcriptome obtained from hemocytes 24 h after injection with the same bacteria, with the goal of establishing a systemic infection (49). Assemblies were constructed as explained in Saco et al. (48) and in Rey-Campos et al. (49), and reads from each transcriptome are accessible through PRJNA638821 and PRJNA466718, respectively. IL-17 and IL-17R protein sequences from the isoforms of the reference genome were used in a tblastn search against those assemblies. Contigs corresponding to the blast hits were obtained and aligned to verify their classification within a specific IL17/IL17-R cluster or isoform. Transcriptomic expression was retrieved in transcripts per million (TPM).

Mussel Bath Infection and Tissue Sampling

Adult *Mytilus galloprovincialis* mussels were obtained from a commercial shellfish farm (Vigo, Galicia, Spain) and acclimatized in tanks with open-circuit filtered seawater (FSW) at 15°C with aeration. Mussels were submitted to a waterborne exposure to 10^8 CFU/ml *Vibrio splendidus* (reference strain LGP32), while control mussels were maintained in FSW.

Samples from hemocytes and gills were obtained after 24h of waterborne exposure to the bacterial infection. Hemolymph was withdrawn from the posterior adductor muscle using a 0.5 mm-diameter (25 G) disposable needle through a hole made in the shell. Hemolymph was centrifuged at 4°C at 1000xg for 20 min, and the hemocyte pellet was suspended and homogenized in the same volume of FSW. Gills were sampled and homogenized in FSW, as well.

RNA Extraction, cDNA Synthesis and Quantitative Real-Time PCR (qRT-PCR)

RNA was extracted from individual triplicates of infected and control hemocyte and gill samples. RNA extraction was carried out using the Maxwell 16 LEV simply RNA kit (Promega, Madison, Wisconsin, United States), and its concentration and purity were measured with a NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, United States). cDNA was synthetized by reverse transcription from 300 ng of total RNA from each sample using an NZY First-Strand cDNA Synthesis Kit (Nzytech, Lisboa, Portugal).

The gene expression of several mussel IL17 and IL17-R genes (**Supplementary Table ST1**) was analyzed in a Step One Plus qPCR System (Applied Biosystems, Foster City, California, United States). The selected isoforms were those showing expression in the previously mentioned gill transcriptome. The cycling conditions were 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 30 s. Reactions were performed in technical triplicates, and the relative expression was normalized using highly stable 18S gene expression as a housekeeping gene, following the delta-delta-CT method (65). Significant differences were analyzed (t-test, p-value ≤ 0.05).

In Situ Hybridization in Tissue Samples

In situ hybridization (ISH) was carried out to localize the expression of the IL17-3 gene in gill histology preparations with digoxigenin-labelled IL17-3-specific RNA probes. This isoform was selected because it was one of the most expressed isoforms (35), and it was modulated in the analyzed gill transcriptome. To obtain the probes, mussel cDNA was amplified in two different PCRs combining pairs of IL17-3 primers, one of which was a normal primer at 10 mM, and the other included the attached SP6/T7-corresponding probe sequence (**Supplementary Table ST1**) at 100 mM (Fsp6 + R; F + RT7). Cycling conditions were set to 94°C 5 min, 35 cycles of 94°C 30 s + 60°C 30 s + 72°C 1 min and a final step of 72°C 7 min. PCR products were filtered and concentrated using Amicon Ultra 0.5-10k devices following kit instructions (Sigma-Aldrich, St. Louis, Missouri, United States). Digoxigenin-labelled IL17-3-specific RNA probes, both antisense (As) and sense (S; control), were obtained by *in vitro* transcription of 1 µg of purified cDNA from each PCR product using the DIG RNA labelling kit (SP6/T7) and following the kit instructions (Roche, Basel, Switzerland). Finally, both sense and antisense probes were purified using SigmaSpin Sequencing Reaction Clean-Up (Sigma-Aldrich). After each step of the probe preparation process, the cDNA/RNA concentration was measured using a NanoDrop ND1000 spectrophotometer, and electrophoresis gels were run to check the specificity.

ISH assays were performed simultaneously on two sets (As and S) of duplicate slides, each with serial sections of gill from the same animal. Briefly, mussels were fixed using formaldehyde and embedded in paraffin. Seven-micrometer cut sections were placed on polylysine-coated glass slides (Thermo Scientific, Waltham, Massachusetts, United States), dried and submitted to deparaffinization and rehydration (passes with xylene, 100%, 95% and 70% ethanol and DEPC-treated ddH_2O). After incubation with DEPC-PBS (2x5 min) and DEPC-PBS with 100 mM glycine (2x5 min), tissue sections were treated with DEPC-PBS 0.3% Tween-20 (15 min) and washed again with DEPC-PBS (2x5 min). Afterward, a permeabilization step with 20 µg/ml proteinase K in TE buffer (100 mM Tris-HCl, 50 mM EDTA, pH 8) at 37°C°C for 30 min was performed. Finally, sections were postfixed with PBS-PFA 4% (5 min, 4°C), washed and acetylated for 20 s with 20% acetic acid.

Prehybridization was carried out by incubating tissue sections for 90 min at 37°C in a humid chamber with hybridization buffer

(40% formamide, 10% dextran sulphate, 1x Denhardt's solution, 4x SSC, 10 mM DTT, 1 mg/ml yeast t-RNA, 1 mg/ml salmon sperm DNA). Hybridization buffer was added again and supplemented with 50 ng of the corresponding digoxigenin (DIG)-labelled RNA probe for each slide. Sections were incubated at 42°C overnight in a humid chamber.

After incubation, the samples were washed at 37°C in 2x SSC (2x15 min) and 1x SSC (2x15 min). To digest any single-stranded nonspecifically bound probes, incubation for 30 min at 37°C with 20 µg/ml RNaseA in NTE buffer (500 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 8) was performed. The slides were again washed at 37°C in 0.1x SSC (2x30 min), incubated in a buffer containing 100 mM Tris-HCl and 150 mM Nacl at pH 7.5 (2x10 min), then incubated for 30 min in a blocking solution with 0.1% Tween-20 and 2% normal sheep serum added to the buffer, and finally incubated for 2 h in a humid chamber after supplementing the buffer with 0.1% Tween-20, 1% normal sheep serum and a dilution 1:250 of anti-DIG-alkaline phosphatase antibody [Fab fragments]. After two washes with the buffer, the slides were incubated for 10 min with 100 mM Tris-HCl, 100 mM NaCl, and 50 mM $MgCl_2$ at pH 9.5, and then 1 mM levamisole, 4.5 µl/ml nitroblue tetrazolium (NBT) and 3.5 µl/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) solutions were added. Finally, the slides were covered with this solution and incubated in darkness in a humid chamber until an easily visible color was found due to alkaline phosphatase detection. The color reaction was stopped with 10 mM Tris-HCl and 1 mM EDTA at pH 8.1, and the slides were washed and finally mounted using DPX.

RESULTS

Screening and Classification of the Mussel IL-17 and IL-17R Repertoire

In total, 552 sequences of IL-17 were identified after applying the screening pipeline to 16 mussel genomes (41), resulting in 379 unique sequences that are deposited in **Supplementary Data 1**. These sequences were clustered based on an identity percentage threshold of 80%, yielding 23 clusters. The Bayesian tree of the 379 unique sequences displayed diversification in the same 23 clusters (**Figure 1**). Six out of 23 IL-17 forms (1-6) were already described by Rosani et al. (35). The new forms were named using the numbers 7 to 23.

As shown in **Table 1**, in each genome, the number of IL-17 genes varied between 29 and 42, with an average repertoire of 34 genes per mussel. These variants were classifiable into the 23 defined clusters. There was no presence/absence phenomenon at the isoform level. Instead, every individual mussel had its own

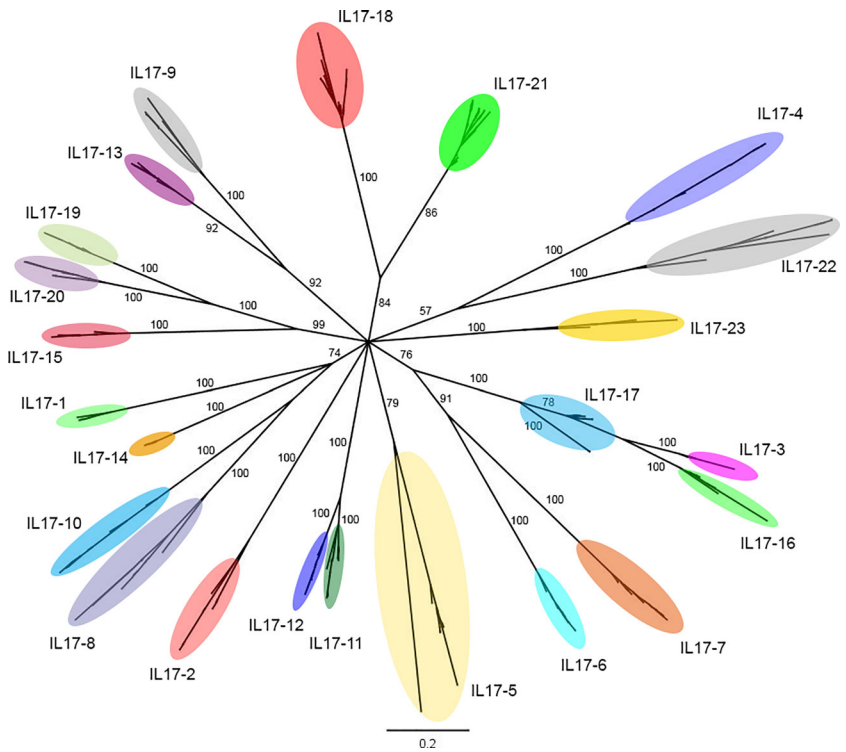


FIGURE 1 | Phylogenetic analysis of the 379 unique IL-17 sequences obtained from the 16 mussel genomes. The 379 unique IL-17 sequences retrieved from 16 mussel genomes were submitted to phylogenetic analysis with Bayesian phylogenetic inference using the Jukes and Cantor (JC) evolution model. The 23 different IL-17 clusters or isoforms obtained are indicated with ellipses that englobe all their unique variants. The figure shows branch posterior probabilities. It can be seen how isoforms IL17-3 and IL17-16 would be diverging from IL17-17, since the branch of the latter contains both isoforms. The distinction between these isoforms was maintained due to the homology criterion and to expression evidence.

TABLE 1 | Sequences present in each of the 16 analyzed mussel genomes that correspond to the different IL-17 and IL-17R isoforms.

	Sequenced mussel genomes															
IL17	**LOLA**	**GALF1**	**GALF2**	**GALF3**	**GALM1**	**GALM2**	**GALM3**	**GALM6**	**GALM11**	**ITAF1**	**ITAF2**	**ITAF3**	**ITAM1**	**ITAM2**	**ITAM3**	**PURA**
IL17-1	1	1	1	1	3	1	2	1	2	1	1	1	2	1	2	1
IL17-2	2	1	2	2	1	1	1	2	1	2	2	1	1	1	1	1
IL17-3	2	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1
IL17-4	2	1	1	1	1	1	1	2	2	2	1	1	2	2	2	1
IL17-5	1	1	1	1	1	1	1	2	2	1	1	1	2	2	1	1
IL17-6	2	2	2	1	1	2	2	2	2	1	1	1	1	1	1	2
IL17-7	2	1	2	1	2	2	2	1	1	2	1	2	2	1	1	1
IL17-8	1	1	2	1	1	2	2	2	2	1	1	2	2	2	2	1
IL17-9	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
IL17-10	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1
IL17-11	2	3	2	2	4	2	3	2	2	2	2	1	2	2	2	2
IL17-12	1	2	2	1	1	1	1	2	1	–	–	1	1	2	1	1
IL17-13	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1
IL17-14	1	1	1	1	2	1	1	2	1	2	1	2	2	2	2	1
IL17-15	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1
IL17-16	1	1	2	1	2	1	1	2	1	–	1	1	1	1	1	1
IL17-17	1	2	1	2	2	1	1	1	1	2	2	2	1	1	2	1
IL17-18	3	4	4	4	5	4	3	4	4	3	3	3	4	2	2	2
IL17-19	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1
IL17-20	1	1	2	2	2	2	2	2	1	2	2	1	1	1	2	2
IL17-21	3	5	5	1	3	2	5	5	5	2	4	2	2	5	3	3
IL17-22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IL17-23	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1
Total	33	36	39	32	40	32	38	42	36	31	31	30	34	35	34	29
IL17R	**LOLA**	**GALF1**	**GALF2**	**GALF3**	**GALM1**	**GALM2**	**GALM3**	**GALM6**	**GALM11**	**ITAF1**	**ITAF2**	**ITAF3**	**ITAM1**	**ITAM2**	**ITAM3**	**PURA**
IL17R-A	2	1	2	2	2	1	1	1	1	1	1	1	2	1	1	1
IL17R-B	1	1	1	1	1	–	1	1	1	1	1	1	1	1	1	1
IL17R-C	1	1	1	1	1	2	2	1	1	2	1	1	1	1	1	1
IL17R-D	2	1	2	2	1	1	2	1	1	2	1	2	1	1	1	1
IL17R-E	1	2	2	2	1	1	1	2	1	1	1	1	1	1	1	1
IL17R-F	1	1	2	1	1	–	2	1	1	1	2	1	1	1	1	1
Total	8	7	10	9	7	5	9	7	6	8	7	7	7	6	6	6

unique sequences belonging to the 23 clusters, characterized by variations in specific positions and resulting in a large number of 379 nonshared sequences. Isoforms IL17-18 and IL17-21 were the most diverse, presenting up to 5 different variants in some mussels.

Considering this conservation at the isoform level, representative sequences were selected from the reference genome for each isoform (listed in **Supplementary Data 2**). As indicated in **Table 2**, they ranged from 126 to 221 amino acids, always with an IL-17-defined domain towards the C-terminus. The signal peptide was predicted in 12 out of 23 consensus isoforms. The multiple sequence alignment of the conserved domains is displayed in **Supplementary Figure S1**. Every isoform presented 6-8 cysteine residues in the functional domain, which would conform to the at least three disulphide bonds typical of invertebrate IL-17s. Isoforms 18 and 21 showed only 4 conserved cysteine residues, and therefore, they could form only the two disulphide bonds characteristic of chordate species.

In total, 115 nucleotide IL-17 receptor sequences were retrieved from the 16 mussel genomes, resulting in 96 unique sequences (deposited in **Supplementary Data 3**). These unique sequences were clustered into 6 isoforms (identity percentage threshold of 80%) and submitted to phylogenetic analysis, which resulted in the same 6 isoforms or clusters (**Figure 2**). Again, as with the IL-17 genes, variability was enormous inside each cluster, but no presence/absence phenomenon was observed at the isoform level (**Table 1**). Three out of six clusters were already described by Rosani et al. (35), which were A, B and C. The three new clusters were named D, E and F to maintain the nomenclature system already in use for mussel, but this is not related to possible vertebrate orthologous. Representative sequences for each isoform were retrieved from the reference genome (**Supplementary Data 4**). SEFIR domains were conserved, as were the transmembrane regions that separate extracellular and cytoplasmic domains, indicating that these receptors would be functional in terms of signaling (**Table 3**).

IL-17 Gene Families Throughout Evolution

Genomic or transcriptomic assemblies from different Mytilidae species were scanned for IL-17 sequences, and those matching the requirements were retrieved (**Table 4**). Every IL-17 gene from the studied Mytilidae species was homologous to a certain isoform of *M. galloprovincialis*. Hence, the resulting phylogenetic tree was clustered by isoform and not by species (**Figure 3**). *Mytilus coruscus* was the closest species, presenting every IL-17 isoform found in *M. galloprovincialis*. The most variable isoforms, IL17-18 and IL17-21, were characterized by two different genes each in the *M. coruscus* genome. Most isoforms were only conserved with homologous sequences in *M.*

TABLE 2 | Sequence and domain information of the 23 isoforms of mussel IL-17.

IL17 cluster representatives	Predicted Protein Length	Signal peptide Sec/SPI Likelihood	Predicted signal peptide cleavage site	IL17 domain e-value (HMMER3)	IL17 domain position
IL17-1	194	0.6047	27 to 28	1.1e-13	102 to 185
IL17-2	192	0.9448	18 to 19	1.2e-16	89 to 168
IL17-3	194	0.9842	23 to 24	3.7e-12	94 to 169
IL17-4	165	0.9053	26 to 27	3.4e-17	82 to 157
IL17-5	221	0.5281	23 to 24	2.2e-12	95 to 173
IL17-6	192	0.7977	24 to 25	8.8e-14	94 to 170
IL17-7	192	0.937	22 to 23	4.4e-13	95 to 170
IL17-8	134	0.0033	-	2.4e-13	46 to 119
IL17-9	169	0.0008	-	2.4e-19	60 to 134
IL17-10	180	0.9751	20 to 21	1.7e-14	85 to 162
IL17-11	179	0.0021	-	4.1e-17	91 to 166
IL17-12	194	0.939	22 to 23	8.1e-19	106 to 181
IL17-13	198	0.4734	-	1.6e-19	87 to 161
IL17-14	185	0.9211	17 to 18	1.7e-16	82 to 168
IL17-15	185	0.0654	-	6.5e-14	85 to 167
IL17-16	194	0.4703	-	1.4e-11	94 to 169
IL17-17	190	0.9332	19 to 20	4.4e-12	90 to 166
IL17-18	126	0.0043	-	3.3e-08	43 to 116
IL17-19	133	0.0019	-	2.8e-15	53 to 127
IL17-20	137	0.001	-	5e-14	55 to 131
IL17-21	116	0.0033	-	6.4e-12	41 to 114
IL17-22	174	0.0009	-	1.9e-17	85 to 169
IL17-23	133	0.996	20 to 21	3.6e-05	53 to 131

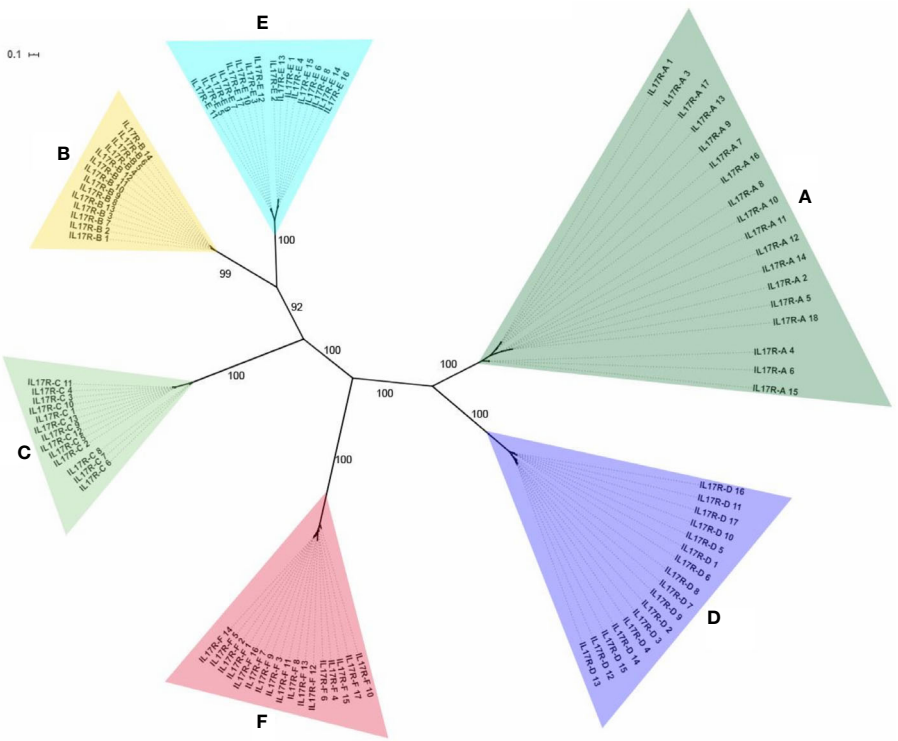


FIGURE 2 | Phylogenetic analysis of the 96 unique IL-17R sequences obtained from the 16 mussel genomes. The 96 unique IL-17R sequences retrieved from 16 mussel genomes were submitted to phylogenetic analysis with Bayesian phylogenetic inference using the GTR+G evolution model. The 6 different IL-17 clusters or isoforms obtained are indicated. The figure shows branch posterior probabilities.

TABLE 3 | Sequence and domain information of the 6 isoforms of mussel IL-17R.

IL17 receptor isoforms in LOLA genome	Predicted Protein Length	Transmembrane region (TMHMM 2.0)	SEFIR domain e-value (HMMER3)	SEFIR domain position
IL17R-A	636	221 to 243	7.8e-12	286 to 437
IL17R-B	727	350 to 372	3.4e-16	401 to 542
IL17R-C	566	13 to 35	9.9e-30	71 to 223
IL17R-D	778	353 to 375	6.2e-13	429 to 580
IL17R-E	685	353 to 375	9.7e-12	407 to 549
IL17R-F	874	386 to 409	6.6e-12	467 to 602

TABLE 4 | Homologous sequences to the *Mytilus galloprovincialis* IL-17 isoforms found in other analyzed Mytilidae species.

Species	Assembly	Project code	Contigs count	Technology	IL17 Forms
Mytilus coruscus	Genomic	PRJEB33342	10484	PromethION DNA sequencer	25
Mytilus californianus	Transcriptomic	PRJNA375125	59027	Illumina NextSeq 500	0
Mytilus edulis	Genomic	PRJNA525607	353272	Illumina HiSeq	7
Perna viridis	Transcriptomic	PRJNA478494	73264	Illumina HiSeq	0
Limnoperna fortunei	Genomic	PRJNA330677	61104	Illumina NextSeq 500; PacBio	3
Mytilus trossulus	Transcriptomic	PRJNA525608	437716	Illumina HiSeq	6
Modiolus philippinarum	Genomic	PRJNA328544	74573	Illumina HiSeq	9
Bathymodiolus platifrons	Genomic	PRJNA328542	272497	Illumina HiSeq	5

galloprovincialis and *M. coruscus*, while others, such as IL17-2, IL17-3, IL17-9 or IL17-14, were quite conserved across the different Mytilidae species studied.

Species belonging to different animal phyla were also analyzed in search of their IL-17 gene repertoires (**Table 5**). The results revealed putative ancestor species and expanded IL-17 repertoires in certain invertebrate species (**Figure 4**). We could not identify any putative IL-17 gene in the sponge *Amphimedon queenslandica* or in the first divergent cnidarian analyzed, *Hydra vulgaris*. Nevertheless, the two other cnidarians analyzed displayed relatively high numbers of IL-17 sequences, reaching 11 genes in *Pocillopora damicornis*, which were classified into 8 clusters. In the present work, Cnidaria emerged as the most ancient divergent phylum with IL-17 gene families. Brachiopoda and Annelida species included in the analysis did show different repertoires present in all species, as did the studied Nematoda species (*T. spiralis* and *C. elegans*). Concerning Mollusca, there was a common evolutionary pattern, with large expanded IL-17 repertoires compared to other phyla, as shown in **Figure 4**. This was observed in all the marine bivalves included in the study (*M. yessoensis*, *C. gigas* and *M. galloprovincialis*) and in the marine cephalopod *Octopus bimaculoides*, as reported by Albertin et al. (32). These species surpassed the 30 genes each and clustered into 23 isoforms, except *C. gigas*, with 24 genes and 16 clusters (**Table 5**). Analyzed mollusc species from the Gastropoda class (*Biomphalaria glabrata* and *Aplysia californica*) did not show these expanded IL-17 families.

Among Arthropoda species, crustaceans such as *Penaeus vannamei* and *Daphnia magna* contained IL-17 repertoires. IL-17 has not been previously described in Insecta, with no reports of these genes in the class (28). However, one sequence with the typical structure of an interleukin, with 218 amino acids, a signal peptide and with the C-terminal IL-17 domain (e-value of 4.7e-05) emerged from the *Drosophila melanogaster* genome in our study. This did not occur in the other analyzed insect, *Apis mellifera*.

The echinoderm *Strongylocentrotus purpuratus* displayed a large repertoire, which was closer to that of the studied marine molluscs. We were able to find 26 sequences in its genome. However, the presence of up to 35 IL-17 sequences in the genome of this species has been reported, and the repertoire must be larger than the revealed from the assembly with which we worked (30, 31). The whole repertoire of 35 genes is reported to cluster into 10 subfamilies or isoforms, which is more consistent with the isoforms that we found (31). The identified sequences may be suitable to represent the diversity of *S. purpuratus* in this study, since the comparative genomics analysis was focused on an approach based on isoforms instead of genes. As in Mollusca, not all echinoderms displayed expanded IL-17 repertoires, since only 4 genes were found in *Asterias rubens*.

We also described, for the first time, an IL-17 repertoire in hemichordates, in particular, *Saccoglossus kowalevskii*. IL-17 had not been previously reported in this phylum. Focusing on Chordata species, we studied the invertebrate *Ciona intestinalis* and some vertebrate species, finding families with similar numbers of IL-17 isoforms (± 6) and being the amphibian *Xenopus laevis* the one with the richest repertoire.

Due to the low sequence identities that IL-17 sequences display, homologous forms can be found only in phylogenetically close species, as seen with Mytilidae or typically reported among mammals or vertebrates. This can be observed in the phylogenetic analysis of IL-17 genes from different phyla, as shown in **Supplementary Figure S2**. Genes from different species do not share high sequence homology but maintain a conserved structure. The IL-17 domain emerges,

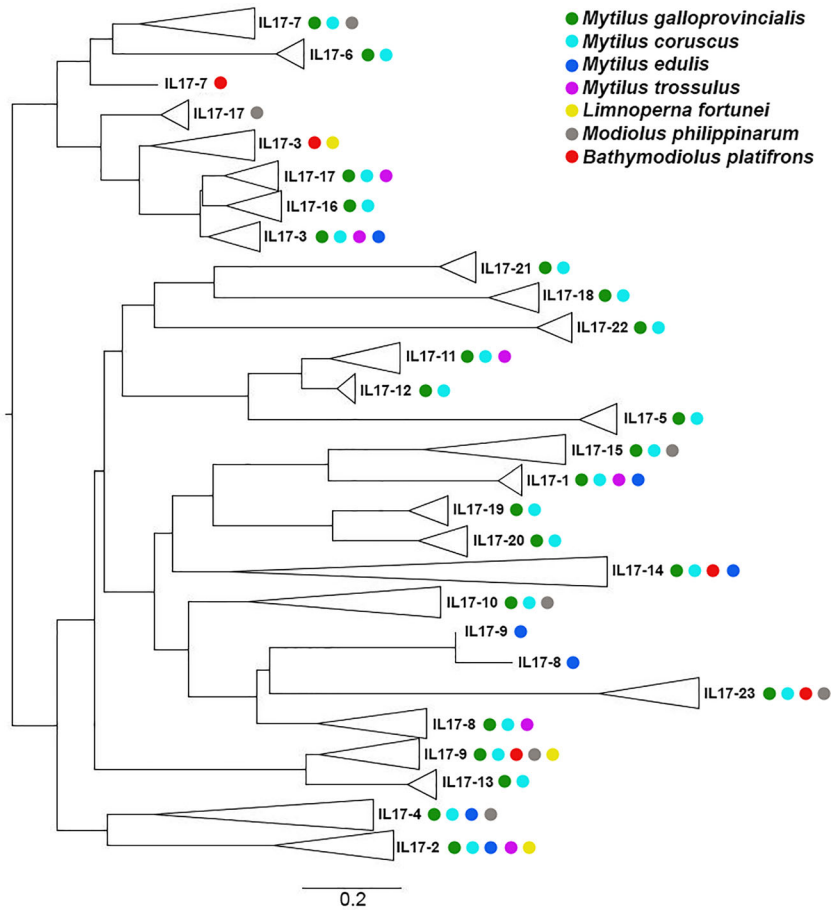


FIGURE 3 | Phylogenetic analysis of the IL-17 families found in Mytilidae species. Homologous sequences to the *M. galloprovincialis* IL-17 isoforms were found in several Mytilidae species and submitted to neighbor joining analysis using the Jukes–Cantor substitution model. Clustering was based on isoform homologs, and species are indicated with colors.

supported by the mentioned cysteine pattern and conserved motifs, as recurrent CPW/SPW in invertebrates and chordates, respectively (**Supplementary Figure S3**).

The study of possible synteny conservation between mussel isoforms and the sequences of other studied species revealed a common pattern in the expanded repertoires, where IL-17 genes appeared repeatedly in tandem duplications in a single scaffold/chromosome. This observation was particularly notorious in the expanded repertoires of marine invertebrates such as *M. galloprovincialis*, *M. yessoensis*, *O. bimaculoides* and *S. purpuratus* and in the hemichordate *S. kowalevskii* (**Figure 5**).

The gene structure and genomic context were also analyzed (**Supplementary Table ST2**). The 33 IL-17 genes found in the mussel reference genome (LOLA) were located on 18 scaffolds, while the 8 receptor genes were located on 5 scaffolds. Gene structure was variable, being 2 exons the most repeated structure in IL-17 and receptors presenting a greater number of exons, up to 11. There was no clear synteny conservation between any mussel isoform and other studied species, which could had suggested a possible conserved ancestral IL-17 gene or isoform (**Supplementary Figure S4**). Synteny conservation was only maintained when studying phylogenetically close species that share homologous IL-17 genes, as mentioned previously with vertebrates (**Supplementary Figure S5**).

IL-17 Modulation After a Bath Infection

IL-17 expression retrieved from transcriptomic data of mussels subjected to bacterial infections differed among isoforms. Two transcriptomic experiments were analyzed: a bath or waterborne infection (24 h) sampled in gills (48) and a systemic infection (24h) sampled in hemocytes (49). As displayed in **Figures 6A, B**, the set of isoforms expressed was stimulus-specific and conditioned by the route of infection. There was a large diversity concerning which isoforms are expressed in each individual animal. Clusters such as IL17-3 or IL17-14 were among the most commonly expressed, while others were detected exclusively in the gill transcriptome samples (IL17-1, IL17-2 or IL17-10) or exclusively in hemocytes (IL17-17, IL17-15 or IL17-23). Some isoforms were even expressed only in a certain sample, i.e., not subjected to any stimulus or tissue (IL17-11,

TABLE 5 | IL-17 genes and isoforms found in the analyzed species from different phyla.

Phylum	Species	NCBI Genome ID	RefSeq/GenBank assembly accession	IL-17 unique sequences	IL-17 clusters (80%)
Porifera	*Amphimedon queenslandica*	2698	GCF_000090795.1	–	–
Cnidaria	*Hydra vulgaris*	12836	GCF_000004095.1	–	–
Cnidaria	*Pocillopora damicornis*	22550	GCF_003704095.1	11	8
Cnidaria	*Nematostella vectensis*	230	GCF_000209225.1	3	3
Brachiopoda	*Lingua anatina*	38582	GCF_001039355.2	4	2
Annelida	*Capitella teleta*	15118	GCA_000328365.1	11	8
Annelida	*Helobdella robusta*	15112	GCF_000326865.1	1	1
Mollusca - Gastropoda	*Biomphalaria glabrata*	357	GCF_000457365.1	8	7
Mollusca - Gastropoda	*Aplysia californica*	443	GCF_000002075.1	4	4
Mollusca - Bivalvia	*Mytilus galloprovincialis*	12190	GCA_900618805.1	33	23
Mollusca - Bivalvia	*Crassostrea gigas*	10758	GCF_902806645.1	24	16
Mollusca - Bivalvia	*Mizuhopecten yessoensis*	12193	GCF_002113885.1	31	23
Mollusca - Cephalopoda	*Octopus bimaculoides*	41501	GCF_001194135.1	36	23
Nematoda	*Caenorhabditis elegans*	41	GCF_000002985.6	4	3
Nematoda	*Trichinella spiralis*	238	GCF_000181795.1	1	1
Arthropoda - Insecta	*Drosophila melanogaster*	47	GCF_000001215.4	1	1
Arthropoda - Insecta	*Apis mellifera*	48	GCF_003254395.2	–	–
Arthropoda - Crustacea	*Penaeus vannamei*	10710	GCF_003789085.1	6	5
Arthropoda - Crustacea	*Daphnia magna*	10953	GCF_003990815.1	2	2
Echinodermata	*Asterias rubens*	83686	GCF_902459465.1	4	4
Echinodermata	*Strongylocentrotus purpuratus*	86	GCF_000002235.5	26	12
Hemichordata	*Saccoglossus kowalevskii*	359	GCF_000003605.2	9	7
Chordata - Tunicata	*Ciona intestinalis*	49	GCF_000224145.3	7	6
Chordata - Actinopterygii	*Salmo salar*	369	GCF_000233375.1	10	8
Chordata - Actinopterygii	*Danio rerio*	50	GCF_000002035.6	5	5
Chordata - Amphibia	*Xenopus laevis*	81	GCF_017654675.1	18	11
Chordata - Aves	*Gallus gallus*	111	GCF_016699485.2	5	5
Chordata - Mammalia	*Mus musculus*	52	GCF_000001635.27	6	6
Chordata - Mammalia	*Homo sapiens*	51	GCF_000001405.39	6	6

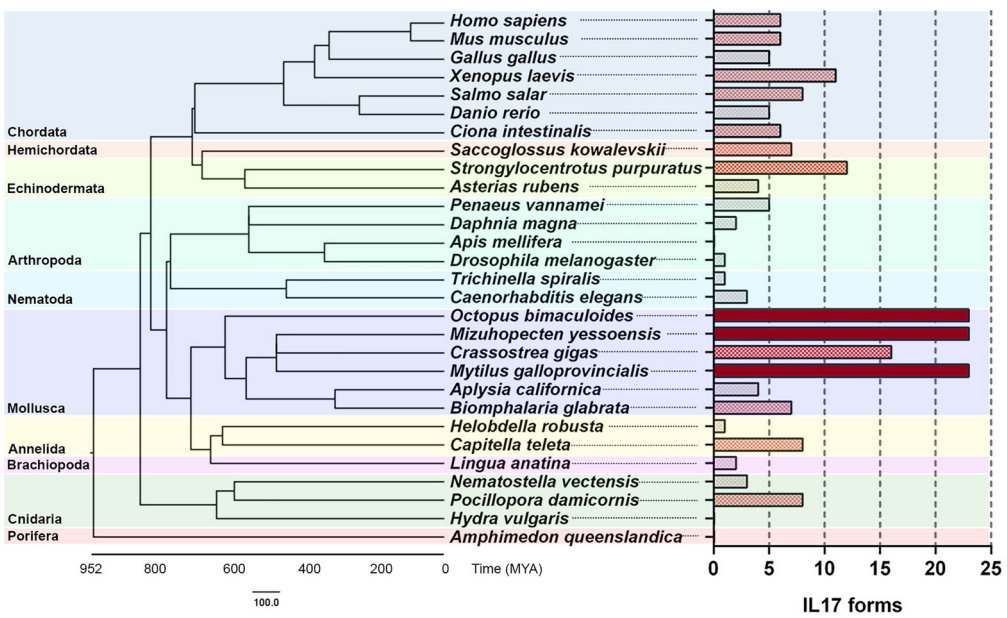


FIGURE 4 | Comparative genomics analysis of IL-17 gene families throughout evolution. IL-17 repertoires were analyzed in several species from different phyla and clustered with an 80% homology criterion. The evolutionary cladogram is presented, along with the extension of the IL-17 family from each species.

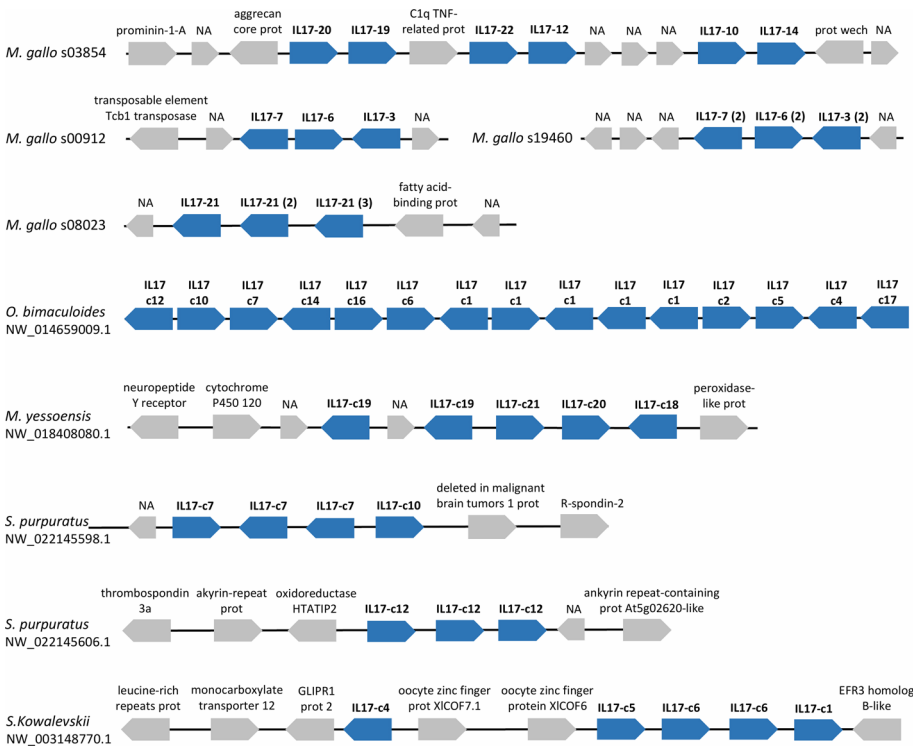


FIGURE 5 | IL-17 genes often appear in tandem repetitions in the genomes of marine invertebrates with expanded IL-17 repertoires. Examples of several scaffolds with IL-17 tandem duplications are presented for these species. IL-17s are indicated with the name of the cluster to which they belong in our analysis. Tandem repetitions can occur with genes from the same or different isoforms. Neighbor genes are represented with their functional annotations.

IL17-6). The bacterial waterborne infection increased the general expression levels of IL-17 forms in gills (**Figure 6A**). In hemocytes, changes in the expressed isoforms between individuals and between control and infected conditions were observed as well, but the differences in expression were not statistically relevant (**Figure 6B**). Receptors did not show the same diversity in terms of the different forms that are expressed in different individuals. However, it was noticeable that IL-17RA was not expressed in gills (**Figure 6C**) while IL-17RF was not expressed in hemocytes (**Figure 6D**).

To explore the modulation under a waterborne infection in more detail, the gene expression of the isoforms previously detected in the transcriptomic approach was studied by qRT-PCR. The expression of these IL-17 isoforms in gills and internal hemocytes differed completely. The general upregulation in gills of all expressed isoforms was confirmed, except for IL17-11 and IL17-8 (**Figure 7A**). Instead, in hemocytes sampled from the adductor muscle of the same individuals, IL-17 isoforms were generally downregulated (**Figure 7B**). The up-regulation in gills was significant for IL17-2 and IL17-3 while the down-regulation in hemocytes was significant for IL17-11 and IL17-2. Receptor isoform IL17R-D was up-regulated, while B and C were down-regulated (significant for IL17R-C in hemocytes). Expression in the isoforms E and F did not show any clear regulation trend. The expression of IL-17 receptors was analogous between gills and hemocytes for every expressed receptor, except for IL17R-F. The expression of isoforms IL17R-B, IL17R-C, IL17R-D and IL17R-E showed significant Spearman correlations between gills and hemocytes of 0.83, 0.77, 0.94 and 0.83 (**Figure 7C**).

The most highly expressed isoform (IL17-3) was studied by *in situ* hybridization in histological preparations of mussel gills. The results shown in **Supplementary Figure S6** may suggest that it is expressed not only in infiltrated hemocytes but also in epithelial gill cells.

DISCUSSION

The IL-17 proteins are key inflammatory cytokines encoded by a diverse, expanded gene family in mussels. As indicated previously, 6 IL-17 sequences and 3 receptors have already been described in mussels (35), but the advantages provided by genome sequencing and resequencing data (41), allowed us to reveal a whole new perspective on this gene family. The analysis of 16 mussel genomes revealed 379 unique IL-17 sequences. A similar level of diversity was observed for the receptors, with 96 unique variants. Once all of these sequences were clustered, there was conservation of almost every isoform among the 16 individual genomes. Therefore, this high variability seems to be different from the phenomenon of presence/absence variation

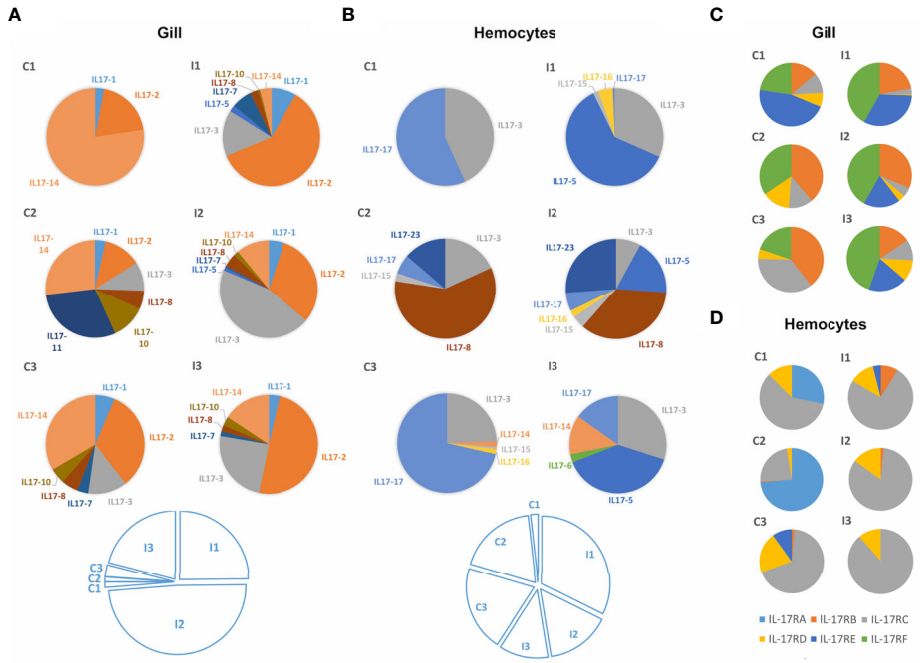


FIGURE 6 | Expression of the different mussel IL-17 isoforms in the gill transcriptome against a waterborne bacterial infection **(A)** and in the hemocyte transcriptome against a bacterial infection injected into the muscle **(B)**. It is shown that the isoforms expressed in each transcriptome changed completely. Colored pie charts represent the diversity of expressed IL-17 isoforms in each individual mussel from the control (C1, C2, C3) and infected conditions (I1, I2, I3). The white pie charts below represent, for each experiment, the weight that the expression values of each of the upper graphs have (calculated by the sum of the expression values of every IL-17 form for each of the individual samples). Clear upregulation of IL-17 expression is seen in the gill transcriptome against infection but not in the hemocyte transcriptome. Expression of IL-17 receptors in the same gill **(C)** and hemocyte **(D)** transcriptomes. Regarding the IL-17 receptors, it can be seen that the expressed isoforms also varied between transcriptomes.

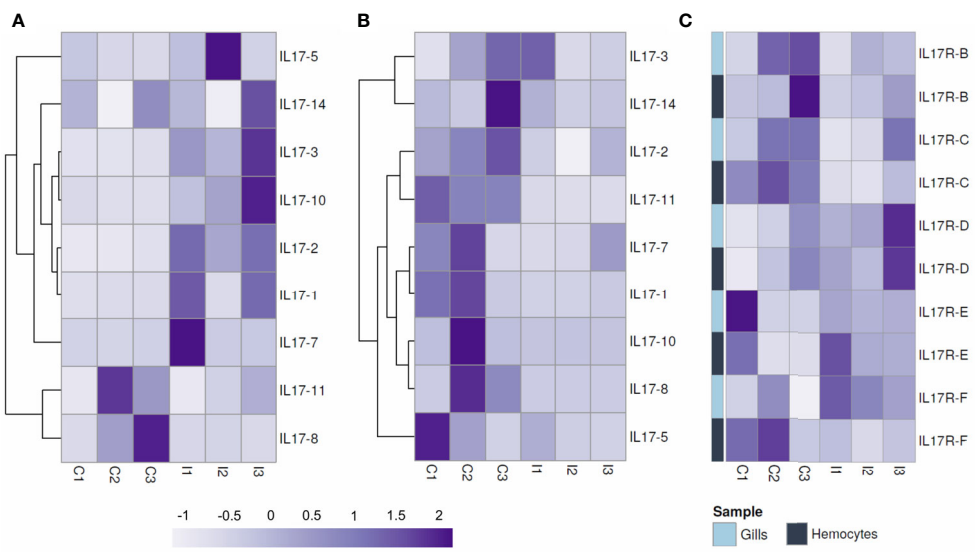


FIGURE 7 | Normalized expression by qPCR of IL-17 isoforms expressed in the gill **(A)** and in the adductor muscle hemocytes of the same individual mussels **(B)** after a 24-hour waterborne infection with *Vibrio splendidus*, as well as the expression of the receptors **(C)**. In gills, upregulation of the expressed IL-17 isoforms occurred, whereas in hemocytes, there was downregulation. Receptors did not show the antagonistic regulation between hemocytes and gills seen with the IL-17 genes, but instead, the expression of each receptor was correlated in both gill and hemocyte samples.

(PAV) recently reported in *M. galloprovincialis* (41) and that characterizes other mussel immune genes. PAV events are associated with pangenomes, which are common among prokaryotes and certain plant species, wherein a core set of indispensable genes is shared among individuals, but there are large variations in genomic content affecting "dispensable" genes (one-third of all the protein-coding genes in mussels) present or absent among individual genomes (66, 67). This phenomenon may represent a breakthrough in the explanation of mussel survival success, since it reveals an adaptive strategy that endows this species with vast levels of variability in key gene families, mainly those related to immunity. In the case of IL-17, unique variants of each isoform were caused by the interindividual variability in specific positions, but every isoform was conserved among individuals. Concerning isoform definition, it was revealed by the phylogenetic analysis that isoforms IL17-3 and IL17-16 would be diverging from IL17-17, since the branch of the latter contains both isoforms. The distinction between these isoforms was maintained due to the homology criterion and the evidence of distinct transcriptomic expression of these isoforms.

Comparative genomics using the same clustering parameters with species from different phyla revealed similarly expanded IL-17 repertoires in other marine invertebrates, as well. This was observed mainly in marine bivalve mollusc species (*C. gigas* and *M. mizuhopecten*, in addition to the mussels *M. galloprovincialis* and *M. coruscus*), the cephalopod *O. bimaculoides* and the echinoderm *S. purpuratus*. Among molluscs, large IL-17 gene families have also been reported in more marine bivalves, such as *Pinctada fucata martensii* (33), while none of the gastropods studied possessed expanded repertoires. There was already previous evidence of IL-17 gene family expansions in octopi (32) and the purple sea urchin (30, 31), which we confirmed in this work.

In these marine invertebrate species, expansions are not restricted to IL-17, as these species are characterized to possess unusual large repertoires of other innate immune genes. This has been extensively reported in marine bivalves with C1q genes, pattern-recognition receptors, TLRs or TNF homologs (41, 68–73), in octopi with protocadherins, among others (32), and in the purple sea urchin with Sp185/333 proteins and TLRs, among other innate immune receptors (30, 74, 75). Expansion phenomena and diversification in innate immune gene families are usually associated with a higher degree of specificity, which endows certain invertebrates with more efficient immune responses. This has also been revealed on several occasions as a cause of the acquisition of innate memory capacity against certain pathogens (76).

The origin of these expanded IL-17 repertoires is not clear. Even if this proinflammatory cytokine is conserved across a wide range of species, the characteristics of the IL-17 family strongly differ among them, and there are no homologous forms that allow tracing throughout evolution. The fact that these expanded repertoires appeared in species from different phyla particularly enriched with highly variable immune genes points to an independent expansion in each lineage. This phenomenon would be supported by the observed tandem duplications of IL-17 genes in these species. The expanded repertoires are conserved within a specific lineage, and homologues are only found in the closest species, as observed between some Mytilidae and strongylocentroid echinoderm species (31). Nonmarine species and gastropods (nonfilter feeder molluscs) lacked these expanded repertoires, which could be related in a certain way to specific species in marine environments and filtering activity, both factors intrinsically linked to an increased exposure to pathogens. The lack of expanded repertoires happened as well in echinoderms as *Asterias rubens*, which is less closely related to *S. purpuratus*, supporting the discussed hypothesis regarding the independence of IL-17 expansions in specific species.

Nematoda had been proposed as the ancestral phylum of origin of IL-17 genes (28), but in the current work, we were able to identify IL-17 families in cnidarians, which emerged 600 million years ago. In the Pfam database, an IL-17 domain sequence was deposited from the poriferan *Amphimedon queenslandica* (ID: A0A1X7 VXB1_AMPQE/359-442), but this interleukin could not be identified in the sponge genome in our work or in previous approaches (28).

IL-17 is typically absent from some invertebrates, including insects and hemichordates (28); however, we found a rich IL-17 family in the hemichordate *S. kowalevskii.* Concerning insects, even if they are phylogenetically related to molluscs as protostomes, there are reports that marine molluscs and annelids are more closely related in terms of genomic organization, gene structure and functional content to some invertebrate deuterostomes, such as the sea urchin (77), which was also evidenced in this genomic study with IL-17 similarities. In the current work, we found three proteins in the *D. melanogaster* genome encoding the IL-17 domain. The three proteins had an amino acidic identity of 100% and differed only in length, therefore the more complete one was selected (NCBI ID: NP_001285550.1). These three proteins had been automatically annotated as prothoracicotropic hormones E, F and G but they share only 23.38% identity with the prothoracicotropic hormone from *D. melanogaster*, which did not contain the IL-17 domain (CAA66841.1). The putative *D. melanogaster* IL-17 should be further analyzed in order to verify its identity.

The study of IL-17 domains highlighted some evolutionary patterns. Invertebrate isoforms contain 6-8 conserved cysteine residues that conform to at least 3 disulphide bridges (28), while vertebrates and chordate invertebrates IL-17s are characterized by 2 bridges. These two missing cysteines have been replaced by serine residues. Motif analysis revealed a conserved motif CPW/SPW (with sporadic variations in the third amino acid) that appeared repeatedly in invertebrate and chordate sequences respectively.

Chordata species, both invertebrates and vertebrates, showed similar repertoires. Only in the Atlantic salmon did we find some additional genes clustered in a family of isoforms similar to humans and quite conserved in other teleosts (19, 21, 23). Expansion in certain marine invertebrates may not be necessary since specificity is guaranteed by their adaptive immune system. Considering that Chordata species contain a

reduced number of IL-17 isoforms that are described to possess functional diversification, it seems reasonable that the same would happen in expanded repertoires as mussels.

Analyses of transcriptomic data indicated the importance of this sequence variability. Expressed isoforms varied among tissues, individuals and stimuli. There were also IL-17 isoforms that lacked the signal peptide, as reported previously in other invertebrate IL-17 repertoires (28). The reason behind this could be a misannotation that does not identify the exon coding for the signal peptide. However, this could be also related to putative functional diversification. In fact, variable exon-intron structures were reported for mussel IL-17 genes, ranging from 1 to 7 exons, although the 2 exon structure was the most repeated one. This is not rare, since invertebrate IL-17 genes are characterized by more variable exon-intron structures than vertebrates (28, 33).

Specific isoforms of the expanded *S. purpuratus* IL-17 repertoire were implicated in the gut epithelium immune response after bacterial infection in sea urchin larvae, activating the modulation of several other innate immunity genes. Another isoform was induced in adult immune cells, representing evidence of functional diversification of the repertoire (31). In mussels, an upregulation of IL-17 in the gills in response to a waterborne infection from the surrounding water has been reported (48). *In situ* hybridization with the most highly expressed isoform (IL17-3) suggested that these interleukins may be expressed in hemocytes and in the epithelial cells of gills. IL-17 isoforms responsive to waterborne bacterial infection could therefore exert specific immune functions in gills, which is in accordance with the role of IL-17 in other species, where they are also implicated in epithelial immunity (31, 78).

IL-17 in gill cells triggers the immune response when coupled to IL-17 receptors, triggering NF-kB signaling and activating antimicrobial peptide expression in molluscs, among other immune genes (37, 79–81). This suggests an ancient role of IL-17 in the mucosal immune responses of gills from marine animals. Indeed, the continuous filtering activity of bivalves could lead to repetitive contact with putative pathogens, for which they would need a strong arsenal of immune molecules and clear recognition of nonself-elements.

The current work indicates the importance of comparative immunology and genomics in studying elements of the innate immune response conserved throughout evolution. IL-17 gene families were described for the first time in several invertebrate species, and Cnidaria was found to be the most ancient diverged phylum. More diverse repertoires were observed in marine animals, which could be related to their environment rich in pathogens. Large expansions of IL-17 gene families were observed in certain phyla of marine invertebrates, revealing rich repertoires with functional diversification. Mussels are species characterized by presence/absence variation in key gene families, but even if they present high sequence variability, IL-17 isoforms are conserved among individuals. This could point to the great importance of these cytokines in the immune response of mussels and other marine invertebrates. We demonstrated the implications of these cytokines in mussel gills, including taking part in immune signaling after the recognition of an infection from the environment. The involvement of Il-17 in mucosal immune signaling seems to be conserved from ancestral lineages to humans.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The Mediterranean mussel, *M. galloprovincialis*, is not considered an endangered or protected species in any international species catalog, including the CITES list (www.cites.org), and it is not included in the list of species regulated by the EC Directive 2010/63/EU. Therefore, no specific authorization is required to work on mussel samples.

AUTHOR CONTRIBUTIONS

BN and AF conceived and designed the project. AS, MR-C, UR and AF performed the bioinformatic analyses. AS and MR-C performed the functional assays. AS wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Spanish AEI/EU-FEDER (RTI2018-095997-B-I00); the European Regional Development Fund (ERDF) Interreg V Spain—Portugal (0474_BLUEBIOLAB); and the Consellería de Economía, Emprego e Industria - GAIN, Xunta de Galicia, project IN607B 2019/01 and EU H2020, project VIVALDI (678589). SA was supported by a Spanish AEI/EU-FSE predoctoral contract PRE2019-090760. MR-C was supported by a Spanish AEI/EU-FEDER predoctoral contract BES-2016-076302.

ACKNOWLEDGMENTS

We want to thank Judit Castro and the aquarium staff at our institution for their technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.692997/full#supplementary-material

REFERENCES

1. Gerdol M, Venier P. An Updated Molecular Basis for Mussel Immunity. *Fish Shellfish Immunol* (2015) 46:17–38. doi: 10.1016/j.fsi.2015.02.013
2. Ottaviani E, Franchini A, Cassanelli S, Genedani S. Cytokines and Invertebrate Immune Responses. *Biol Cell* (1995) 85:87–91. doi: 10.1016/0248-4900(96)89130-1
3. Malagoli D. Cytokine Network in Invertebrates: The Very Next Phase of Comparative Immunology. *Invertebr Surviv J* (2010) 7:146–8.
4. Saenz SA, Taylor BC, Artis D. Welcome to the Neighborhood: Epithelial Cell-Derived Cytokines License Innate and Adaptive Immune Responses at Mucosal Sites. *Immunol Rev* (2008) 226:172–90. doi: 10.1111/j.1600-065X.2008.00713.x
5. Turner MD, Nedjai B, Hurst T, Pennington DJ. Cytokines and Chemokines: At the Crossroads of Cell Signalling and Inflammatory Disease. *Biochim Biophys Acta - Mol Cell Res* (2014) 1843:2563–82. doi: 10.1016/j.bbamcr.2014.05.014
6. Rouvier E, Luciani MF, Mattéi MG, Denizot F, Golstein P. CTLA-8, Cloned From an Activated T Cell, Bearing AU-Rich Messenger RNA Instability Sequences, and Homologous to a Herpesvirus Saimiri Gene. *J Immunol* (1993) 150:5445–56.
7. Yao Z, Painter SL, Fanslow WC, Ulrich D, Macduff BM, Spriggs MK, et al. Human IL-17: A Novel Cytokine Derived From T Cells. *J Immunol* (1995) 155:5483–6.
8. Zhang X, Angkasekwinai P, Dong C, Tang H. Structure and Function of Interleukin-17 Family Cytokines. *Protein Cell* (2011) 2:26–40. doi: 10.1007/s13238-011-1006-5
9. Cua DJ, Tato CM. Innate IL-17-Producing Cells: The Sentinels of the Immune System. *Nat Rev Immunol* (2010) 10:479–89. doi: 10.1038/nri2800
10. Benderdour M, Tardif G, Pelletier JP, Di Battista JA, Reboul P, Ranger P, et al. Interleukin 17 (IL-17) Induces Collagenase-3 Production in Human Osteoarthritic Chondrocytes *via* AP-1 Dependent Activation: Differential Activation of AP-1 Members by IL-17 and IL-1β. *J Rheumatol* (2002) 29:1262–72.
11. Hata K, Andoh A, Shimada M, Fujino S, Bamba S, Araki Y, et al. IL-17 Stimulates Inflammatory Responses *via* NF-κb and MAP Kinase Pathways in Human Colonic Myofibroblasts. *Am J Physiol - Gastrointest Liver Physiol* (2002) 282:1035–44. doi: 10.1152/ajpgi.00494.2001
12. Gaffen SL. Biology of Recently Discovered Cytokines: Interleukin-17 - A Unique Inflammatory Cytokine With Roles in Bone Biology and Arthritis. *Arthritis Res Ther* (2004) 6:240–7. doi: 10.1186/ar1444
13. Gaffen SL. Structure and Signalling in the IL-17 Receptor Family. *Nat Rev Immunol* (2009) 9:556–67. doi: 10.1038/nri2586
14. Kolls JK, Lindén A. Interleukin-17 Family Members and Inflammation. *Immunity* (2004) 21:467–76. doi: 10.1016/j.immuni.2004.08.018
15. Starnes T, Broxmeyer HE, Robertson MJ, Hromas R. Cutting Edge: IL-17D, a Novel Member of the IL-17 Family, Stimulates Cytokine Production and Inhibits Hemopoiesis. *J Immunol* (2002) 169:642–6. doi: 10.4049/jimmunol.169.2.642
16. Li H, Chen J, Huang A, Stinson J, Heldens S, Foster J, et al. Cloning and Characterization of IL-17B and IL-17C, Two New Members of the IL-17 Cytokine Family. *Proc Natl Acad Sci USA* (2000) 97:773–8. doi: 10.1073/pnas.97.2.773
17. Hymowitz SG, Filvaroff EH, Yin JP, Lee J, Cai L, Risser P, et al. IL-17s Adopt a Cystine Knot Fold: Structure and Activity of a Novel Cytokine, IL-17F, and Implications for Receptor Binding. *EMBO J* (2001) 20:5332–41. doi: 10.1093/emboj/20.19.5332
18. Secombes CJ, Wang T, Bird S. The Interleukins of Fish. *Dev Comp Immunol* (2011) 35:1336–45. doi: 10.1016/j.dci.2011.05.001
19. Gunimaladevi I, Savan R, Sakai M. Identification, Cloning and Characterization of Interleukin-17 and its Family From Zebrafish. *Fish Shellfish Immunol* (2006) 21:393–403. doi: 10.1016/j.fsi.2006.01.004
20. Kono T, Korenaga H, Sakai M. Genomics of Fish IL-17 Ligand and Receptors: A Review. *Fish Shellfish Immunol* (2011) 31:635–43. doi: 10.1016/j.fsi.2010.11.028
21. Korenaga H, Kono T, Sakai M. Isolation of Seven IL-17 Family Genes From the Japanese Pufferfish *Takifugu Rubripes*. *Fish Shellfish Immunol* (2010) 28:809–18. doi: 10.1016/j.fsi.2010.01.016
22. Dong C, Kong S, Zheng X, Zhang J, Nie G, Li X, et al. Genome-Wide Identification of Interleukin-17 (IL17) in Common Carp (*Cyprinus Carpio*) and its Expression Following *Aeromonas Hydrophila* Infection. *Gene* (2019) 686:68–75. doi: 10.1016/j.gene.2018.10.038
23. González-Fernández C, Chaves-Pozo E, Cuesta A. Identification and Regulation of Interleukin-17 (IL-17) Family Ligands in the Teleost Fish European Sea Bass. *Int J Mol Sci* (2020) 21:1–15. doi: 10.3390/ijms21072439
24. Han Q, Das S, Hirano M, Holland SJ, McCurley N, Guo P, et al. Characterization of Lamprey IL-17 Family Members and Their Receptors. *J Immunol* (2015) 195:5440–51. doi: 10.4049/jimmunol.1500892
25. Monte MM, Wang T, Holland JW, Zou J, Secombes CJ. Cloning and Characterization of Rainbow Trout Interleukin-17A/F2 (IL-17a/F2) and IL-17 Receptor A: Expression During Infection and Bioactivity of Recombinant IL-17a/F2. *Infect Immun* (2013) 81:340–53. doi: 10.1128/IAI.00599-12
26. Du L, Qin L, Wang X, Zhang A, Wei H, Zhou H. Characterization of Grass Carp (*Ctenopharyngodon Idella*) IL-17D: Molecular Cloning, Functional Implication and Signal Transduction. *Dev Comp Immunol* (2014) 42:220–8. doi: 10.1016/j.dci.2013.09.015
27. Du L, Feng S, Yin L, Wang X, Zhang A, Yang K, et al. Identification and Functional Characterization of Grass Carp IL-17a/F1: An Evaluation of the Immunoregulatory Role of Teleost IL-17a/F1. *Dev Comp Immunol* (2015) 51:202–11. doi: 10.1016/j.dci.2015.03.014
28. De Huang X, Zhang H, He MX. Comparative and Evolutionary Analysis of the Interleukin 17 Gene Family in Invertebrates. *PloS One* (2015) 10:1–15. doi: 10.1371/journal.pone.0132802
29. Vizzini A, Di Falco F, Parrinello D, Sanfratello MA, Mazzarella C, Parrinello N, et al. C*iona Intestinalis* Interleukin 17-Like Genes Expression is Upregulated by LPS Challenge. *Dev Comp Immunol* (2015) 48:129–37. doi: 10.1016/j.dci.2014.09.014
30. Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, et al. The Immune Gene Repertoire Encoded in the Purple Sea Urchin Genome. *Dev Biol* (2006) 300:349–65. doi: 10.1016/j.ydbio.2006.08.065
31. Buckley KM, Ho ECH, Hibino T, Schrankel CS, Schuh NW, Wang G, et al. IL17 Factors are Early Regulators in the Gut Epithelium During Inflammatory Response to Vibrio in the Sea Urchin Larva. *Elife* (2017) 6:1–24. doi: 10.7554/eLife.23481
32. Albertin CB, Simakov O, Mitros T, Wang ZY, Pungor JR, Edsinger-Gonzales E, et al. The Octopus Genome and the Evolution of Cephalopod Neural and Morphological Novelties. *Nature* (2015) 524:220–4. doi: 10.1038/nature14668
33. Cao Y, Yang S, Feng C, Zhan W, Zheng Z, Wang Q, et al. Evolution and Function Analysis of Interleukin-17 Gene From *Pinctada Fucata* Martensii. *Fish Shellfish Immunol* (2019) 88:102–10. doi: 10.1016/j.fsi.2019.02.044
34. Li J, Zhang Y, Zhang Y, Xiang Z, Tong Y, Qu F, et al. Genomic Characterization and Expression Analysis of Five Novel IL-17 Genes in the Pacific Oyster, *Crassostrea Gigas*. *Fish Shellfish Immunol* (2014) 40:455–65. doi: 10.1016/j.fsi.2014.07.026
35. Rosani U, Varotto L, Gerdol M, Pallavicini A, Venier P. IL-17 Signaling Components in Bivalves: Comparative Sequence Analysis and Involvement in the Immune Responses. *Dev Comp Immunol* (2015) 52:255–68. doi: 10.1016/j.dci.2015.05.001
36. Roberts S, Gueguen Y, de Lorgeril J, Goetz F. Rapid Accumulation of an Interleukin 17 Homolog Transcript in *Crassostrea Gigas* Hemocytes Following Bacterial Exposure. *Dev Comp Immunol* (2008) 32:1099–104. doi: 10.1016/j.dci.2008.02.006
37. Wu SZ, De Huang X, Li Q, He MX. Interleukin-17 in Pearl Oyster (*Pinctada Fucata*): Molecular Cloning and Functional Characterization. *Fish Shellfish Immunol* (2013) 34:1050–6. doi: 10.1016/j.fsi.2013.01.005
38. Romero A, Del Mar Costa M, Forn-Cuni G, Balseiro P, Chamorro R, Dios S, et al. Occurrence, Seasonality and Infectivity of Vibrio Strains in Natural Populations of Mussels *Mytilus Galloprovincialis*. *Dis Aquat Organ* (2014) 108:149–63. doi: 10.3354/dao02701
39. Benabdelmouna A, Garcia C, Ledu C, Lamy P, Maurouard E, Dégremont L. Mortality Investigation of Mytilus Edulis and Mytilus Galloprovincialis in France: An Experimental Survey Under Laboratory Conditions. *Aquaculture* (2018) 495:831–41. doi: 10.1016/j.aquaculture.2018.06.075
40. Domeneghetti S, Varotto L, Civettini M, Rosani U, Stauder M, Pretto T, et al. Mortality Occurrence and Pathogen Detection in *Crassostrea Gigas* and *Mytilus Galloprovincialis* Close-Growing in Shallow Waters (Goro Lagoon, Italy). *Fish Shellfish Immunol* (2014) 41:37–44. doi: 10.1016/j.fsi.2014.05.023
41. Gerdol M, Moreira R, Cruz F, Gómez-Garrido J, Vlasova A, Rosani U, et al. Massive Gene Presence-Absence Variation Shapes an Open Pan-Genome in the Mediterranean Mussel. *Genome Biol* (2020) 21:275. doi: 10.1186/s13059-020-02180-3
42. Zhang Y, Li J, Yu F, He X, Yu Z. Allograft Inflammatory Factor-1 Stimulates Hemocyte Immune Activation by Enhancing Phagocytosis and Expression of Inflammatory Cytokines in *Crassostrea Gigas*. *Fish Shellfish Immunol* (2013) 34:1071–7. doi: 10.1016/j.fsi.2013.01.014
43. Gerdol M, Gomez-Chiarri M, Castillo MG, Figueras A, Fiorito G, Moreira R, et al. Immunity in Molluscs: Recognition and Effector Mechanisms, With a

Focus on Bivalvia. In: EL Cooper, editor. *Advances in Comparative Immunology*, vol. . p . Cham, Switzerland: Springer International Publishing (2018). p. 225–341. doi: 10.1007/978-3-319-76768-0_11
44. Li Y, Jiang S, Li M, Xin L, Wang L, Wang H, et al. A Cytokine-Like Factor Astakine Accelerates the Hemocyte Production in Pacific Oyster *Crassostrea Gigas*. *Dev Comp Immunol* (2016) 55:179–87. doi: 10.1016/j.dci.2015.10.025
45. Zhang L, Zhao J, Li C, Su X, Chen A, Li T, et al. Cloning and Characterization of Allograft Inflammatory Factor-1 (AIF-1) From Manila Clam *Venerupis Philippinarum*. *Fish Shellfish Immunol* (2011) 30:148–53. doi: 10.1016/j.fsi.2010.09.021
46. Balseiro P, Falcó A, Romero A, Dios S, Martínez-López A, Figueras A, et al. *Mytilus Galloprovincialis* Myticin C: A Chemotactic Molecule With Antiviral Activity and Immunoregulatory Properties. *PloS One* (2011) 6:1–14. doi: 10.1371/journal.pone.0023140
47. Rey-Campos M, Novoa B, Pallavicini A, Gerdol M, Figueras A. Comparative Genomics Reveals a Significant Sequence Variability of Myticin Genes in Mytilus Galloprovincialis. *Biomol* (2020) 10:6. doi: 10.3390/biom10060943
48. Saco A, Rey-Campos M, Novoa B, Figueras A. Transcriptomic Response of Mussel Gills After a Vibrio Splendidus Infection Demonstrates Their Role in the Immune Response. *Front Immunol* (2020) 11:615580. doi: 10.3389/fimmu.2020.615580
49. Rey-Campos M, Moreira R, Valenzuela-Muñoz V, Gallardo-Escárate C, Novoa B, Figueras A. High Individual Variability in the Transcriptomic Response of Mediterranean Mussels to Vibrio Reveals the Involvement of Myticins in Tissue Injury. *Sci Rep* (2019) 9:3569. doi: 10.1038/s41598-019-39870-3
50. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for Clustering the Next-Generation Sequencing Data. *Bioinformatics* (2012) 28:3150–2. doi: 10.1093/bioinformatics/bts565
51. Li W, Godzik A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* (2006) 22:1658–9. doi: 10.1093/bioinformatics/btl158
52. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: A Web Server for Clustering and Comparing Biological Sequences. *Bioinformatics* (2010) 26:680–2. doi: 10.1093/bioinformatics/btq003
53. Fu Y-X. Linear Invariants Under Jukes' and Cantor's One-Parameter Model. *J Theor Biol* (1995) 173:339–52. doi: 10.1006/jtbi.1995.0067
54. Yang Z. Estimating the Pattern of Nucleotide Substitution. *J Mol Evol* (1994) 39:105–11. doi: 10.1007/BF00178256
55. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian Inference of Phylogenetic Trees. *Bioinformatics* (2001) 17:754–5. doi: 10.1093/bioinformatics/17.8.754
56. Rambaut A. *FigTree. Tree Figure Drawing Tool* (2009). Available at: http://tree.bio.ed.ac.uk/software/figtree/.
57. Letunic I, Bork P. Interactive Tree Of Life (iTOL) V4: Recent Updates and New Developments. *Nucleic Acids Res* (2019) 47:256–9. doi: 10.1093/nar/gkz239
58. Durbin R, Eddy SR, Krogh A, Mitchison G. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press (1998). 356 p. doi: 10.1017/CBO9780511790492
59. Sonnhammer EL, von Heijne G, Krogh A. A Hidden Markov Model for Predicting Transmembrane Helices in Protein Sequences. *Proc Int Conf Intell Syst Mol Biol* (1998) 6:175–82.
60. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting Transmembrane Protein Topology With a Hidden Markov Model: Application to Complete Genomes. *J Mol Biol* (2001) 305:567–80. doi: 10.1006/jmbi.2000.4315
61. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of Prokaryotic and Eukaryotic Signal Peptides and Prediction of Their Cleavage Sites. *Protein Eng* (1997) 10:1–6. doi: 10.1093/protein/10.1.1
62. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 Improves Signal Peptide Predictions Using Deep Neural Networks. *Nat Biotechnol* (2019) 37:420–3. doi: 10.1038/s41587-019-0036-z
63. Mello B. Estimating TimeTrees With MEGA and the TimeTree Resource. *Mol Biol Evol* (2018) 35:2334–42. doi: 10.1093/molbev/msy133
64. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME Suite: Tools for Motif Discovery and Searching. *Nucleic Acids Res* (2009) 37:202–8. doi: 10.1093/nar/gkp335
65. Pfaffl MW. A New Mathematical Model for Relative Quantification in Real-Time RT–PCR. *Nucleic Acids Res* (2001) 29:e45. doi: 10.1093/nar/29.9.e45
66. McInerney JO, McNally A, O'Connell MJ. Why Prokaryotes Have Pangenomes. *Nat Microbiol* (2017) 2:1–5. doi: 10.1038/nmicrobiol.2017.40
67. Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, et al. Insights Into the Maize Pan-Genome and Pan-Transcriptome. *Plant Cell* (2014) 26:121–35. doi: 10.1105/tpc.113.119982
68. Gao D, Qiu L, Gao Q, Hou Z, Wang L, Song L. Repertoire and Evolution of TNF Superfamily in *Crassostrea Gigas*: Implications for Expansion and Diversification of This Superfamily in Mollusca. *Dev Comp Immunol* (2015) 51:251–60. doi: 10.1016/j.dci.2015.04.006
69. Wang L, Zhang H, Wang M, Zhou Z, Wang W, Liu R, et al. The Transcriptomic Expression of Pattern Recognition Receptors: Insight Into Molecular Recognition of Various Invading Pathogens in Oyster *Crassostrea Gigas*. *Dev Comp Immunol* (2019) 91:1–7. doi: 10.1016/j.dci.2018.09.021
70. Gerdol M, Venier P, Pallavicini A. The Genome of the Pacific Oyster *Crassostrea Gigas* Brings New Insights on the Massive Expansion of the C1q Gene Family in Bivalvia. *Dev Comp Immunol* (2015) 49:59–71. doi: 10.1016/j.dci.2014.11.007
71. Gerdol M, Manfrin C, De Moro G, Figueras A, Novoa B, Venier P, et al. The C1q Domain Containing Proteins of the Mediterranean Mussel *Mytilus Galloprovincialis*: A Widespread and Diverse Family of Immune-Related Molecules. *Dev Comp Immunol* (2011) 35:635–43. doi: 10.1016/j.dci.2011.01.018
72. Romero A, Dios S, Poisa-Beiro L, Costa MM, Posada D, Figueras A, et al. Individual Sequence Variability and Functional Activities of Fibrinogen-Related Proteins (FREPs) in the Mediterranean Mussel (*Mytilus Galloprovincialis*) Suggest Ancient and Complex Immune Recognition Models in Invertebrates. *Dev Comp Immunol* (2011) 35:334–44. doi: 10.1016/j.dci.2010.10.007
73. Toubiana M, Rosani U, Giambelluca S, Cammarata M, Gerdol M, Pallavicini A, et al. Toll Signal Transduction Pathway in Bivalves: Complete Cds of Intermediate Elements and Related Gene Transcription Levels in Hemocytes of Immune Stimulated. *Mytilus galloprovincialis*. *Dev Comp Immunol* (2014) 45:300–12. doi: 10.1016/j.dci.2014.03.021
74. Ghosh J, Buckley KM, Nair SV, Raftos DA, Miller C, Majeske AJ, et al. Sp185/333: A Novel Family of Genes and Proteins Involved in the Purple Sea Urchin Immune Response. *Dev Comp Immunol* (2010) 34:235–45. doi: 10.1016/j.dci.2009.10.008
75. Nair SV, Del Valle H, Gross PS, Terwilliger DP, Smith LC. Macroarray Analysis of Coelomocyte Gene Expression in Response to LPS in the Sea Urchin. Identification of Unexpected Immune Diversity in an Invertebrate. *Physiol Genomics* (2005) 22:33–47. doi: 10.1152/physiolgenomics.00052.2005
76. Milutinović B, Kurtz J. Immune Memory in Invertebrates. *Semin Immunol* (2016) 28:328–42. doi: 10.1016/j.smim.2016.05.004
77. Simakov O, Marletaz F, Cho SJ, Edsinger-Gonzales E, Havlak P, Hellsten U, et al. Insights Into Bilaterian Evolution From Three Spiralian Genomes. *Nature* (2013) 493:526–31. doi: 10.1038/nature11696
78. Pappu R, Rutz S, Ouyang W. Regulation of Epithelial Immunity by IL-17 Family Cytokines. *Trends Immunol* (2012) 33:343–9. doi: 10.1016/j.it.2012.02.008
79. Xin L, Zhang H, Zhang R, Li H, Wang W, Wang L, et al. CgIL17-5, an Ancient Inflammatory Cytokine in Crassostrea Gigas Exhibiting the Heterogeneity Functions Compared With Vertebrate Interleukin17 Molecules. *Dev Comp Immunol* (2015) 53:339–48. doi: 10.1016/j.dci.2015.08.002
80. Oyanedel D, Gonzalez R, Flores-Herrera P, Brokordt K, Rosa RD, Mercado L, et al. Molecular Characterization of an Inhibitor of NF-κb in the Scallop *Argopecten Purpuratus*: First Insights Into its Role on Antimicrobial Peptide Regulation in a Mollusk. *Fish Shellfish Immunol* (2016) 52:85–93. doi: 10.1016/j.fsi.2016.03.021
81. Xin L, Zhang H, Du X, Li Y, Li M, Wang L, et al. The Systematic Regulation of Oyster CgIL17-1 and CgIL17-5 in Response to Air Exposure. *Dev Comp Immunol* (2016) 63:144–55. doi: 10.1016/j.dci.2016.06.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be considered as a conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 30 July 2021
doi: 10.3389/fimmu.2021.715245

Two Amphioxus ApeC-Containing Proteins Bind to Microbes and Inhibit the TRAF6 Pathway

Jin Li[1,2,3], Yuhui Li[1,2,3], Zhaoyu Fan[1], Shenghui Chen[1], Xinyu Yan[1], Zirui Yue[1], Guangrui Huang[4], Shumin Liu[1], Hao Zhang[1], Shangwu Chen[1], Meiling Dong[1], Anlong Xu[1,4] and Shengfeng Huang[1,2,3]**

[1] Guangdong Province Key Laboratory of Pharmaceutical Functional Genes, State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, [2] Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai, China, [3] Laboratory for Marine Biology and Biotechnology, Qingdao National Laboratory for Marine Science and Technology, Qingdao, China, [4] School of Life Sciences, Beijing University of Chinese Medicine, Beijing, China

OPEN ACCESS

Edited by:
Ioannis Eleftherianos, George Washington University, United States

Reviewed by:
Jianmin Ye, South China Normal University, China
Yong-hua Hu, Chinese Academy of Tropical Agricultural Sciences, China

***Correspondence:**
Anlong Xu
lssxal@mail.sysu.edu.cn
Shengfeng Huang
hshengf2@mail.sysu.edu.cn

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *26 May 2021*
Accepted: *12 July 2021*
Published: *30 July 2021*

Citation:
Li J, Li Y, Fan Z, Chen S, Yan X, Yue Z, Huang G, Liu S, Zhang H, Chen S, Dong M, Xu A and Huang S (2021) Two Amphioxus ApeC-Containing Proteins Bind to Microbes and Inhibit the TRAF6 Pathway. Front. Immunol. 12:715245. doi: 10.3389/fimmu.2021.715245

The apextrin C-terminal (ApeC) domain is a class of newly discovered protein domains with an origin dating back to prokaryotes. ApeC-containing proteins (ACPs) have been found in various marine and aquatic invertebrates, but their functions and the underlying mechanisms are largely unknown. Early studies suggested that amphioxus ACP1 and ACP2 bind to bacterial cell walls and have a role in immunity. Here we identified another two amphioxus ACPs (ACP3 and ACP5), which belong to the same phylogenetic clade with ACP1/2, but show distinct expression patterns and sequence divergence (40-50% sequence identities). Both ACP3 and ACP5 were mainly expressed in the intestine and hepatic cecum, and could be up-regulated after bacterial challenge. Both prokaryotic-expressed recombinant ACP3 and ACP5 could bind with several species of bacteria and yeasts, showing agglutinating activity but no microbicidal activity. ELISA assays suggested that their ApeC domains could interact with peptidoglycan (PGN), but not with lipoteichoic acid (LTA), lipopolysaccharides (LPS) and zymosan A. Furthermore, they can only bind to Lys-type PGN from *Staphylococcus aureus*, but not to DAP-type PGN from *Bacillus subtilis* and not to moieties of PGN such as MDPs, NAMs and NAGs. This recognition spectrum is different from that of ACP1/2. We also found that when expressed in mammalian cells, ACP3 could interact with TRAF6 *via* a conserved non-ApeC region, which inhibited the ubiquitination of TRAF6 and hence suppressed downstream NF-κB activation. This work helped define a novel subfamily of ACPs, which have conserved structures, and have related yet diversified molecular functions. Its members have dual roles, with ApeC as a lectin and a conserved unknown region as a signal transduction regulator. These findings expand our understanding of the ACP functions and may guide future research on the role of ACPs in different animal clades.

Keywords: ApeC, ACP, microbial binding, PGN, NF-κB, TRAF6, Amphioxus (*Branchiostoma floridae*)

INTRODUCTION

The apextrin C-terminal (ApeC) is a new class of protein domains, characterized by a sequence of about 200 amino acid residues, comprising eight conserved Cysteine residues and three relatively conserved DXED motifs (1). Collectively, here we refer to the proteins containing ApeC domains as ACP (ApeC-containing protein) (2). Comparative genomic analyses show that ACPs exhibit diverse architectures and are widely distributed in many invertebrates. Most ACPs are present in aquatic invertebrates or invertebrates from moist environments, including cnidarians, mollusks, echinoderms, cephalochordates, flatworms, water bears, nematodes and annelids. However, there is no ACPs found in vertebrates and in major arthropod lineages (e.g. insects and crustaceans) except arachnids (2). Distant ApeC homologs were also found in bacteria, hence the origin of ApeC could be traced back to the prokaryotes. Despite their wide distribution, no ACP orthologs could be found between any phyla or sub-phyla, suggesting that ACPs obviously underwent rapid turnover and diversification.

The animal ApeC domain has eight conserved cysteines, while the bacterial ApeC-like domain has only four of them. In bacteria, ApeC-like seems to exist as single-domain proteins. In animals, ApeC actively participated in domain shuffling, which gave rise to many novel domain architectures. So far, more than twenty different domain architectures involving ApeC have been identified, suggesting great architectural diversity. This is reminiscent of other versatile domains like immunoglobulin (IG) and C-type lectin (CLECT), which are capable of exerting different functions in various domain architectures (3, 4).

The first ACP was discovered in the sea urchin *Heliocidaris erythrogramma* (5–7), which is a secreted protein with a MACPF-ApeC domain architecture. This protein is concentrated in the apical extracellular matrix in the columnar cells of the larval ectoderm, hence it was named apextrin. This apextrin is present in eggs in a type of secretory vesicles and this maternal pool, and after fertilization it will be gradually secreted to form the extracellular matrix. It is proposed that this apextrin is involved in apical cell adhesion and that its high level of expression may be necessary for strengthening the large *H. erythrogramma* embryo. Another two ACPs from the mussel *Mytilus galloprovincialis* may also have a role in embryogenesis (8): they were expressed rapidly after fertilization (up by thousands of times), and the expression decreased in later development but remained in a level in the adult stage. In Pacific oysters, the expression of an ACP was significantly up-regulated under hypoxic condition, suggesting that it may be involved in anti-stress responses (9). There are also reports suggest that ACPs may have immune functions in different species. In oysters, sea urchins and amphioxus, the expression of ACP genes could be significantly up-regulated after bacterial stimulation (10–12). A mussel ACP was highly expressed in gills and blood cells and could be further up-regulated after pathogen invasion (8, 13).

Recently, two ACPs (bjACP1 and bjACP2) from the amphioxus *Branchiostoma japonicum* have been functionally characterized (1). They were concentrated in gill and skin, and could be dramatically upregulated during acute antibacterial responses. Both them can aggregate bacteria by using their ApeC domain to bind with the cell wall component peptidoglycan (PGN). Further analysis indicates that they may interact with the muramyl dipeptide (MDP), the minimal bioactive motif of PGN. In addition to being a pattern-recognition protein, bjACP2 could regulate the TRAF6-NF-κB pathway when present in cytosol.

Previous studies have shown that some ApeC domains have carbohydrate binding capacity, and some ACPs are involved in embryonic development and host defense. However, the functions of most ACPs and their ApeC domains still remain unknown. Amphioxus (cephalochordate) is a marine chordate invertebrate, which represents the basal living chordate lineage and is therefore considered as an important proxy to study the evolution from invertebrates to vertebrates (14–19). Amphioxus has more than twenty ACP genes, hence it can be served as a valuable model to understand the functions and evolution of ACPs (2). In this study, we characterized another two ACPs (bfACP3 and bfACP5) from another amphioxus species (*Branchiostoma floridae*), which provides *a new line* of mechanistic evidence for the immune functions of ACPs.

MATERIALS AND METHODS

Animals and Cells

Adult amphioxus (*Branchiostoma floridae*) were obtained from GL's lab in Xiamen University (China), which derives from a stock maintained by Jr-Kai Yu originating from Tampa, Florida. The culture was maintained under previously described conditions (20), cultured in aquaria with aeration and supplied with fresh seawater (1.9%–2.9% salinity) at 20-25°C and fed with *Oocystis sp*. daily. HEK293T and Hela cell lines from ATCC, were grown in DMEM (Gibco) supplemented with 10% FCS at 37°C under 5% CO_2.

RNA Isolation and cDNA Synthesis

Total RNA was extracted using TRIzol Reagent (Invitrogen) and isopropanol was used for precipitation. RNA quality was determined by agarose gel electrophoresis and spectrophotometer. The isolated RNA was reverse-transcribed to synthesize the first strand cDNA using PrimeScript first Strand cDNA Synthesis Kit (Takara) using the oligo d(T) primer according to the manufacturer's protocol. The cDNA was stored at -80°C.

Cloning of B. floridae ACP3, ACP5, MyD88 and TRAF6

Amphioxus ACP genes were identified in the *B. floridae* genome from the Joint Genome Institute (JGI, http://genome.jgi-psf.org/Brafl1/Brafl1.home.html) database in our previous work (2). Using these sequences as baits, we blasted the transcript database of *B. floridae* (http://genome.bucm.edu.cn/lancelet/index.php) and obtained the consensus sequences of *B. floridae ACP3* (*bfACP3*, with gene ID of 102545) and *B. floridae ACP5* (*bfACP5*, with gene ID of 95439). Using human MyD88 and TRAF6 as baits, we

blasted the *B. floridae* genome from the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/genbank/) and obtained the consensus sequences of *B. floridae MyD88* (*bfMyD88*, with gene ID of 118403304) and *B. floridae TRAF6* (b*fTRAF6*, with gene ID of 118414689). To obtain complete cDNA sequences, four pairs of gene-specific primers were designed for each. The amplified fragments were cloned into the pGEX-Teasy vector (Promega) and verified by sequencing at The Beijing Genomics Institute (BGI). The primers were shown in **Table 1**. These sequences have been submitted to the NCBI database under the accession number MZ004941, MZ004942, MZ442381, and MZ442382 respectively.

Bioinformatic Analysis

The domain structure was predicted on the SMART Website (http://smart.embl-heidelberg.de). BLASTP was performed to analyze the sequence identities. The isoelectric point (pI) and m.w. were estimated on ExPASy Website (http://www.expasy.org/tools/). Multiple sequence alignments were analyzed using MEGA-X (21) by using the ClustalW algorithm and were

TABLE 1 | Primers used for PCR amplification.

Primers	Primer sequence (5′ to 3′)
Gene-specific primers	
bfACP3-F	5'-ATGTTAGCACTCAAGCTCATT-3'
bfACP3-R	5'-TCACTAGGCAGGCTGGTAGTA-3'
bfACP5-F	5'-AACAGGTGTGGAGTTAAAGGTG-3'
bfACP5-R	5'-TGTGTAGGCAGCATTACGA-3'
bfMyD88-F	5'-GCAAGAATCCAGCCTTTGATCTG-3'
bfMyD88-R	5'-AAGACTGCAACGGAGGCTAA -3'
bfTRAF6-F	5'-TCTTTCCCTGTCTTCATACTTT-3'
bfTRAF6-R	5'-ATGAAACTCAGAAGTTTTTCTGTG-3'
Primers for Q-PCR	
gapdh-F	5'-CAAGGCTGTAGGCAAGGTCAT-3'
gapdh-R	5'-CTTCTTCAGTCGGCAGGTCAG-3'
bfACP1-qF	5'-CTTCGGAAGAAACAACAT-3'
bfACP1-qR	5'-ATCTTCATCGTCCCAATA-3'
bfACP2-qF	5'-GACGATAATAGCAACCAAT-3'
bfACP2-qR	5'-GTTTCCCTTCTTAAAGATACA-3'
bfACP3-qF	5'-AGTGATGGCTCCATTTAC-3'
bfACP3-qR	5'-ATCGTTGTAGGTATTCTCAT-3'
bfACP5-qF	5'-TCCTACAGAACACCGATA-3'
bfACP5-qR	5'-CTAGCTCGTTATGGTTGA-3'
Primers for Recombinant proteins	
bfACP3-32aF	5'-gccatggctgatatcggatccGACAATTTTGAGAAAACTCCAGTGG-3'
bfACP3-32aR	5'-acggagctcgaattcggatccgcGGCAGGCTGGTAGTAGCAGTACC-3'
bfACP5-32aF	5'-gccatggctgatatcggatccGACAATGTCTGTGGCGATGATC-3'
bfACP5-32aR	5'-acggagctcgaattcggatccgcTTCATCCCGTTGGTAGAAGCA-3'
Primers for construction of expression vector	
bfACP3-HA-F	5'-actactggtacctctggatccATGTTAGCACTCAAGCTCATTGTGC-3'
bfACP3-HA-R	5'-cttaccgaattctgtggatccGGCAGGCTGGTAGTAGCAGTACC-3'
Myc-bfACP3-F	5'-tggccatggaggcccgaattcggATGTTAGCACTCAAGCTCATTGTGC-3'
Myc-bfACP3-R	5'-gatccccgcggccgcggtaccCTAGGCAGGCTGGTAGTAGCAGT-3'
Flag-bfMyD88-F	5'-gacgatgacaagggcggtaccATGGCAACAAACGCGCCA-3'
Flag-bfMyD88-R	5'-ttctgtggatccagaggtaccTCACGGGCGAGAGAGGGC-3'
Flag-bfTRAF6-F	5'-gacgatgacaagggcggtaccATGAAGCCAGGAGGGAGGG-3'
Flag-bfTRAF6-R	5'-ttctgtggatccagaggtaccCTATTGTGGCTGCACCGTACAT-3'
bfACP3-24-561-F	5'-accgagatctctcgaggtaccGACAATTTTGAGAAAACTCCAGTGG-3'
bfACP3-24-561-R	5'-gatccccgcggccgcggtaccCTAGGCAGGCTGGTAGTAGCAGT-3'
bfACP3-24-204-F	5'-accgagatctctcgaggtaccGACAATTTTGAGAAAACTCCAGTGG-3'
bfACP3-24-204-R	5'-gatccccgcggccgcggtaccCTACCCCACCTCCACTGCCG-3'
bfACP3-205-561-F	5'-accgagatctctcgaggtaccAATCAGTGGATGGAGGGCG-3'
bfACP3-205-561-R	5'-gatccccgcggccgcggtaccCTAGGCAGGCTGGTAGTAGCAGT-3'
bfACP3-205-358-F	5'-accgagatctctcgaggtaccAATCAGTGGATGGAGGGCG-3'
bfACP3-205-358-R	5'-gatccccgcggccgcggtaccCTATTTTTTCACATGTGGCTCCAG-3'
bfACP3-165-358-F	5'-accgagatctctcgaggtaccGGCACCAGCAGTGACCAACA-3'
bfACP3-165-358-R	5'-gatccccgcggccgcggtaccCTATTTTTTCACATGTGGCTCCAG-3'
bfACP3-124-358-F	5'-accgagatctctcgaggtaccAAGGAAACAGGCAACGAGAACG-3'
bfACP3-124-358-R	5'-gatccccgcggccgcggtaccCTATTTTTTCACATGTGGCTCCAG-3'
bfACP3-24-358-F	5'-accgagatctctcgaggtaccGACAATTTTGAGAAAACTCCAGTGG-3'
bfACP3-24-358-R	5'-gatccccgcggccgcggtaccCTATTTTTTCACATGTGGCTCCAG-3'
bfACP3-359-561-F	5'-accgagatctctcgaggtaccTGGCCTACTGGAACCTATGGC-3'
bfACP3-359-561-R	5'-gatccccgcggccgcggtaccCTAGGCAGGCTGGTAGTAGCAGT-3'

manually corrected using GeneDoc software (22). The Neighbor-joining phylogenetic tree was built using MEGA-X (21) with the JTT matrix-based method (23), 1000 bootstrap tests and pairwise deletion of gaps. The Maximum likelihood phylogenetic tree was built using MEGA-X with the WAG model, Gamma distribution of rates across sites model and 1000 bootstrap tests (24).

Real-Time Quantitative RT-PCR

Real-time quantitative RT-PCR (Q-PCR) was performed and analyzed as described (25). For tissue expression profiles, five healthy adults were chosen and not fed for 3 days to empty the gut. Then different tissues (muscle, skin, gill, ovary, hepatic cecum, intestine) were dissected under an optical microscope. Total RNA was extracted and reversely transcribed to the first strand cDNA according to the method in 2.2. Q-PCR was performed on the Roche LightCycler 480 instrument (using the 384-well module). SYBR® Green Realtime PCR Master Mix (Toyobo) was used for the assays according to the manufacturer's protocol. The reaction volume was 10μl, with 40ng 1st-strand cDNA and a primer concentration of 0.5 μM. The PCR program is 95 °C for 3 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 20 s. Reaction of each sample was performed in triplet. Amphioxus *gapdh* was used as the internal control. The cycle threshold values were calculated by the Roche LightCycler 480 software. Data were quantified using the $2^{-\Delta\Delta Ct}$ method based on the cycle threshold values. All results were confirmed by repeating the assays by one or two more times. Primers used for Q-PCR were listed in **Table 1**.

For monitoring the gene expression changes after immune stimulation, ten adults were collected for each treatment and not fed for 24h in filtered seawater. Each amphioxus was injected into the coelom with 20 ul LPS (1 mg/mL in PBS, from *E. coli* O111:B4, Sigma), LTA (1 mg/mL in PBS, from *S. aureus*, Sigma), inactivated bacteria or PBS (as control). The bacteria were 1:1 mixture (vol/vol) of formalin-inactivated *Staphylococcus* aureus (2×10^8 cells/mL) and *Vibrio anguillarum* (2×10^8 cells/mL). Then, the gut (hepatic cecum and intestine) and the gill were harvested at 0, 2, 4, 8, 12, 24, and 48h after injection and immediately frozen using liquid nitrogen.

Preparation of Recombinant Proteins

To obtain soluble proteins, we expressed recombinant ACP proteins using the pET32a system, a thioredoxin (TRX) fusion system containing a 6×His tag to facilitate the purification on a Ni^2-chelating Sepharose column and a partner TRX to help the proteins fold correctly. The ClonExpress®II One Step Cloning Kit (Vazyme) was used to insert the coding regions of mature bfACP3 (Asp24 to Ala561) and bfACP5(Asp21 to Glu636) into the pET32a plasmids. The plasmids were transformed into *E. coli* BL21 (DE3). The transformed bacteria were cultured to an OD=0.6-0.8 and induced with 1 mM IPTG at 37°C for four hours. After induction, the bacterial cells were collected and sonicated for lysis. The cell lysis supernatant was purified through a Ni^2-chelating Sepharose column (GE Healthcare). The recombinant proteins were eluted with 250 mM imidazole, dialyzed in PBS buffer at 4°C for 12 hours three times and concentrated by ultrafiltration using an Ultrafree centrifugal filter device (Millipore). Pierce™ BCA Protein Assay Kit (Thermo) was used to determine the protein concentration according to the manufacturer's protocol. Primers used for preparation of recombinant proteins were listed in **Table 1**.

Microbial Binding Assays

Bacteria Staphylococcus aureus, Enterococcus faecalis, Escherichia coli, Vibrio anguillarum, Vibrio parahemolyticus, Acinetobacter calcoaceticus and Saccharomyces cerevisiae were respectively inoculated into suitable liquid medium and cultivated overnight under suitable conditions. V. anguillarum was cultured with high-salt Luria-Bertani (LB) medium at 28°C, V. parahemolyticus was cultured with seawater medium at 28°C, S. cerevisiae was cultured with YPD medium at 28°C, and the remaining bacteria were cultured with LB medium at 37°C. The next day, the cells were collected and washed with PBS buffer and resuspended in PBS. Approximately 2×10^6 microbes were incubated with 1 μg of purified recombinant proteins in 1 mL PBS at 4°C overnight with gentle orbital rotation. Microbes were centrifuged at 12,000 × g for 1 min at 4°C and the pellets were washed five times with 1 mL of PBST (0.05% Tween-20 in PBS). The washed pellets were then suspended in 100μl PBS and 20μl 6×loading buffer and boiled at 100°C for 10 min. Western blot was performed with an anti-His mouse monoclonal antibody (Sigma) to validate the binding proteins.

Microbial Aggregation Assays

Microbes collected from liquid cultures were suspended in 1 mL PBS and mixed with 50μl Fluorescein isothiocyanate (FITC) (sigma, 10 mg/mL in DMSO). The reaction was incubated at room temperature in the dark for 3h with gentle agitation. Then the microbes were washed five times and resuspended in PBS. 10μg TRX fusion proteins were incubated with 50 μl FITC-labeled *S. aureus* (2×10^8 cells/mL), *E. coli* (2×10^8 cells/mL), *V. parahemolyticus* (2×10^8 cells/mL) or *S. cerevisiae* (2×10^7 cells/mL) at room temperature in the dark for 2h, respectively. The agglutinating reaction was examined immediately under fluorescence microscopy (Carl Zeiss).

Binding Assays of ACPs With Microbial Cell Wall Components

ELISA was used to analyze the binding of ACP proteins with soluble microbial cell wall components as previously described (26). In brief, 20 μg of lipopolysaccharide (LPS, from *E. coli*, Sigma), lipoteichoic acid (LTA, from *S. aureus*, Sigma), peptidoglycan (PGN, from *S. aureus* or *Bacillus subtilis*, Sigma), D-Mannose (Sigma), zymosan A (from *S. cerevisiae* Sigma), muramyl dipeptide (MDP, Sigma), GlcNAc (NAG, Sigma) and MurNAc (NAM, Sigma) were used to coat 96-well microplate (Corning 96-well Clear Polystyrene High-Bind Strip well Microplate) at 37°C for 3 h (PGN and zymosan A are ultrasonically solubilized). After washing with PBS three times, the wells were blocked with PBST (0.05% Tween 20 in PBS) containing 10% (wt/vol) skim milk overnight at 4°C. Several concentrations of ACP proteins were added to the well and incubated at 37°C for 2 h. Binding proteins were detected with mouse anti-His mAb (sigma, diluted 10000-fold) at room

temperature for 1h, followed by an hour incubation with HRP-labeled anti-mouse Secondary antibodies (Abmart, diluted 5000-fold). Incubating plate with 100ul TMB Substrate Solution (Thermo) at room temperature for 15 minutes and stop reaction by adding 50μL of 2M sulfuric acid. The absorbance was read at 450 nm immediately. Reaction of each sample was performed in triplet and the assay was repeated at least three times. The insoluble PGN-binding activities of ACP proteins were detected by pull-down assays as previously described (26). In brief, various concentrations of insoluble PGNs were incubated with 5 μg of ACP proteins in 1 mL PBS at 4°C for 1 h. The samples were centrifuged at 15,000 × g for 10 min and pellets were washed with PBST (0.05% Tween 20 in PBS) four times, suspended in 100 μl PBS and 20 μl 6×loading buffer, and then boiled at 100°C for 10 min. Western blot analysis of the binding proteins was performed using mouse anti-His mAb (sigma).

Antimicrobial Activity Assays

The growth curves of *Staphylococcus aureus* and *Vibrio parahemolyticus* cultured with recombinant ACP proteins were tested as follows. Two single colonies were picked up separately and transferred into 1 mL of LB or Sea water broth. A volume of 50 μl of cell suspension was mixed with purified recombinant proteins and added to 1 mL broth. *S. aureus* was incubated at 200 rpm at 37°C and *V. parahemolyticus* was incubated at 200 rpm at 28°C. OD_{600} of each sample was measured every 1 h. The Oxford cup method was performed on a petri dish. Twenty milliliters of warm nutrient agar (1.5%) were poured into a 90-mm plate and cooled to form the base medium. Ten milliliters of 0.8% warm nutrient agar mixed with *S. aureus* or *V. parahemolyticus* were poured onto the base medium and then quickly placed the Oxford cup and gently press. Then, the targeted protein or antibiotic in 100 μL of PBS was added to the pores at a final concentration of 1 μg/μL. The plates were incubated at a suitable temperature for 16h or 40 h. A transparent ring around the pores indicated antibacterial activity.

Construction of the Expression Vectors

For the study of the subcellular localization and coimmunoprecipitation (Co-IP) test between bfACP3 and bfTRAF6, full-length bfACP3 was inserted into the pcDNA3.0 vector (Clontech) with a C-terminal HA Tag (transformed by our laboratory, the HA coding sequence was inserted after the *XbaI* restriction site) and bfTRAF6 was fused with Flag tag and inserted into the pcDNA 3.0 vector (Clontech, transformed by our laboratory, the Flag coding sequence was inserted in front of the *Kpn I* restriction site). For the expression of the truncated mutants of bfACP3, PCR fragments encoding amino acids 24-561, 24-204, 205-561, 205-358, 165-358, 124-358, 24-358 and 359-561 were fused with myc tag, and inserted into the expression plasmid pCMV-Myc vector (Clontech). For the reporter assays and ubiquitination experiment, full-length bfACP3 was cloned into the pCMV-Myc vector (Clontech) and bfMyD88 was constructed in the same way as bfTRAF6. The full-length sequences of bfMyD88 and bfTRAF6 are shown in the **Supplementary Figures 6** and **7**, respectively. The ClonExpress® II Kit (Vazyme) was used for the construction of recombinant expression vectors. The vectors were verified by sequencing and the expression of proteins were confirmed by western blot. Primers were described in **Table 1**.

Luciferase Reporter Assays

HEK293T cells were digested by trypsin and seeded in 48-well plates. After 24h, the cells were transfected with equivalent mixed expression plasmids as previously descripted (27), which consist of the indicated amount of expression vectors, 50 ng per well of the NF-κB response promoter luciferase reporter plasmid pNF-κB-Luc (StrataGene), and 5 ng *Renilla* luciferase reporter plasmid pRLTK (Promega) per well to normalize the data due to different transfection efficiency between wells, and empty vectors to complement the total plasmid quantity of each well to the same. After 24 h, HEK293T cells were harvested and measured by dual-luciferase reporter assay system (Promega). Each experiment was performed in triplicate and was repeated at least three times in all cases. The data is shown as the fold change relative to that measured in cells transfected with empty vector.

Western Blot and Co-Immunoprecipitation (Co-IP) Assays

Western blot and Co-IP assays were performed to demonstrate the interaction of amphioxus bfACP3 and bfTRAF6. HEK293T cells in six-well dishes were transfected with 6μg constructed plasmid (3 μg/each expression vector) after 24 h seeded. At 24h post-transfection, whole-cell extracts were prepared in IP lysis buffer (Cell Signaling Technology) supplemented with cocktail protease inhibitor (Roche) on ice. Then the cell lysates were fractionated by 10% SDS-PAGE and then transferred onto a nitrocellulose filter membrane. The membrane was blocked with 5% skim milk in PBST (0.05%Tween-20 in PBS) at room temperature for 1 h and then incubated with the corresponding primary antibody (Sigma) at 4°C overnight. Wash the blot in PBST three times and incubated with HRP-conjugated Goat Anti-Mouse IgG (Abmart) at room temperature for 1 h and detected with enhanced chemiluminescence reagent (Millipore, USA). For Co-IP assays, the cell lysates were incubated with 1μg primary antibodies (Sigma) at 4°C overnight then incubated with Protein G Sepharose (Roche) at 4°C for 4-6 h. The mixture was washed three times with cell lysis buffer and analyzed by Western blot.

Subcellular Localization Analysis

Hela cells were seeded on coverslips (10 mm×10 mm) in a 24-well plate. After 24 h, cells were transfected with 2μg indicated expression plasmids (1 μg/each expression vector) by Lipofectamine 3000 (Thermo) according to the manufacturer's instructions. After 24 h since transfection, cells we washed twice in PBS, fixed in 4% paraformaldehyde solution, and permeabilized by 0.1% Triton X-100 in PBST (0.05%Tween-20 in PBS). After washing with PBST three times, the wells were blocked with PBST containing 10% (wt/vol) BSA at room temperature for 1 h. Cells were further incubated with primary mAb (Sigma) at 4°C overnight, washed three times in PBST, and incubated with the second antibody Alexa Fluor 568 Goat anti-

rabbit IgG (Invitrogene) or Alexa Fluor 488 Goat anti-mouse IgG (Invitrogene) for 1 h. Following triple washing in PBS, cells were stained with 0.2 μg/mL DAPI for 5min. The slips were washed three times in PBS, mounted in ProLong™ Glass Antifade Mountant (Thermo) and photographed with a Carl Zeiss Axio Imager Z1 fluorescence microscope.

Statistical Analysis

Student's t-tests were performed to compare the means of three sample groups. In all cases, differences of $p<0.05$ were considered significant. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$ indicate statistical differences. All experiments were repeated at least three times.

RESULTS

Cloning and Sequence Analysis of Amphioxus ACP3 and ACP5

We previously identified 21 ACP gene models from the genome and transcriptomes of the amphioxus *Branchiostoma floridae* (bf) (2). In this study, we cloned the full-length cDNA sequences for bfACP3 and bfACP5, which encode a 561aa protein and a 636aa protein, respectively (**Supplementary Figures 1** and **2**). ACP3 has a signal peptide, an unknown conserved region (**Supplementary Figure 3**) and a C-terminal ApeC domain. ACP5 structurally resembles ACP3, but has a complement-related CUB domain right behind the signal peptide (**Figure 1A**). Phylogenetic analysis based on the ApeC domains suggests that ACP3 and ACP5 are close paralogs to the previously-reported ACP1 and ACP2 (**Figure 1B**, **Supplementary Figures 4** and **5**). But differences are also obvious. For example, ACP3/ACP5 have a longer middle spacer region than ACP1/ACP2. ACP5 even has a CUB domain (**Figure 1A**). The CUB domain (for complement C1r/C1s, Uegf, Bmp1) is almost exclusively found in extracellular matrix or plasma proteins (28). Based on the multiple alignment of the ApeC domain, ACP3 shared only 50% sequence identity with ACP1 and 47% with ACP2, while ACP5 shared only 47% sequence identity with ACP1 and 46% with ACP2. The sequence identity decreases in the region outside the ApeC domain. A typical ApeC domain has eight conserved cysteines and three DXED motifs. The ApeC domains of amphioxus ACP1 and ACP2, as well as many other invertebrate ACPs, preserve all these conserved features. Although ACP3 and ACP5 preserve all eight cysteines, ACP3 has mutations in the first and third DEXD motifs, and ACP5 has mutations in all three DXED motifs (**Figure 1C**). We speculate that the deviation from the canonical motif may have functional implications, but so far the function of these motifs remains unknown. Overall, amphioxus ACP3 and ACP5 have their own sequence and structural characteristics, though belonging to the same class with ACP1 and ACP2.

The Expression Patterns of *bfACP3* and *bfACP5*

Realtime quantitative RT-PCR analyses showed that *bfACP1* and *bfACP2* had substantial expressions in at least four of six examined tissues of adult amphioxus (**Figures 2A, B**). In particularly, they exhibited high expressions in the gill and skin, which are in direct contact with the environmental microbial pathogens. On the other hand, the mRNA of *bfACP3* and *bfACP5* had limited tissue distribution, with *bfACP3* predominantly expressed in the intestine and *bfACP5* predominantly in the intestine and hepatic cecum (**Figures 2C, D**). After challenged with formalin-inactivated bacteria (the Gram-positive *Staphylococcus aureus* plus the Gram-negative *Vibrio anguillarum*), the mRNA of all four amphioxus ACPs could be upregulated in the gut and gill in adult amphioxus (**Figures 2E–H**). However, *bfACP3* and *bfACP5* were less responsive in the gill than in the gut, despite that their basal expression in the gill was extremely low (0.5-0.8% of the *gapdh* expression). The expression of *bfACP3* and *bfACP5* also responded to the major components (LPS & LTA) of the bacterial cell walls, but the profiles were quite complicated (**Figures 2I–L**). Taken together, while amphioxus *ACP1* and *ACP2* have substantial expressions in multiple tissues and are capable of dramatic upregulation during the acute immune response, the expression and function of *ACP3* and *ACP5* appear to mainly locate in the gut (intestine and hepatic cecum).

BfACP3 and bfACP5 Bind and Aggregate Different Microbes

Here we showed that bfACP3 and bfACP5 could be greatly upregulated in the gut by bacterial stimulation (**Figure 2**), so they might also have the capacity to bind microbes. To test this, we prepared purified recombinant His-tagged TRX-bfACP3 and TRX-bfACP5 fusion proteins (**Figure 3A**). Since bfACP3 and bfACP5 have many cysteines for potential intermolecular disulfide bonds (**Figure 1C**), we performed reduced and non-reduced SDS-PAGE assays and found that both proteins could form dimers or tetramers (**Figure 3B**). We then incubated the recombinant proteins with different microbes. The microbial pellets were assessed by western blot using anti-His monoclonal antibodies. The results showed that bfACP3 bound strongly to yeasts (*S. cerevisiae*) and Gram-positive bacteria, but bound to Gram-negative bacteria with weaker affinity (**Figure 3C**). Besides, based on the band intensity estimation, up to 50% of the bfACP3 recombination proteins in this assay bound to the microbes. As for bfACP5, it could bind with all the tested microbes (**Figure 3C**). Moreover, fluorescence microscopy showed that both bfACP3 and bfACP5 could aggregate FITC-labeled *S. aureus*, *E. coli*, *V. parahaemolyticus* and yeast *S. cerevisiae* (**Figure 3D**). Quantification of the diameter of green microbial puncta showed a significant agglutination effect (**Figure 3E**). In a previous study, ACP1 and ACP2 were shown to bind and aggregate Gram-positive bacteria but not Gram-negative bacteria (1). Therefore, the microbial binding spectrum of the amphioxus ACP3 and ACP5 seems quite different from that of ACP1 and ACP2. We wondered if such differences were related to the less conserved DEXD motifs in the ApeC domain of ACP3 and ACP5 (**Figure 1B**).

BfACP3 and bfACP5 Have No Inhibitory Activity Against *S. aureus* and *V. parahaemolyticus*

We monitored the growth curves of *S. aureus* and *V. parahaemolyticus* in the presence of bfACP3 or bfACP5

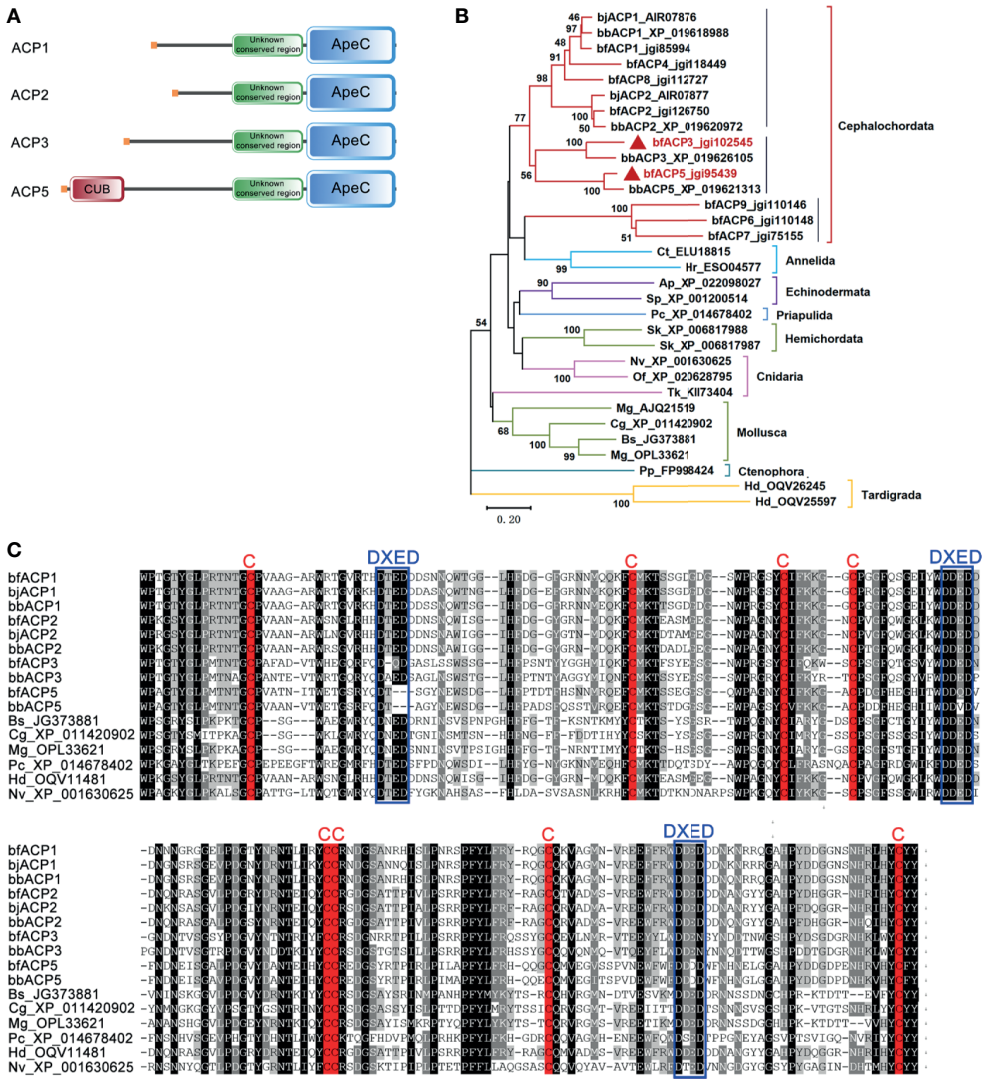


FIGURE 1 | Comparison of the amino acid sequences of bfACP3 and bfACP5 to other ACPs. **(A)** The domain architectures of amphioxus ACP3 and ACP5 compared with ACP1 and ACP2. **(B)** The phylogenetic tree of representative ACPs from different aquatic animal phyla based on the sequence of ApeC domains. The tree was constructed using the neighbor-joining method. Numbers on the lines indicate the percentage bootstrap values for 1000 replicates. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. A Maximum Likelihood tree based on the same alignment is provided in **Supplementary Figure 4** and a more detailed version is shown in **Supplementary Figure 5**. **(C)** Multiple alignment of the ApeC domains of ACPs from amphioxus and other aquatic species. The conserved Cysteine residues and DXED motifs are marked with red and blue boxes, respectively. bj, *Branchiostoma japonicum*; bf, *Branchiostoma floridae*; bb, *Branchiostoma belcheri*; Bs, *Botryllus schlosseri*; Cg, *Crossostrea gigas*; Mg, *Mytilus galloprovincialis*; Pc, *Priapulus caudatus*; Hd, *Hypsibius dujardini*; Nv, *Nematostella vectensis*; Ap, *Acanthaster planci*; Sp, *Strongylocentrotus purpuratus*; Sk, *Saccoglossus kowalevskii*; Tk, *Thelohanellus kitauei*; Of, *Orbicella faveolata*; Ct, *Capitella teleta*; Hr, *Helobdella robusta*; Pp, *Pleurobrachia pileus*.

proteins (with Ampicillin as a positive control). Both proteins had a negligible effect on the growth rates of the bacteria under these experimental conditions (**Figures 4A, B**). Moreover, the Oxford cup experiments showed that TRX-bfACP3 and TRX-bfACP5 fusion proteins had no discernable inhibitive effect on *S. aureus* and *V. parahaemolyticus* (**Figures 4C, D**). These results are reminiscent of a previous study, in which amphioxus ACP1 and ACP2 showed no detectable bacteriostatic or bactericidal activity against *S. aureus* (1). In conclusion, amphioxus ACP1, ACP3, ACP3 and ACP5 may all have microbial binding and aggregating activities, but have no killing activity.

Binding Specificity of bfACP3 and bfACP5 to Microbial Cell Wall Components

We performed ELISA assays to detect which microbial cell wall component was recognized by bfACP3 and bfACP5. The results indicated that both bfACP3 and bfACP5 recombinant proteins had a high binding affinity with the soluble PGN

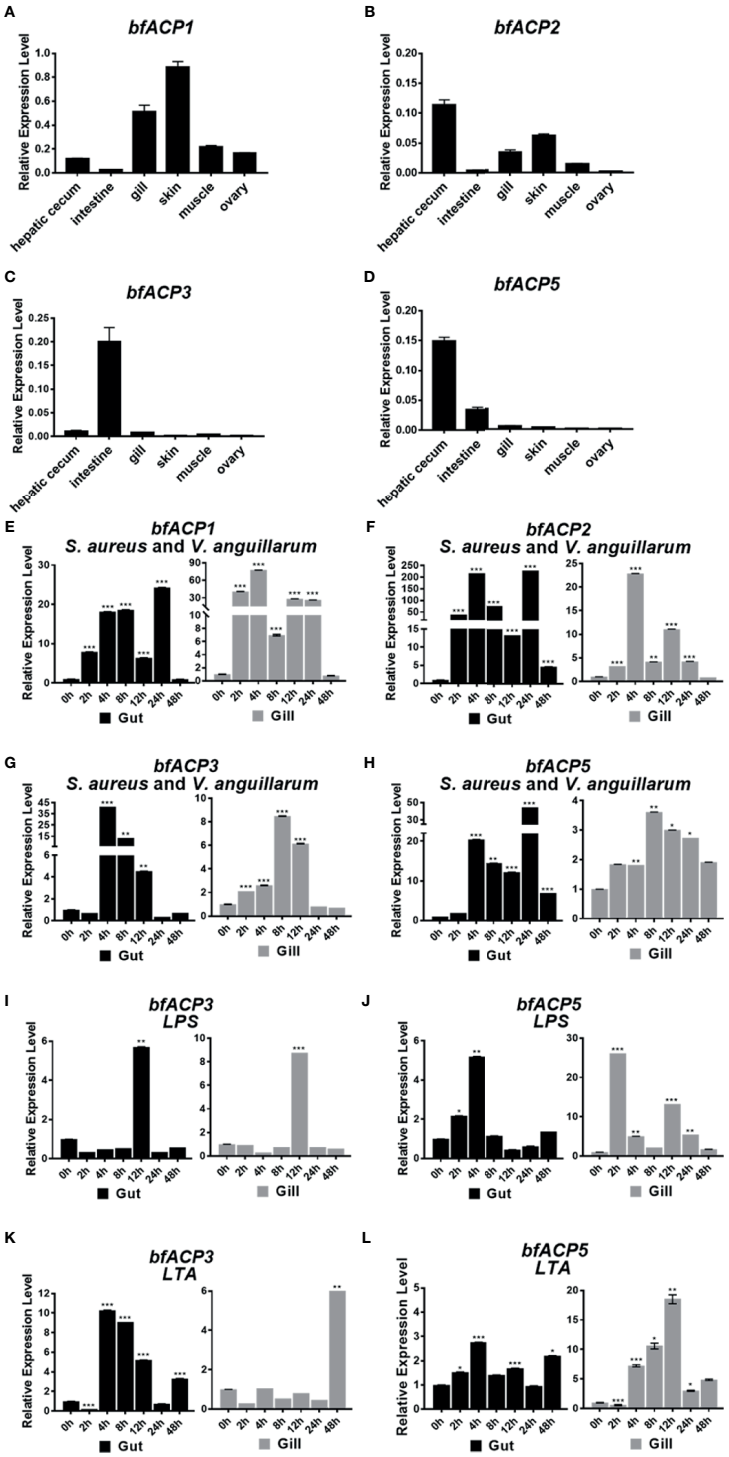


FIGURE 2 | Q-PCR analysis of the expression profile of *bfACP3* and *bfACP5*. The relative level of *bfACP1* **(A)**, *bfACP2* **(B)**, *bfACP3* **(C)** and *bfACP5* **(D)** mRNA in different tissues, respectively. Experiments were performed with five amphioxus. Data were expressed as a ratio to the *gapdh* mRNA expression and were plotted as the mean ± SD. **(E–H)** The relative expression level of *ACP* mRNA in gut (hepatic cecum and intestine) and gill after challenge with mixed inactivated bacteria (*S. aureus* and *V. anguillarum* in 1:1 ratio). **(I–L)** The relative expression level of *bfACP3 and bfACP5* mRNA in gut (hepatic cecum and intestine) and gill after challenge with Gram-negative cell wall component LPS (1mg/mL) and Gram-positive bacteria cell wall component LTA (1mg/mL) for different time. Data were expressed as a ratio to the ACP mRNA expression level of unchallenged naive animals and were normalized to the *gapdh* expression. All the samples were analyzed in three replicates and mean ± SD is plotted, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. One representative experiment out of three is shown.

from *S. aureus*, but not with LPS, LTA, mannose, zymosan A and the PGN from *B. subtilis* (**Figures 5A, B**). PGN can be classified into the DAP-type and the Lys-type according to the difference in amino acid residues and cross-linking methods (29). PGN from *S. aureus* is Lys-type and PGN from *B. subtilis* is DAP-type. Therefore, it seems that bfACP3 and bfACP5 could only recognize the Lys-type PGN. Pull-down assays confirmed that bfACP3 and bfACP5 were directly bound to insoluble PGN from *S. aureus* in a dose-dependent manner but not to PGN from *B. subtilis* (**Figure 5C**). The basic building blocks of PGN include N-acetylglucosamine (GlcNAc; NAG), N-acetylmuramic acid (MurNAc; NAM) and Muramyl dipeptide (MDP), but we detected no interaction between these motifs with our recombinant bfACP3 and bfACP5 (**Figures 5D, E**). This contrasted the finding that the recombinant proteins of amphioxus ACP1 and ACP2 could bind with MDP (1). We suspect that the intact molecule or certain conformation of Lys-type PGN are required for the stable interaction with bfACP3 and bfACP5. So far, we may conclude that bfACP3 and bfACP5 could agglutinate Gram-positive bacteria by recognizing their cell-wall component PGN, though there is more to be done to explain how bfACP3/bfACP5 could bind with yeast and how bfACP5 could bind with Gram-negative bacteria.

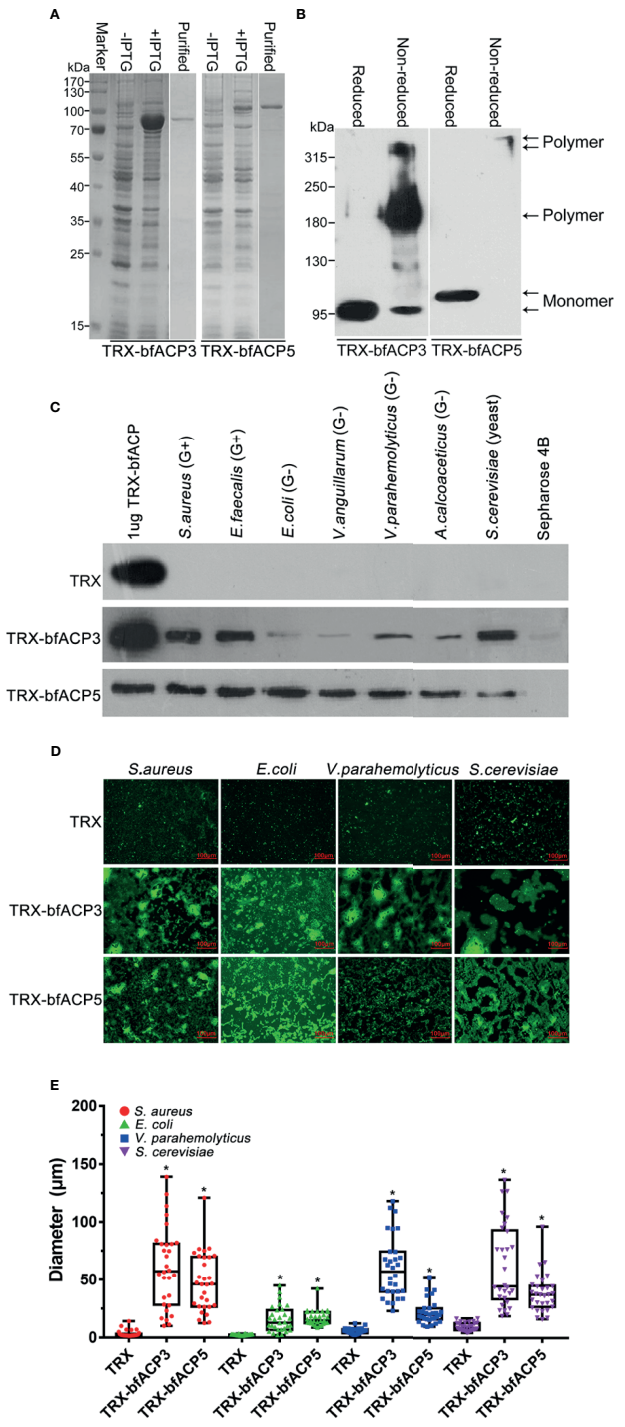


FIGURE 3 | Continued

FIGURE 3 | Binding and aggregation of the microbes by bfACP3 and bfACP5. **(A)** SDS-PAGE analysis of samples taken during the purification of recombinant TRX-bfACP3 and TRX-bfACP5. **(B)** Reducing and Non-reducing SDS-PAGE of ACPs. The bands corresponding to the monomer or oligomer were marked. **(C)** The binding of microorganisms by recombinant TRX-bfACP3 and TRX-bfACP5 protein. Approximately 2×10^6 living microbes were incubated with 1μg TRX fusion proteins in 1 mL PBS at 4°C overnight and the stirringly washed pellets were subjected to the SDS-PAGE and detected by Western blot with anti-6×His monoclonal antibody. One representative experiment out of three is shown. **(D)** Aggregation of the microbes by TRX-bfACP3 and TRX-bfACP5. 10μg TRX fusion proteins were incubated with 50 μl FITC-labeled *S. aureus* (2×10^8 cells/mL), *E coli* (2×10^8 cells/mL), *V. parahemolyticus* (2×10^8 cells/mL) or *S. cerevisiae* (2×10^7 cells/mL) at room temperature in the dark for 2h, respectively. The agglutinating reaction was examined using fluorescence microscopy. **(E)** Box plot showing the diameters of green puncta in microbial aggregation assays. Box plot explanation: upper horizontal line of box, 75th percentile; lower horizontal line of box, 25th percentile; horizontal bar within box, median; upper horizontal bar outside box, maximum value; lower horizontal bar outside box, minimum value. $^*p < 0.05$ *versus* TRX control.

BfACP3 Suppressed MyD88 and TRAF6 Mediated NF-κB Activation

It was shown that when retained in cytoplasm, amphioxus ACP2 could interact with TRAF6 and suppress TRAF6 mediated NF-κB activation (1). We wondered if bfACP3 has similar intracellular functions. We transfected the HEK293T cells with the NF-κB luciferase reporter plasmid and the bfACP3 expression plasmid, together with the bfMyD88 or bfTRAF6 expression plasmids. Luciferase reporter assays showed that bfACP3 alone could not affect the NF-κB signal, but could inhibit the NF-κB signal activated by bfMyD88 and bfTRAF6 in a dose-dependent form (**Figures 6A–C**). Co-IP assays using 293T cells showed that bfACP3 could physically interact with bfTRAF6 (but not with bfMyD88) in our experimental condition (**Figure 6D**). Immunofluorescence imaging analysis using HeLa cells confirmed that bfACP3 and bfTRAF6 could co-localize in the same subcellular punctate structures (**Figure 6E**). These results suggested that when presented in cytoplasm, bfACP3

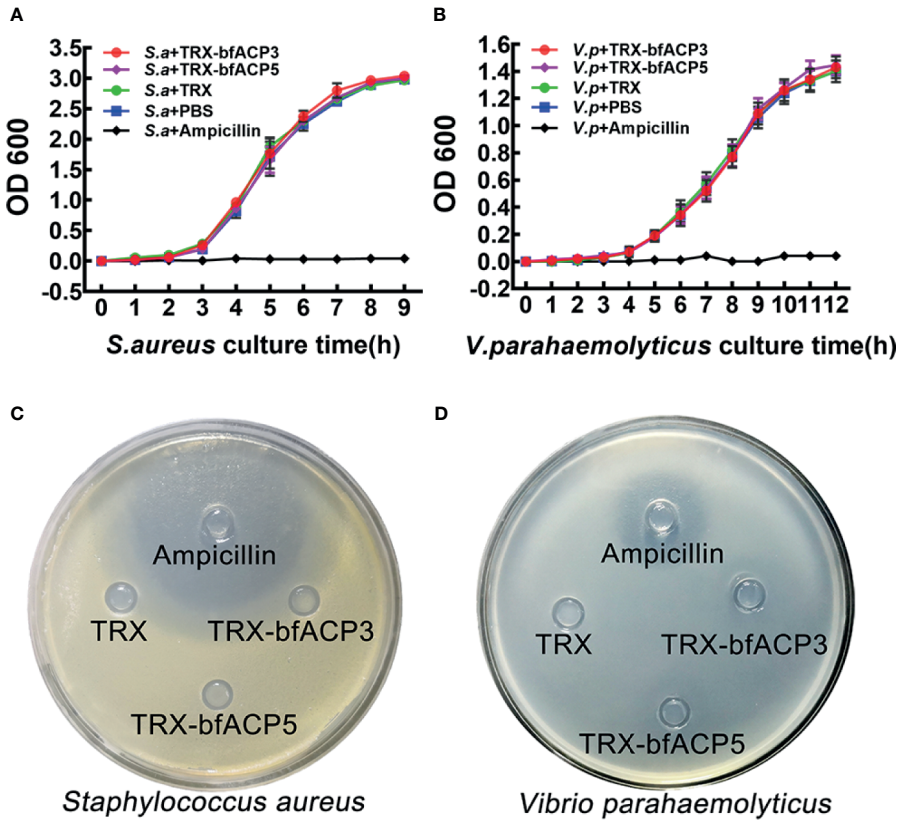


FIGURE 4 | BfACP3 and bfACP5 are not antibacterial against *S. aureus* and *V. parahaemolyticus*. Growth curves of *S. aureus* **(A)** and *V. parahaemolyticus* **(B)** in the presence of TRX-bfACP3 or TRX-bfACP5 in medium while being shaken. OD_{600} was measured every 1 h after starting the culture (mean ± SD, n = 3). Oxford Cup experiments cultured with *S. aureus* **(C)** or *V. parahaemolyticus* **(D)** were performed with 100ul Ampicillin (100μg/mL), TRX (0.2μg/μl), TRX-bfACP3 (0.2μg/μl) and TRX-bfACP5 (0.2μg/μl). Then, the plates were incubated at 37°C for 16 h (*S. aureus*) or 28°C for 40 h (*V. parahaemolyticus*). A transparent ring around the cups signifies antibacterial activity.

could negatively regulate the MyD88-TRAF6-NF-κB pathway by interacting with bfTRAF6.

BfACP3 Interfered With the Self-Ubiquitination of TRAF6

K63-linked polyubiquitination of TRAF6 is essential for NF-κB signaling (30). Such self-linked polyubiquitin chains serve as a scaffold to assemble the downstream signaling complex consisting of TAB2, NEMO, TAK1, and IKKs, which subsequently leads to IκB degradation and NF-κB activation (31). K48-linked polyubiquitination of TRAF6, however, might lead to protein degradation and hence negatively regulate the NF-κB pathway (32). We performed ubiquitination assays by transfecting 293T cells with Flag-tagged bfTRAF6 and HA-tagged wild-type or mutant ubiquitin (Ub, Ub-K63 and Ub-K48) in the presence or absence of bfACP3 proteins. We found that bfACP3 could significantly inhibit the ubiquitination of bfTRAF6, including both K63- and K48-ubiquitination (**Figure 6F**). This suggests that like amphioxus ACP2, bfACP3 might suppress TRAF6 functions by interfering with its ubiquitination.

BfACP3 Used an Unknown Conserved Region to Interact With TRAF6

The amphioxus ACP2 could use its N-terminal non-ApeC region (**Figure 1A**) to interact with TRAF6, and putative TRAF6-binding motif could be found in this region (1). However, we did not find any TRAF6-binding motif in bfACP3, though it has an even longer N-terminal non-ApeC region (**Figure 1A**). We guessed that other sequences might be responsible for the interaction between bfACP3 and bfTRAF6. To determine which sequence of bfACP3 was used to recognize bfTRAF6, we constructed different truncated forms of bfACP3 (**Figure 6G**). Co-IP assays revealed that the 205-358aa region played the main function of interacting with TRAF6 (**Figure 6H**). This region is conserved among ACP1, ACP2, ACP3 and ACP5, located in the N-terminal non-ApeC region and adjacent to the ApeC domain (**Figure 1A** and **Supplementary Figure 3**). The identities of this region between different ACPs are modest (43-64%). It is rich in leucine but does not have a typical leucine-rich repeat (LRR). These results suggested that bfACP3 used an unknown conserved sequence in its N-terminal non-ApeC region to interact with bfTRAF6.

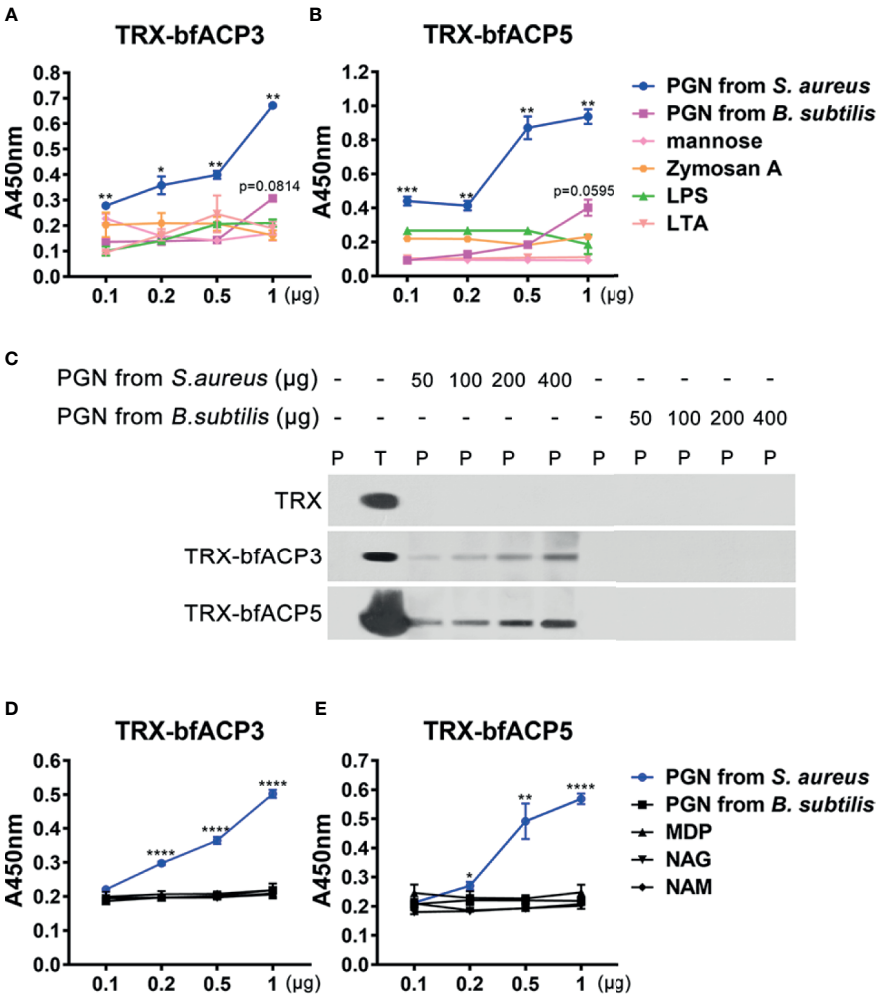


FIGURE 5 | BfACP3 and bfACP5 directly interacted with the components of the microorganism cell walls. **(A, B)** ELISA analysis of the interaction between recombinant fusion TRX-bfACP3 and TRX-bfACP5 to the components, respectively. Plates were coated with 20μg components, incubated with TRX-bfACP3 or TRX-bfACP5 at 37°C overnight and detected with anti-6×His monoclonal antibody. Three biological replicates were designed for each experiment, and three technical replicates were performed, showing one of the representative results. Background absorbance with TRX was subtracted. **(C)** Pull-down analysis of the binding of 5μg recombinant fusion ACPs to PGN from *S. aureus* or *B subtilis*. P, pellet protein; T, total protein. **(D, E)** ELISA analysis of the interaction between recombinant fusion ACPs and 20μg monomers that are parts of peptidoglycan. NAG, GlcNAc; NAM, MurNAc. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$.

DISCUSSION AND CONCLUSIONS

ApeC Is a Novel Protein Domain Widely Distributed in Invertebrates

ApeC-containing proteins consist of a large protein family which are widely distributed in invertebrates, especially in invertebrates from aquatic and humid environments (2). ApeC-like domains were also found in bacteria, suggesting the ancient origins and broad distribution of ApeC (2). ApeC is a typical promiscuous domain like the IgSF domains and C-type lectin domains, capable of forming domain combinations with various domains. So far tens of different domain architectures have been documented in ACPs. Gene expression regulation and tissue distribution of some ACPs have been examined in different invertebrates, including sea urchins, sea cucumbers, corals, amphioxus, oysters and mussels, etc. (8–13, 33–36). Current data suggest that some ACPs may have an important role in embryogenesis, development and immune responses (5, 7–13, 33–41), but in most cases, the underlying mechanisms have not been investigated. In fact, most ACPs have not been functionally investigated.

Apextrin is the first ACP subfamily received detailed study. A typical apextrin has a signal peptide, an N-terminal membrane attack complex/perforin (MACPF) domain and a C-terminal ApeC domain, though the ApeC domain had not been recognized in early days. An apextrin from the sea urchin was found to be expressed in large quantities from fertilization to metamorphosis, and was proposed to be involved in the adhesion of apical cells and strengthening the outer layer of the embryo (6, 7), despite that the underlying molecular mechanism remains

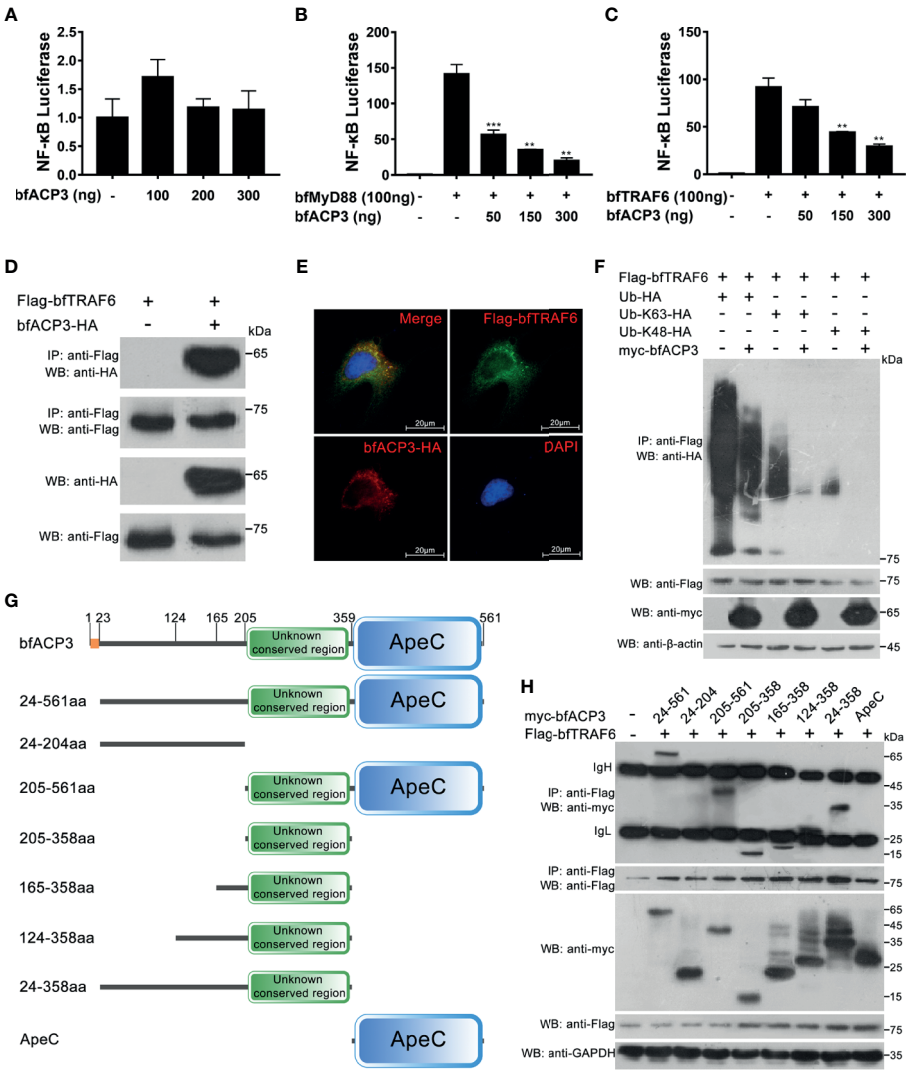


FIGURE 6 | BfACP3 negatively regulated TRAF6-NF-κB pathway by suppressing the ubiquitination of bfTRAF6. For luciferase reporter assays, HEK293T cells were co-transfected with NF-κB transcriptional luciferase reporter plasmid, *Renilla* luciferase reporter plasmid, bfMyD88 or bfTRAF6 vectors, together with bfACP3 vector. For Co-IP assays and colocalization assay, HEK293T cells and Hela cells were used, respectively. **(A)** BfACP3 hardly activated NF-κB signal. **(B, C)** BfACP3 negatively regulated bfMyD88-induced **(B)** and bfTRAF6-induced **(C)** NF-κB activation. The representative results are shown as means ± SD (n=3) of three experiments. $^{**}p < 0.01$, $^{***}p < 0.001$. **(D)** Co-IP assay showing that bfACP3 interacts with amphioxus bfTRAF6 when overexpressed in HEK293T cells. **(E)** Immunofluorescence analysis of the subcellular co-localization of bfTRAF6 and bfACP3. HeLa cells were co-transfected with HA-tagged bfACP3 and Flag-tagged bfTRAF6, then stained with rabbit anti-HA and mouse anti-Flag antibody, followed by incubating with the Alexa Fluor 568 goat anti-rabbit and Alexa Fluor 488 goat anti-mouse secondary antibody respectively. **(F)** Ubiquitination assays indicate that bfACP3 suppressed the polyubiquitin chains of bfTRAF6. **(G)** The full-length and truncated mutants of bfACP3 used in this study. The amino acids were numbered according to bfACP3 sequence. **(H)** Co-IP assay between bfACP3 mutants and bfTRAF6 indicated that the unknown conserved non-ApeC region of bfACP3 is responsible for the interaction with bfTRAF6.

unclear. Later, apextrins have been found in at least five animal phyla (2).

A Subfamily of Amphioxus ACPs Act as Pattern Recognition Proteins

An early study suggested that amphioxus ACP1 and ACP2 (both from *B. japonicum*) could function as pathogen-associated molecular pattern (PAMP) recognition proteins (1). In this study, we characterized another two amphioxus ACPs (ACP3 and ACP5 from *B. floridae*). ACP3/5 and ACP1/2 may have related functions, because the similar domain architecture and the close phylogenetic relation suggest that they belong to the same subfamily (**Figures 1A, B**). But they also have distinct differences. ACP3 and ACP5 only share 40-50% sequence identities with ACP1 and ACP2. ACP3 and ACP5 have longer N-terminal sequences than ACP1 and ACP2. ACP5 even has an additional CUB domain, making it more akin to extracellular matrix proteins or humoral proteins. In the complement

proteases, CUB domains can mediate dimerization and binding to collagen regions of partner proteins (28).

ACP1 and ACP2 are highly expressed in several mucosal and non-mucosal tissues, with the highest concentration in the gill and skin. While ACP3 and ACP5 have a more limited tissue distribution, predominantly expressed in the gut. This suggests that though they all function in mucosal tissues, ACP3/5 and ACP1/2 specialize in different niches. In line with this, they have different microbial binding spectrum. Recombinant ACP1 and ACP2 were shown to bind Gram-positive bacteria (1), while recombinant ACP5 exhibited binding capacity with yeasts, Gram-positive and Gram-negative bacteria (**Figure 3**). Recombinant ACP3 had a similar spectrum with ACP5, but its binding strength to Gram-negative bacteria was much weaker (**Figure 3**). Despite all these, no recombinant proteins (ACP1, ACP2, ACP3 and ACP5) showed killing or inhibitive effects on microbes (**Figure 4**), suggesting that they more likely act as mere lectins.

The bacterial cell wall component PGN and its basic active motif MDP have been identified as binding ligands for recombinant ACP1 and ACP2 (1). Though recombinant ACP3 and ACP5 were also found to recognize PGN, they displayed a more selective specificity. They only bound to the Lys-type PGN of Gram-positive bacteria, and had no detectable affinity for three basic motifs of PGN, including MDP, NAM and NAG (**Figure 5**). Moreover, recombinant ACP3 and ACP5 were also found to bind yeasts and some Gram-negative bacteria (**Figures 3C, D**), but we failed to determined what ligands they used to recognize yeasts and Gram-negative bacteria (**Figures 3–5**). There are several possibilities related to the observed binding specificities: ACP3/5 have different structural characteristics than ACP1/2, the recombinant ACP3/5 might not preserve all the properties of the native proteins, and the purified microbial cell wall components might lose their native conformation.

This Subfamily of Amphioxus ACPs Regulate the TRAF6-NF-κB Pathway

In addition to being an extracellular effector, intracellular ACP2 was found capable of modulating the TRAF6-NF-κB pathway, possibly as a feedback inhibitory mechanism to control the magnitude of the immune response (1). It could interact with TRAF6 and prevent TRAF6 from self-ubiquitination and hence from activating NF-κB. Here we found that ACP3 also has this intracellular function, to suppress the TRAF6 induced NF-κB activation by binding TRAF6 and interfering with the ubiquitination of TRAF6 (**Figure 6**). Here we further pinpointed the region responsible for the TRAF6 binding, and this region is adjacent to the ApeC domain and relatively conserved in ACP1, 2, 3 and 5 (**Figure 1A** and **Supplementary Figure 3**).

CONCLUDING REMARKS

Taken together, this study provides a second but different line of mechanistic evidence for the immune role of ACPs in amphioxus. This study, together with the previous studies (1, 15, 38), defines a novel subfamily of ACPs, which share a conserved structure and have similar yet diversified molecular functions. The members of this ACP subfamily adopt a dual-functional mode: the ApeC domain serves the lectin role, while a conserved unknown region serves as a signal transduction regulator for the TRAF6-NF-κB pathway. This work broadens our understanding of the ACP functions and may facilitate further research efforts on the role of ACPs in other animal clades.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

JL, SH and AX designed the study. JL and YL performed the experiments. JL, YL and SH analyzed the data. ZF, SheC, XY, ZY, GH, SL, MD, ShaC, and HZ also conducted some experiments, provided reagents or analyzed the data. JL and SH drafted the manuscript. AX revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by NNSF Projects (31872595 & 31722052 & 31971107), National Key R&D Program of China (2018YFD0900503), Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (2018SDKJ0302-2), Project supported by Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (No. 311021006), and projects from Guangdong and Guangzhou (2021A1515012380 & 2020B1212060031).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.715245/full#supplementary-material

REFERENCES

1. Huang G, Huang S, Yan X, Yang P, Li J, Xu W, et al. Two Apextrin-Like Proteins Mediate Extracellular and Intracellular Bacterial Recognition in Amphioxus. *Proc Natl Acad Sci USA* (2014) 111(37):13469–74. doi: 10.1073/pnas.1405414111
2. Li Y, Li J, Yan X, Chen S, Wu C, Huang H, et al. Broad Distribution, High Diversity and Ancient Origin of the ApeC-Containing Proteins. *Mol Phylogenet Evol* (2021) 155:107009. doi: 10.1016/j.ympev.2020.107009
3. Drickamer K, Taylor ME. Recent Insights Into Structures and Functions of C-Type Lectins in the Immune System. *Curr Opin Struct Biol* (2015) 34:26–34. doi: 10.1016/j.sbi.2015.06.003

4. Williams AF, Barclay AN. The Immunoglobulin Superfamily–Domains for Cell Surface Recognition. *Annu Rev Immunol* (1988) 6:381–405. doi: 10.1146/annurev.iy.06.040188.002121
5. Minsuk SB, Raff RA. Pattern Formation in a Pentameral Animal: Induction of Early Adult Rudiment Development in Sea Urchins. *Dev Biol* (2002) 247 (2):335–50. doi: 10.1006/dbio.2002.0704
6. Haag ES, Sly BJ, Andrews ME, Raff RA. Apextrin, a Novel Extracellular Protein Associated With Larval Ectoderm Evolution in Heliocidaris Erythrogramma. *Dev Biol* (1999) 211(1):77–87. doi: 10.1006/dbio.1999.9283
7. Haag ESaR RA. Isolation and Characterization of Three mRNAs Enriched in Embryos of the Direct-Developing Sea Urchin Heliocidaris Erythrogramma: Evolution of Larval Ectoderm. *Dev Genes Evol* (1998) 208(4):188–204. doi: 10.1007/s004270050173
8. Estevez-Calvar N, Romero A, Figueras A, Novoa B. Involvement of Pore-Forming Molecules in Immune Defense and Development of the Mediterranean Mussel (Mytilus Galloprovincialis). *Dev Comp Immunol* (2011) 35(10):1017–31. doi: 10.1016/j.dci.2011.03.023
9. David E, Tanguy A, Pichavant K, Moraga D. Response of the Pacific Oyster Crassostrea Gigas to Hypoxia Exposure Under Experimental Conditions. *FEBS J* (2005) 272(21):5635–52. doi: 10.1111/j.1742-4658.2005.04960.x
10. McDowell IC, Modak TH, Lane CE, Gomez-Chiarri M. Multi-Species Protein Similarity Clustering Reveals Novel Expanded Immune Gene Families in the Eastern Oyster Crassostrea Virginica. *Fish Shellfish Immunol* (2016) 53:13–23. doi: 10.1016/j.fsi.2016.03.157
11. Dheilly NM, Haynes PA, Bove U, Nair SV, Raftos DA. Comparative Proteomic Analysis of a Sea Urchin (Heliocidaris Erythrogramma) Antibacterial Response Revealed the Involvement of Apextrin and Calreticulin. *J Invertebr Pathol* (2011) 106(2):223–9. doi: 10.1016/j.jip.2010.09.008
12. Huang GH, Liu H, Han Y, Fan L, Zhang Q, Liu J, et al. Profile of Acute Immune Response in Chinese Amphioxus Upon Staphylococcus Aureus and Vibrio Parahaemolyticus Infection. *Dev And Comp Immunol* (2007) 31(10):1013–23. doi: 10.1016/j.dci.2007.01.003
13. Gerdol M, Venier P. An Updated Molecular Basis for Mussel Immunity. *Fish Shellfish Immunol* (2015) 46(1):17–38. doi: 10.1016/j.fsi.2015.02.013
14. Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and Not Cephalochordates Are the Closest Living Relatives of Vertebrates. *Nature* (2006) 439(7079):965–8. doi: 10.1038/nature04336
15. Huang S, Chen Z, Yan X, Yu T, Huang G, Yan Q, et al. Decelerated Genome Evolution in Modern Vertebrates Revealed by Analysis of Multiple Lancelet Genomes. *Nat Commun* (2014) 5:5896. doi: 10.1038/ncomms6896
16. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The Amphioxus Genome and the Evolution of the Chordate Karyotype. *Nature* (2008) 453(7198):1064–71. doi: 10.1038/nature06967
17. Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, et al. The Amphioxus Genome Illuminates Vertebrate Origins and Cephalochordate Biology. *Genome Res* (2008) 18(7):1100–11. doi: 10.1101/gr.073676.107
18. Schubert M, Escriva H, Xavier-Neto J, Laudet V. Amphioxus and Tunicates as Evolutionary Model Systems. *Trends Ecol Evol* (2006) 21(5):269–77. doi: 10.1016/j.tree.2006.01.009
19. Louis A, Roest Crollius H, Robinson-Rechavi M. How Much Does the Amphioxus Genome Represent the Ancestor of Chordates? *Briefings Funct Genomics* (2012) 11(2):89–95. doi: 10.1093/bfgp/els003
20. Li G, Shu Z, Wang Y. Year-Round Reproduction and Induced Spawning of Chinese Amphioxus, Branchiostoma Belcheri, in Laboratory. *PloS One* (2013) 8(9):e75461. doi: 10.1371/journal.pone.0075461
21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. Mega X: Molecular Evolutionary Genetics Analysis Across Computing Platforms. *Mol Biol Evol* (2018) 35(6):1547–9. doi: 10.1093/molbev/msy096
22. Nicholas K, Nicholas HBJ. *GeneDoc: a Tool for Editing and Annotating Multiple Sequence Alignments. Distributed by the authors* (1997). http://www.psc.edu/biomed/genedoc.
23. Jones DT, Taylor WR, Thornton JM. The Rapid Generation of Mutation Data Matrices From Protein Sequences. *Comput Appl Biosci* (1992) 8(3):275–82. doi: 10.1093/bioinformatics/8.3.275
24. Whelan S, Goldman N. A General Empirical Model of Protein Evolution Derived From Multiple Protein Families Using a Maximum-Likelihood Approach. *Mol Biol Evol* (2001) 18(5):691–9. doi: 10.1093/oxfordjournals.molbev.a003851
25. Yan X, Chen S, Huang H, Peng T, Lan M, Yang X, et al. Functional Variation of IL-1R-Associated Kinases in the Conserved Myd88-TRAF6 Pathway During Evolution. *J Immunol (Baltimore Md 1950)* (2020) 204(4):832–43. doi: 10.4049/jimmunol.1900222
26. Yu Y, Yu Y, Huang H, Feng K, Pan M, Yuan S, et al. A Short-Form C-type Lectin From Amphioxus Acts as a Direct Microbial Killing Protein Via Interaction With Peptidoglycan and Glucan. *J Immunol (Baltimore Md 1950)* (2007) 179(12):8425–34. doi: 10.4049/jimmunol.179.12.8425
27. Yuan S, Huang S, Zhang W, Wu T, Dong M, Yu Y, et al. An Amphioxus TLR With Dynamic Embryonic Expression Pattern Responses to Pathogens and Activates NF-kappaB Pathway Via Myd88. *Mol Immunol* (2009) 46(11-12):2348–56. doi: 10.1016/j.molimm.2009.03.022
28. Blanc G, Font B, Eichenberger D, Moreau C, Ricard-Blum S, Hulmes DJ, et al. Insights Into How CUB Domains can Exert Specific Functions While Sharing a Common Fold: Conserved and Specific Features of the CUB1 Domain Contribute to the Molecular Basis of Procollagen C-Proteinase Enhancer-1 Activity. *J Biol Chem* (2007) 282(23):16924–33. doi: 10.1074/jbc.M701610200
29. Schleifer KH, Kandler O. Peptidoglycan Types of Bacterial Cell Walls and Their Taxonomic Implications. *Bacteriol Rev* (1972) 36(4):407–77. doi: 10.1128/br.36.4.407-477.1972
30. Deng L, Wang C, Spencer E, Yang LY, Braun A, You JX, et al. Activation of the I Kappa B Kinase Complex by TRAF6 Requires a Dimeric Ubiquitin-Conjugating Enzyme Complex and a Unique Polyubiquitin Chain. *Cell* (2000) 103(2):351–61. doi: 10.1016/s0092-8674(00)00126-4
31. Skaug B, Jiang XM, Chen ZJ. The Role of Ubiquitin in NF-Kappa B Regulatory Pathways. *Annu Rev Biochem* (2009) 78:769–96. Annual Review of Biochemistry. Palo Alto: Annual Reviews. doi: 10.1146/annurev.biochem.78.070907.102750
32. Zhou L, Ma Q, Shi HL, Huo KK. NUMBL Interacts With TRAF6 and Promotes the Degradation of TRAF6. *Biochem Biophys Res Commun* (2010) 392(3):409–14. doi: 10.1016/j.bbrc.2010.01.037
33. Puill-Stephan E, Seneca FO, Miller DJ, van Oppen MJ, Willis BL. Expression of Putative Immune Response Genes During Early Ontogeny in the Coral Acropora Millepora[J]. *PloS One* (2012) 7. doi: 10.1371/journal.pone.0039099
34. Hayward DC, Hetherington S, Behm CA, Grasso LC, Foret S, Miller DJ, et al. Differential Gene Expression at Coral Settlement and Metamorphosis - A Subtractive Hybridization Study[J]. *PloS One* (2011) 6(10):e26411–. doi: 10.1371/journal.pone.0026411
35. Grasso LC, Maindonald J, Rudd S, Hayward DC, Saint R, Miller DJ, et al. Microarray Analysis Identifies Candidate Genes for Key Roles in Coral Development. *BMC Genomics* (2008) 9:540. doi: 10.1186/1471-2164-9-540
36. Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, et al. The Innate Immune Repertoire in Cnidaria—Ancestral Complexity and Stochastic Gene Loss. *Genome Biol* (2007) 8(4):R59. doi: 10.1186/gb-2007-8-4-r59
37. Meyer GVA E, Matz MV. Profiling Gene Expression Responses of Coral Larvae (Acropora Millepora) to Elevated Temperature and Settlement Inducers Using a Novel RNA-Seq Procedure[J]. *Mol Ecol* (2011) 2011:20. doi: 10.1111/j.1365-294X.2011.05205.x
38. Huang S, Wang X, Yan Q, Guo L, Yuan S, Huang G, et al. The Evolution and Regulation of the Mucosal Immune Complexity in the Basal Chordate Amphioxus. *J Immunol* (2011) 186(4):2042–55. doi: 10.4049/jimmunol.1001824
39. Kang V, Lengerer B, Wattiez R, Flammang P. Molecular Insights Into the Powerful Mucus-Based Adhesion of Limpets (Patella Vulgata L.). *Open Biol* (2020) 10(6):18. doi: 10.1098/rsob.200019
40. van de Water J, Leggat W, Bourne DG, van Oppen MJH, Willis BL, Ainsworth TD. Elevated Seawater Temperatures Have a Limited Impact on the Coral Immune Response Following Physical Damage. *Hydrobiologia* (2015) 759 (1):201–14. doi: 10.1007/s10750-015-2243-z
41. Jiang K, Yin Z, Zhang Y, Xu Q, Yu Y, Cong W, et al. Genome-Wide Investigation and Expression Analysis of MACPF Gene Family Reveals Its Immune Role in Response to Bacterial Challenge of Manila Clam. *Genomics* (2021) 113(3):1136–45. doi: 10.1016/j.ygeno.2021.02.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fimmu.2021.747737

Convergent Loss of the Necroptosis Pathway in Disparate Mammalian Lineages Shapes Viruses Countermeasures

*Ana Águeda-Pinto[1,2], Luís Q. Alves[3], Fabiana Neves[1], Grant McFadden[4], Bertram L. Jacobs[4,5], L. Filipe C. Castro[2,3], Masmudur M. Rahman[4] and Pedro J. Esteves[1,2,6]**

[1] CIBIO/InBio-Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Vairão, Portugal, [2] Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal, [3] CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Matosinhos, Portugal, [4] Center for Immunotherapy, Vaccines and Virotherapy, The Biodesign Institute, Arizona State University, Tempe, AZ, United States, [5] School of Life Sciences Center for Immunotherapy, Vaccines and Virotherapy, Biodesign Institute, Arizona State University, Tempe, AZ, United States, [6] CITS-Centro de Investigação em Tecnologias da Saúde, Instituto Politécnico de Saúde do Norte (IPSN), Cooperativa de Ensino Superior Politécnico e Universitário (CESPU), Gandra, Portugal

OPEN ACCESS

Edited by:
Bostjan Kobe,
The University of Queensland,
Australia

Reviewed by:
Tudor Moldoveanu,
St. Jude Children's Research Hospital,
United States
Sarah J. Poynter,
Wilfrid Laurier University, Canada

***Correspondence:**
Pedro J. Esteves
pjesteves@cibio.up.pt

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: 26 July 2021
Accepted: 17 August 2021
Published: 01 September 2021

Citation:
Águeda-Pinto A, Alves LQ, Neves F, McFadden G, Jacobs BL, Castro LFC, Rahman MM and Esteves PJ (2021) Convergent Loss of the Necroptosis Pathway in Disparate Mammalian Lineages Shapes Viruses Countermeasures. Front. Immunol. 12:747737. doi: 10.3389/fimmu.2021.747737

Programmed cell death is a vital process in the life cycle of organisms. Necroptosis, an evolutionary form of programmed necrosis, contributes to the innate immune response by killing pathogen-infected cells. This virus-host interaction pathway is organized around two components: the receptor-interacting protein kinase 3 (RIPK3), which recruits and phosphorylates the mixed lineage kinase-like protein (MLKL), inducing cellular plasma membrane rupture and cell death. Critically, the presence of necroptotic inhibitors in viral genomes validates necroptosis as an important host defense mechanism. Here, we show, counterintuitively, that in different mammalian lineages, central components of necroptosis, such as RIPK3 and MLKL, are deleted or display inactivating mutations. Frameshifts or premature stop codons are observed in all the studied species of cetaceans and leporids. In carnivores' genomes, the MLKL gene is deleted, while in a small number of species from afrotheria and rodentia premature stop codons are observed in RIPK3 and/or MLKL. Interestingly, we also found a strong correlation between the disruption of necroptosis in leporids and cetaceans and the absence of the N-terminal domain of E3-like homologs (responsible for necroptosis inhibition) in their naturally infecting poxviruses. Overall, our study provides the first comprehensive picture of the molecular evolution of necroptosis in mammals. The loss of necroptosis multiple times during mammalian evolution highlights the importance of gene/pathway loss for species adaptation and suggests that necroptosis is not required for normal mammalian development. Moreover, this study highlights a co-evolutionary relationship between poxviruses and their hosts, emphasizing the role of host adaptation in shaping virus evolution.

Keywords: necroptosis, mammalian lineages, RIPK3, MLKL, convergent gene loss, host-range interactions

INTRODUCTION

Sensing of viral pathogens by the host cells is critical for animal survival. Thus, a variety of molecular responses, including the induction of inflammatory cytokines, chemokines and interferons, as well as the activation of cell-death pathways that provide clearance of pathogen-infected cells, require sensing of pathogen associated molecular patterns for activation. Although apoptosis has long been considered a critical clearance mechanism to control viral spread, caspase-independent cell death, or programmed necrosis, has recently emerged as an alternative death pathway that dominates under specific conditions (1).

Necroptosis is an inflammatory form of regulated necrosis that acts as an alternative host defense pathway during some viral infections and plays a major role in the killing and removal of pathogen-infected cells (1–3). Activation of necroptosis follows an intracellular signaling cascade that is dependent on the receptor-interacting serine/threonine-protein kinase 3 (RIPK3) and its substrate, the mixed lineage kinase like protein (MLKL) downstream of death receptors (DRs) and pattern-recognition receptors (PPRs) (**Figure 1**) (4, 5). Several pathway-specific adaptor proteins that contain a RIP homotypic interaction motif (RHIM-domain) can activate RIPK3-induced necroptosis (**Figure 1**). For example, when there is an interference or loss of function of caspase-8, the induction of necroptosis through the use of DRs results in the recruitment of RIPK1, which subsequently exposes its RHIM-domain to recruit RIPK3 (6–8). Apart from RIPK1, the TIR-domain-containing adaptor-inducing IFN β (TRIF), an essential protein downstream of Toll-like receptor (TLR)3/4 and the Z-DNA binding protein (ZBP1), also directly activate RIPK3 (9, 10) (**Figure 1**). Exposure of RIPK3 to a RHIM adaptor (RIPK1, TRIF or ZBP1) is a crucial step in the initiation of necroptosis, as these proteins activate the downstream executor of necroptosis, MLKL, that destabilizes the plasma membrane integrity leading to cell swelling followed by membrane rupture of infected cells and release of damage-associated molecular patterns (DAMPs) (3, 11). Thus, necroptosis provides a critical extra defense mechanism against pathogen infection, facilitating the elimination of virus-infected cells before the production of progeny virions. The importance of necroptosis for host defense is further supported by the identification of viral inhibitors of necroptosis, like is the case of Vaccinia virus (VACV) E3 protein and the murine cytomegalovirus (MCMV) M45 protein (9, 11).

Necroptosis has a major role in protecting cells against viral infection (1–3). However, despite the recent advances to understand the molecular regulation of this unique pathway, it

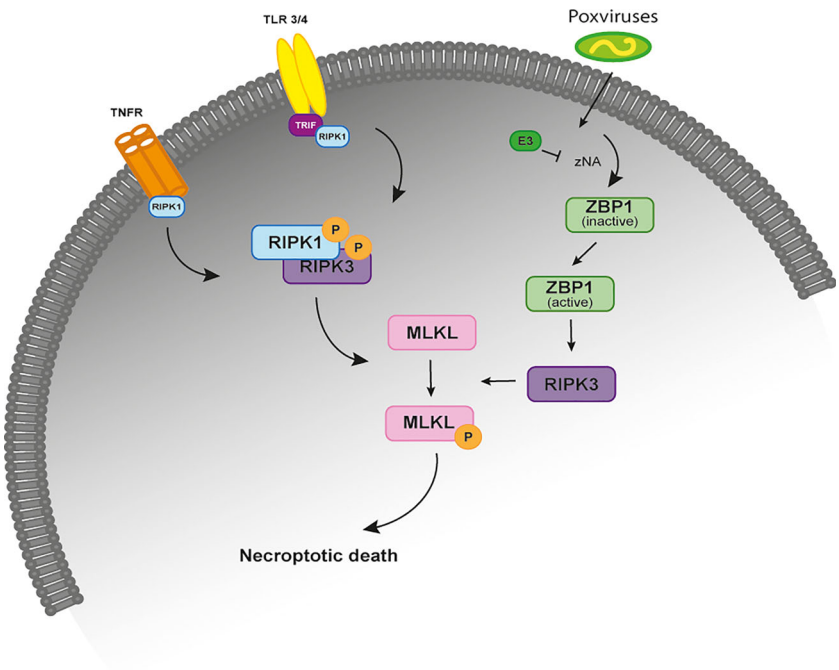


FIGURE 1 | The necroptosis signaling pathway. Simplified schematic representation of the necroptosis signaling pathway upon stimulation of the TNFR, TRF3/4 and infection by poxviruses. All of these necroptosis-inducing signals converge on the kinase RIPK3, which is activated through the homotypic interaction with RIPK1 or other RHIM-containing proteins, such as TRIF and DAI. When the activity of caspase-8 is inhibited, binding of TNF to TNFR1 leads to the phosphorylation and activation of RIPK1 that binds to RIPK3 through their RHIM domains to form a protein complex (necrosome). Activated RIPK3 recruits MLKL that oligomerizes and translocates to the plasma membrane to cause necroptosis. In TLR3- and TLR4-induced necroptosis, TRIF is required for the activation of RIPK3. ZBP1 is required for the activation of RIPK3 in response to the presence of Z-form nucleic acids. In VACV-infected cells, the poxviral E3 protein binds to VACV-induced Z-form nucleic acid, preventing RIPK3-induced necroptosis. TNFR, tumor necrosis factor receptor; TLR 3/4, toll-like receptor; TRIF, TIR-domain-containing adaptor-inducing IFN β; RIP, receptor-interacting protein kinase; ZBP1, Z-DNA binding protein; MLKL, mixed-lineage kinase domain like.

is still unclear whether the necroptotic cell death pathway acts as a universal cell death program in mammals. Previous studies suggested that components of necroptosis are absent in the genomes of extant birds and marsupials. Interestingly, within the Mammalia class, it was previously reported that order Carnivora lost the *MLKL* gene (12). Taken together these reports suggest some degree of plasticity in the conservation of necroptosis responses. Here, we address the molecular evolution of the necroptotic pathway in multiple mammalian lineages. We show that during mammalian evolution, necroptosis was convergently inactivated several times in mammalian lineages. Remarkably, we also report that mammalian orders that lost the necroptotic pathway display infection episodes by poxviruses that have lost the ability to inhibit this pathway, showing a co-evolutionary relationship between host adaptation in shaping virus evolution.

METHODS

Genomic Approach to Detect Genes Associated With the Necroptosis Pathway

To detect intact and inactivated genes, we first identified the key genes of the necroptotic pathway (i.e., *RIPK1*, *RIPK3* and *MLKL*) in the human (*H. sapiens*) and mouse (*Mus musculus*) reference genomes and looked for the presence of orthologues in existing genome sequence databases from 67 different species that belong to the 9 main mammalian orders/superorders: primates, rodents, lagomorpha, chiroptera, carnivora, perissodactyla, artiodactyla, cetacea, and afrotheria (**Supplementary File 1**). We did not only search for the complete loss of exons or entire genes, but also searched for insertions and deletions that shift the reading frame, frame-preserving insertions that create a premature stop codon and substitutions that create an in-frame stop codon. The respective search methodology had been previously applied to the identification of different homologues of different annotated mammalian genomes (13, 14). To further ensure that all gene loss events discussed in this study are real and not due to sequencing errors, we validated them either by sequencing of samples or by using a curated bioinformatic pipeline (see below).

Amplification and Sequencing of RIPK1 and RIPK3 Nucleic Acid Sequences From Lagomorpha Species

In contrast to the majority of mammalian orders, lagomorpha only presents three annotated genomes: the European rabbit (*Oryctolagus cuniculus*, accession # GCA_000003625.1), the American Pika genome (*Ochotona princeps*, accession # GCA_014633375.1) and the plateau pika (*O. curzoniae*, accession # GCA_017591425.1). Given the importance of lagomorphs for this study, samples from different lagomorpha species were used to obtain the nucleic coding sequence from RIPK1. For that, RNA was extracted from tissues of *O. cuniculus cuniculus*, *O. cuniculus algirus*, *Lepus americanus*, *L. europaeus*, *L. timidus*, *L. granatensis*, *Sylvilagus floridanus*, *S. bachmanis*, *O. princeps* and *O. collaris* samples, using the Qiagen DNeasy Blood & Tissue kit (Qiagen, USA) following manufacturer's instructions. Synthesis of cDNA was achieved by using SuperScript III Reverse Transcriptase (Invitrogen, USA). Primers were designed according to the *RIPK1* transcript from *O. cuniculus* [Accession # XM_017350509.1] (Forward 5'-ATGTCTTTGGATGACATTAAAATG-3' and Reverse 5'-CTACTTCTGGCTGAGCTGTATC-3') and used to amplify the samples mentioned before. Phusion® High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland) was used in the PCR amplification and the conditions included an initial denaturation (98°C for 3min), 35 cycles of denaturation (98°C for 30s), annealing (60°C for 15s) and extension (72°C for 30s) followed a final extension (72°C for 5 min).

From our initial search, the *RIPK3* gene was not annotated in the European rabbit. However, after mapping the location of RIPK3 based on its location in *H. sapiens*, *M. musculus* and *O. curzoniae*, we were able to identify a partial RIPK3 sequence in the European rabbit genome that presented an early stop codon. To exclude potential artifacts that can mimic real gene-inactivating mutations, a forward (5'-ATGTCTTCTGTCAAATTGTGG-3') and a reverse (5'-ACTGCCTGCATCAGGATC-3') primer were designed based on the parcial *RIPK3* sequence and were used to amplify the same region in the genomes from *O. cuniculus cuniculus*, *S. floridanus*, *L. americanus*, *L. europaeus* and *L. saxatilis*. For that, genomic DNA was extracted using the Qiagen DNeasy Blood & Tissue kit (Qiagen, USA) according to the manufacturer's instructions. Phusion® High-Fidelity DNA Polymerase (Finnzymes, Espoo, Finland) was used in the PCR amplification and the conditions included an initial denaturation (98°C for 3min), 9 cycles of denaturation (98°C for 30s), annealing (66°C for 15s) and extension (72°C for 30s) followed by more 25 cycles of denaturation (98°C for 30s), annealing (61°C for 15s) and extension (72°C for 30s) and a final extension (72°C for 5min).

Amplicons sequencing from RIPK1 and RIPK3 was performed with the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit and according to manufacturer's protocol; reactions were cleaned with Sephadex™ (GE Healthcare Life Sciences, UK) and applied on an ABI PRISM 310 Genetic Analyser (Life Technologies, Applied Biosystems, Carlsbad, CA, USA). The obtained RIPK1 coding sequences and the partial RIPK3 sequences from the different Lagomorphs have been deposited in the GenBank database under the accession numbers that are shown in **Supplementary File 2**. All samples were supplied by CIBIO/InBIO, Vairão, Portugal and used in previous studies (15, 16). No animals were captured, handled, or killed specifically for the purpose of this study.

Detailed Analysis on the Cetacean Genomes

Briefly, NCBI gene annotations for the gene orthologues of *MLKL* and *RIPK3* were initially screened *via* PseudoChecker (pseudochecker.ciimar.up.pt), which evaluates the coding condition of a gene (17, 18). For each gene, a PseudoChecker analysis was run (default parameters), using the *Bos taurus* (cow) gene orthologue as a comparative input (NCBI Accession ID

regarding cow MLKL: XM_002694707.6; RIPK3: XM_024997365.1), as well as the genomic sequences encompassing the putative ORF of the orthologous counterpart of each target species, directly exported from the NCBI genome browser. Through PseudoIndex, a built-in assistant metric, we quickly assessed the erosion status of the tested genes on a discrete scale ranging from 0 (coding) to 5 (pseudogenized) (17). Subsequent manual annotation was performed by importing the previously collected genomic sequences into Geneious Prime 2020 software (www.geneious.com) (19) and determining each gene's CDS using as reference cow's *MLKL* and *RIPK3* orthologues sequences. In detail, per gene and species, using the built-in map to reference tool (highest sensitivity parameter selected), each (3' and 5' untranslated region-flanked) reference coding-exon was mapped against each target genomic sequence. Exons alignments were further screened for gene disruptive mutations, including in-frame premature stop codons, frameshift, and splice site mutations (any deviation from the consensus donor splice site GT/GC or the consensus acceptor splice site AG).

To inspect if the identified genetic lesions were not rendered as result of sequencing and/or genome assembly artifacts, we performed mutational validation (one per gene and species), resorting of raw genomic sequencing reads, retrieved from two independent genomic projects from the NCBI sequence read archive (SRA), when available. Explicitly, blastn searches were directed to the selected SRA projects, using the nucleotide sequence portion containing the selected mutation(s) as a query. The matching sequencing reads were downloaded into Geneious Prime 2020 (19) software and mapped against the manually annotated mutation (highest sensibility parameter selected), confirming, or not, the presence of the identified mutation.

Phylogenetic and Molecular Evolutionary Analyses

The complete dataset of RIPK1, RIPK3 and MLKL proteins was aligned in BioEdit Sequence Alignment Editor using Clustal W (20), followed by manual corrections when necessary. Amino acid alignments were then used to infer Maximum Likelihood (ML) phylogenetic trees using MEGA X (21), with the substitution models JTT+G+F, JTT+G and HKY+G+I, respectively; determined using ProtTest (22).

Given the fact that RIPK3 and MLKL proteins are highly divergent across the studied mammalian species, we decided not to perform any evolutionary analysis using these alignments. To look for signatures of natural selection operating in the RIPK1 alignment, we used HyPhy software implemented in the Datamonkey Web server (23), to detect codons under selection: the Single Likelihood Ancestor Counting (SLAC) model, the Fixed Effect Likelihood (FEL) method (24), the Random Effect Likelihood, the Mixed Effects Model of Evolution (MEME) (25) and Fast Unbiased Bayesian AppRoximation (FUBAR) (26) methods were used. To avoid a high false positive rate, codons with p-values <0.05 for SLAC, FEL and MEME models and a posterior probability >0.95 for FUBAR were accepted as candidates for selection. For a more conservative approach, only residues identified as being under positive selection in three or more ML methods were considered.

Analysis of VACV E3 Homologues

VACV E3 homologues encoded by different poxviruses (=11) were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/) and aligned in the BioEdit Sequence Alignment Editor using Clustal W (20), followed by manual corrections when necessary. Amino acid alignments of the representative E3-like proteins were used to generate schematic diagrams using the COBALT program from the NCBI database.

RESULTS

Necroptosis has a significant role in protecting cells against viral infection (1–3). However, despite recent advances to understand the molecular regulation of this unique pathway, it is still unclear whether the necroptotic cell death pathway acts as a universal cell death program in mammals. Here, taking advantage of genomic collection databases and the use of Leporid samples (see Methods section for more information), the goal is to better understand the molecular evolution of the necroptotic pathway in different mammalian lineages.

RIPK1 Protein Is Under Evolutionary Conservation in Mammals

Human RIPK1 is a multidomain protein composed of an N-terminal Ser/Thr kinase, a C-terminal death binding domain that mediates binding to DRs and an intermediate domain that includes a K377 ubiquitination site and an RHIM motif that binds to other RHIM-containing proteins (27). Due to the importance of RIPK1 as an adaptor molecule, previous phylogenetic analysis suggested that RIPK1 probably emerged in the ancestor of vertebrates (12). In accordance, our search detected the presence of *RIPK1* homologues in all the studied mammalian lineages (**Figure 2A** and **Supplementary Files 1–3**).

To look for evidence of potential selection pressures acting on the different domains of RIPK1 protein, we used the dataset of mammalian sequences mentioned above and implemented an ML approach, by using the Datamonkey software (see Methods section for more information). For most protein-coding genes, the rate between nonsynonymous and synonymous substitutions (dN/dS) is a measure of natural selection, with positive selection (dN/dS > 1) acting against the common genotype (28). In this study, we deduced ten sites that reflect strong positive selection pressure in RIPK1, while more than 200 amino acids were under negative selection (**Figure 2B** and **Supplementary File 4**). As seen in **Figure 2B**, residues identified as being under positive selection fall within or very close to the kinase domain (4 residues), the RHIM domain (5 residues) and the death domain (1 residue) of RIPK1 (residues under positive selection are marked as red circles). The N-terminal kinase domain is known to present several essential residues for phosphorylation and ubiquitination (Ser14/15, 20, 161 and 166 and Lys115 and

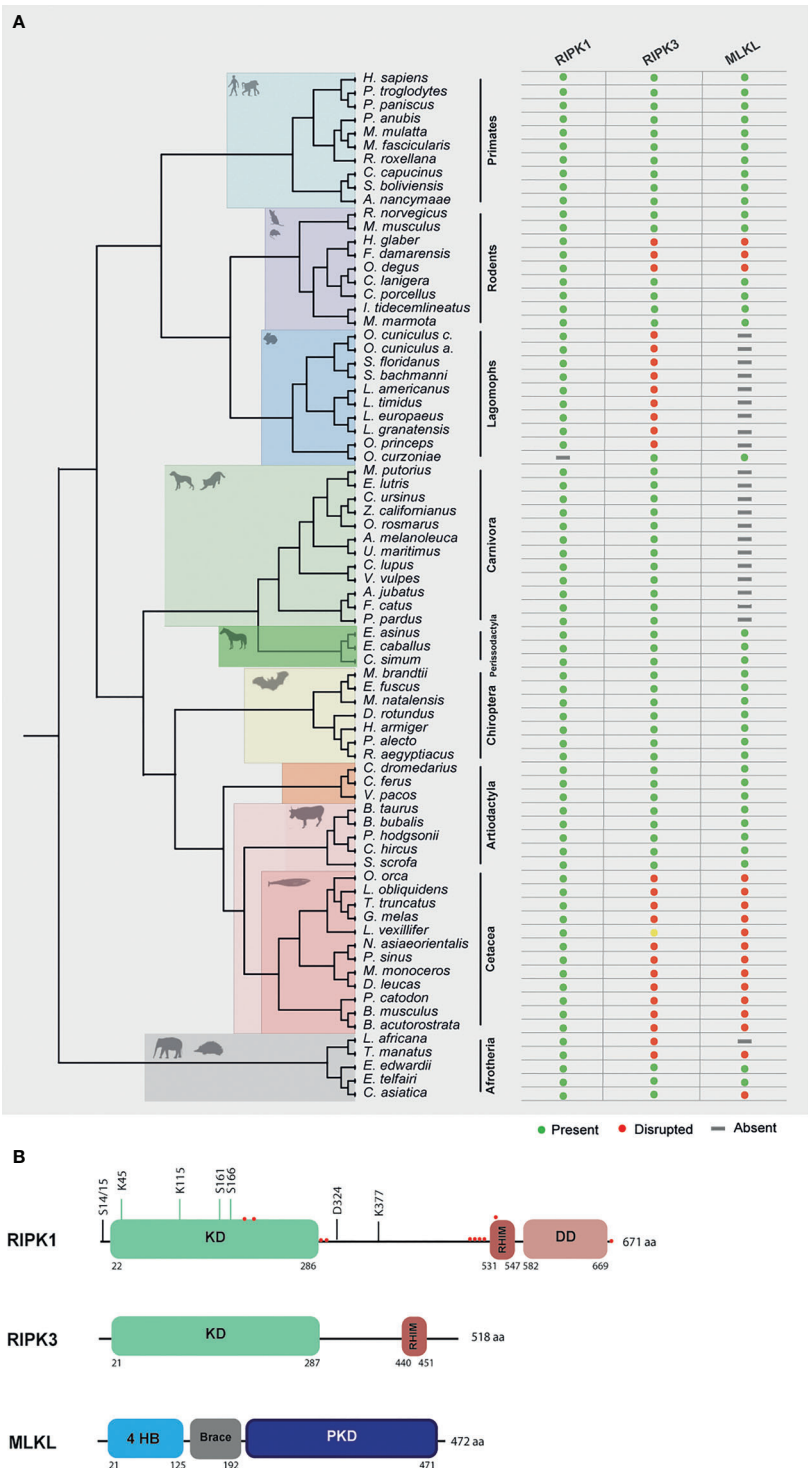


FIGURE 2 | Evolution of RIPK1, RIPK3 and MLKL in different mammalian lineages. **(A)** Phylogenetic tree showing the independent lineages that lost necroptotic core proteins (RIPK3 and MLKL) during evolution. Green circles represent genes that are present in the studied species, red circles represent genes that are disrupted, yellow circles represent genes that have incomplete assemblies and grey rectangles indicate that genes were not found in those species genomes. **(B)** A schematic diagram of RIPK1, RIPK3 and MLKL domains. RIPK1 contains an N-terminal kinase domain (KD), an intermediate domain with a RIP homotypic interaction motif (RHIM), and a C-terminal death domain (DD). The phosphorylation and ubiquitination sites are indicated above the RIPK1 domains. Red circles represent residues that are under positive selection. RIPK3 contains a KD and a RHIM domain. MLKL is composed of an N-terminal bundle four-helix bundle (4HD) domain that is regulated by the C-terminal pseudokinase domain (PKD).

163), regulation of necroptosis and RIPK1-dependent apoptosis (29). Interestingly, the codons under positive selection fall only at the end of the N-terminal domain. A considerable portion of the negatively selected sites fall in the critical regions for phosphorylation and ubiquitination (**Supplementary File 4**), suggesting that the beginning of the RIPK1 protein is under strong purifying selection. The same was also observed for the rapidly evolving sites of the RHIM domain, which were grouped only at the beginning of this domain (**Figure 2B**). It was previously shown that the RHIM domain has a crucial conserved core motif of 12 amino acids that resides at the end of this domain (30, 31). Indeed, changing four consecutive amino acids to alanine within this core region abrogates interaction between RHIM domains and, as a consequence, necroptosis (30, 31). Our findings that the positively selected residues did not overlap with the core motif of 12 amino acids further support the importance of the conservation of this region. Interestingly, in the RIPK1 death domain that is known to mediate homodimerization as well as heterodimerization with other DD-containing proteins, such as FADD, TNFR1 and Fas (29, 32), no positively selected sites were found, with most of the domain being under negative pressure (**Supplementary File 4**). Collectively, our results show signatures of positive selection occurring at the end of the N-terminal kinase domain and at the beginning of the RHIM domain of RIPK1 proteins. Interestingly, these residues do not overlap with domains known to be fundamental for RIPK1-dependent apoptosis and necroptosis, suggesting that these domains might be under evolutionary conservation and possibly functional constraint for the studied mammals.

Convergent Erosion of RIPK3 and MLKL in Mammalian Lineages

RIPK3 and MLKL form the core of the necroptotic machinery and both are, as a consequence, important for necroptosis induction in mice and humans downstream of PRRs and DRs (1, 4, 5). To further understand the evolutionary history of the necroptotic pathway in mammals, we performed detailed sequence and phylogenetic analyses for RIPK3 and MLKL homologous proteins (**Supplementary File 3**). Our screens for *RIPK3* and *MLKL* genes revealed evidence of pseudogenization in five mammalian lineages: order rodentia, lagomorpha, carnivora, cetacean and in the superorder afrotheria (**Figure 2A**). Given the fact that *MLKL* pseudogenes have been previously identified in carnivores, they will not be discussed in detail here (12).

Variation of the Necroptotic Pathway Within Afrotherian and Rodent Families

In afrotherian and rodent lineages, we found that the necroptotic pathway was missing in some species (**Figure 2A**). In rodents, two species from the Bathyergidae family and one species from the Octodontidae family presented early stop codons in both RIPK3 and MLKL, resulting in the disruption of necroptosis. In the naked mole-rat (*Heterocephalus glaber*), RIPK3 presented a premature stop codon in exon 6 resulting in a shorter version of this protein (**Supplementary File 5**). In the common degu (*Octodon degus*) and in the damaraland mole-rat (*Fukomys damarensis*), both RIPK3 and MLKL proteins presented several premature stop codons (**Supplementary File 5**). However, disruption of these proteins appears to have occurred in an independent way, rather than in a common ancestral. Given the fact that *RIPK3* and *MLKL* present signs of pseudogenization, it is expected that in the naked mole-rat, common degu and damaraland mole-rat necroptosis is disrupted. Our studies also revealed that species from Afrotherian families, including the African bush elephant (*Loxodonta africana*, family Elephantidae) and the Cape golden mole (Chrysochloris asiatica, family Chrysochloridae) presented early stop codons in RIPK3 and MLKL, respectively, while the West Indian manatee (*Trichechus manatus*, family Trichechidae) present early stop codons in both genes (**Supplementary File 5**). However, our results also show that the Cape elephant (*Elephantulus edwardii*, family Macroscelididae) and the lesser hedgehog tenrec (*Echinops telfairi*, family Tenrecidae) present intact copies of the *RIPK3* and *MLKL* genes, indicating that RIPK3 and MLKL were present in early stages of Afrotheria evolution, but must have been lost later in specific lineages, resulting in the existence of alternative modes of necroptosis inactivation.

The Necroptotic Pathway Is Disrupted in Lagomorphs

The Order Lagomorpha is divided into two families, Ochotonidae and Leporidae, which diverged around 30–35 Mya (33). While Ochotonidae is only composed of one extant genus, *Ochotona*, the Leporidae family includes 11 genera, including *Lepus*, *Sylvilagus* and *Oryctolagus* (34). Using the methods described previously, we were only able to identify *RIPK3* and *MLKL* transcripts for plateau Pika (*O. curzoniae*), while no *RIPK3* and *MLKL* transcripts were found for the European rabbit (*O. cuniculus*). For the American Pika (*O. princeps*), incomplete genome assemblies in the vicinity of the *RIPK3* and *MLKL* regions made retrieving the sequence of these genes impossible. By evaluating *RIPK3* gene from human and mouse and its genomic context, we were able derive a partial *RIPK3* nucleotide sequence from the European rabbit genome, which displays a frameshift mutation caused by the insertion of a single nucleotide. It is well known that accurately detecting gene-inactivation mutations in these alignments poses a number of challenges like, for example, sequencing errors and cases of assembly incompleteness. For this reason, we assessed the accuracy of our database prediction by sequencing that same genomic region in different Leporid species, representative of different genera (*Lepus*, *Sylvilagus* and *Oryctolagus*). From the obtained results, we confirmed the insertion of 1 nucleotide (+G, exon 3) not only in the European rabbit RIPK3, but also in species from genus *Lepus* and *Sylvilagus*, suggesting that disruption of *RIPK3* gene occurred in a common ancestral and was maintained throughout Leporid evolution (**Figure 3A**). During necroptosis, activated RIPK3 phosphorylates and activates MLKL, which results in its recruitment and oligomerization in the plasma membrane leading to rupture

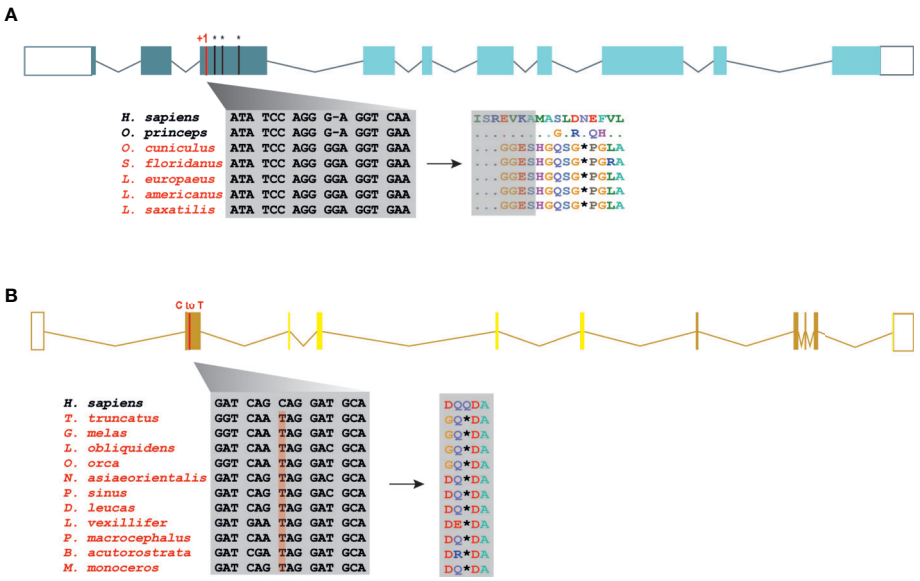


FIGURE 3 | Loss of *RIPK3* and *MLKL* genes in the steam lineage of Leporids and Cetaceans. **(A)** Genomic analysis of the first tree exons from Leporids. In Leporids, RIPK3 was lost as a result of a shared insertion (+G) in the third exon that resulted in the appearance of several premature stop codons. **(B)** A point mutation (C to T) in all the studied cetacean species indicates that MLKL inactivation occurred in Cetacea steam lineage. Moreover, 9 out of the 11 studied species (excluding *B. acutorostrata* and *M. Monoceros*) lost exon 2, 3, 4 and 5 throughout evolution (represented by faint yellow). Premature stop codons are represented by an asterisk (*).

and cell death (1, 4). Interestingly, and despite all of our efforts, no *MLKL* transcripts or MLKL protein accumulation were found in any of the studied Lagomorphs (data not shown). Detailed analysis of the upstream and downstream MLKL flanking genes in both human and mouse genomes reveal that MLKL resides between the ring finger and WD repeat domain 3 protein (*RFWD3*) and fatty acid 2-hydroxylase (*FA2H*) genes (**Figure 4**). Accordingly, although there are no gaps or incomplete genomic assemblies surrounding that region in the European rabbit genome, we were not able to retrieve a complete or partial *MLKL* gene, suggesting once more that this gene is not present in these mammals. Together, our results suggest that the studied leporid species are deficient in the core proteins of the necroptotic pathway, and that RIPK3 inactivation occurred at the stem Leporid branch and was maintained during evolution.

Inactivation of Necroptosis Components in Cetacea

Modern Cetacea comprises Mysticete (or baleen whales) and Odontocete (or toothed whales) and are the most specialized and diversified group of mammals (35). Comparative analysis of cetacean genomes has already provided important insights into the unique cetacean traits and aquatic specializations (13, 36, 37). For our screen, Odontocetes were represented by 12 species belonging to five different families (Delphinidae, Phocoenidae, Monodontidae, Lipotidae and Physeteridae), and Mysticetes were represented by the common minke whale from Balaenopteridae family. In Cetacean species, the disruption of *RIPK3* occurred at different positions depending on the studied species: a frameshift mutation was identified in exon 6 of two Delphinidae species, as well as in exon 8 of two Phocoenidae species, one Monodontidae species and one Balaenopteridae species. There was also evidence of two species (one species from Monodontidae and another from Lipotidae families) with RIPK3 pseudogenes based on the presence of a stop codon located in exon 2 (**Supplementary File 6**). Interestingly, while Cetacean RIPK3 inactivation appears to be a result of different mutations depending on the studied species, our results show a shared mutation in the exon 1 of *MLKL* in all cetacean species (**Figure 3B**). Moreover, this premature stop codon leads to the absence of exon 2, 3, 4 and 5 in most cetacean species, which very likely results in this gene inactivation. Given the presence of an inactivating mutation that is shared between mysticetes and odontocetes, the most parsimonious hypothesis suggests that they occurred before the split of these two clades in the common ancestral branch of Cetacea.

Diversity Among the Poxvirus Encoded E3-Like Necroptosis Antagonists

Previously, it was shown that the N-terminus of VACV E3 competes with ZBP1 for binding to virus-induced Z-nucleic acid, being a key component to inhibit the action of IFN and induction of necroptosis (38) (**Figure 1**). E3-like encoded proteins are composed of a carboxy (C)-terminal double-stranded RNA binding domain (dsRNA-BD) and an amino (N)-terminal Z-nucleic acid-binding domain (zNA-BD) (39, 40) (**Figure 5A**). Given the importance of the N-terminus region from VACV E3 protein against virus infection, we hypothesize that poxviruses that lack this region in their E3 homologs can still successfully replicate in their natural host

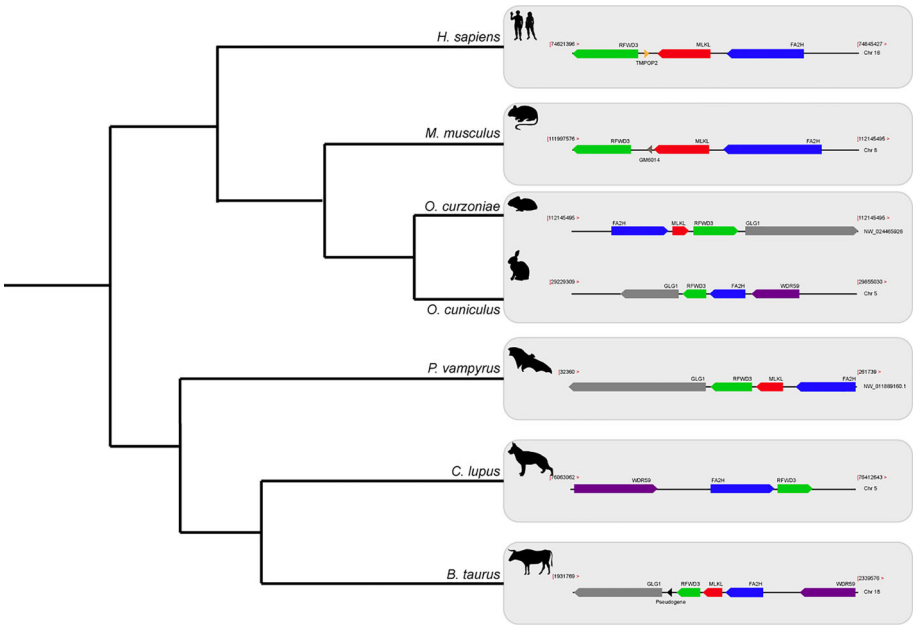


FIGURE 4 | Gene synteny of the genome regions containing MLKL gene in different mammals. Genomic regions containing the *MLKL* gene or its flanking genes in *H. sapiens*, *M. musculus*, *O. curzoniae*, *O. cuniculus*, *P. vampyrus*, *C. lupus* and *B. taurus*. Horizontal lines indicate chromosome fragments and coloured arrows identify genes and their orientation in the genome. Orthologous genes are indicated in the same colour and their names are indicated above/below. Black arrows indicate the presence of pseudogenes. RFWD3, ring finger and WD repeat domain 3; MLKL, mixed-lineage kinase domain like; FA2H, fatty acid 2-hydroxylase; GLG1, golgi glycoprotein 1; WDR59, WD repeat domain 59; *TMPOP2*, thymopoietin pseudogene; *GM6014*, ubiquitin-40S ribosomal protein S27a pseudogene; *LOC788457*, translationally-controlled 1 pseudogene.

because of a compromised necroptotic pathway. Among the E3L proteins, E3 from VACV is the best studied protein. However, E3 homologs can be found in orthopoxviruses, clade II poxviruses and parapoxviruses (39, 40). Recently, the genome characterization of CePV-TA identified two novel E3L homologs: CePV-TA-20 and CePV-TA-21 (41).

Our analysis on 11 different E3 homologues revealed that these are highly divergent: while CPXV 069 (Cowpoxvirus E3 homolog) and TATV 060 (Tateropoxvirus E3 homolog) presented identities of >90% to VACV E3, E3L homologs from poxviruses like the Deerpoxvirus, Sheeppoxvirus and Yaba monkey poxvirus presented less than 40% identity (**Figure 5B**). Analysis of the two newly identified E3 homologs from CePV-TA shows that both present low identity to VACV E3, with CePV-TA-20 and CePV-TA-21 proteins only presenting sequence identity of 37% and 34%, respectively (**Figure 5B**). It is known that at the amino acid level, the C-terminal of E3-like proteins display a higher level of sequence similarity than the N-terminal domain (42). Accordingly, the dsRNA-BD domain from CePV-TA-20 and CePV-TA-21 proteins also display a higher level of sequence similarity compared to other E3 homologs (**Figure 5C**), suggesting that in CeTV this domain might also target conserved antiviral dsRNA-activated pathways. Similar to what is observed for Monkeypox virus (MPXV) and Myxoma virus (MYXV) E3 homologues, both E3 homologs from CePV-TA present incomplete or disrupted N-terminal zNA-BDs. As shown in **Figure 5C**, CePV-TA-20 is missing 20 amino acids in its N-terminal domain. However, this region still retains the conserved LY and PPXW motifs, as well the basic KKCINR motif (**Figure 5C**) that are known to bind with Z-NA (40). Interestingly, MPXV F3 protein, lacking 37 amino acids from the N-terminal domain, is not able to compete with ZBP1 and inhibit sensing (unpublished data), even though it retains the key residues important for binding to Z-NA (**Figure 5C**). While CePV-TA-20 and F3 proteins contain an incomplete zNA-BD, M029 and CePV-TA-21 proteins are missing most of their N-terminal zNA-BD (**Figures 5B, C**), suggesting a total inactivation of this domain and a loss of function regarding the inhibition of ZBP1-dependent necroptosis. Overall, our results show that the novel CePV-TA presents two E3L homologues that, like E3L homologues from MPXV and MYXV, present incomplete or disrupted N-terminal zNA-BD. The presence of an incomplete or disrupted zNA-BD in E3L homologues highly suggests that these proteins cannot fully compete with ZBP1 to inhibit necroptosis induction. However, further studies will be necessary to fully comprehend the action of these proteins regarding complete necroptosis inhibition.

DISCUSSION

Necroptosis is an inflammatory form of cell death that is mediated by RIPK3 and MLKL and provides an extra defense

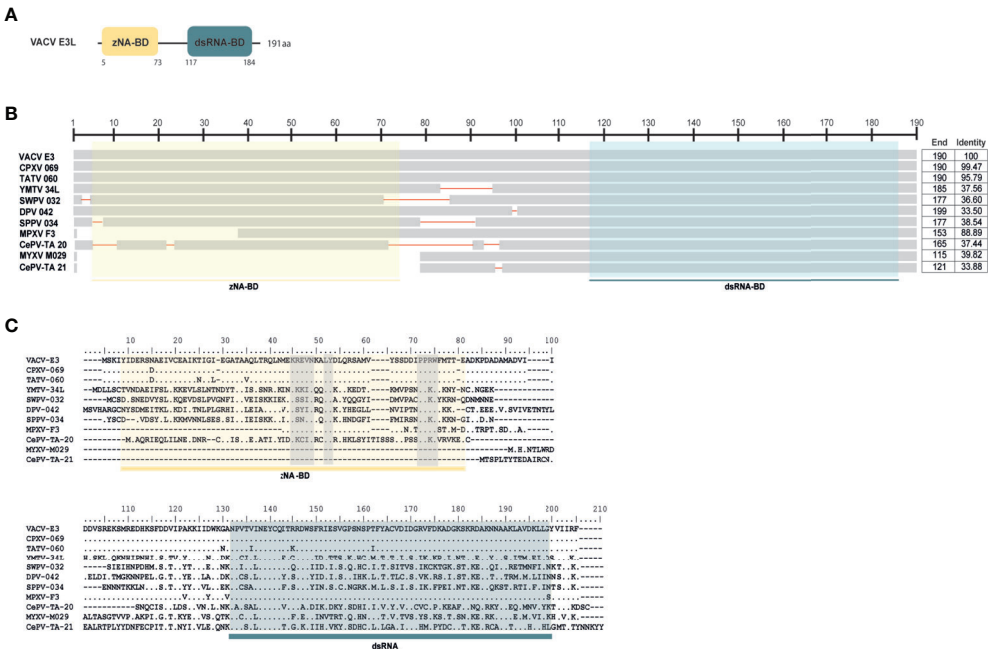


FIGURE 5 | Protein sequence alignment of E3L proteins. **(A)** Schematic diagram of VACV E3 protein binding domains: yellow box represents the zNA-BD and blue box represents the dsRNA-BD. The same color scheme is used in **(B, C)**. **(B)** E3L homologues from 11 poxviruses (VACV E3, Cowpoxvirus (CPXV) 069, Tateropoxvirus (TATV) 060, Yaba monkey tumor virus (YMTV) 034, swinepoxvirus (SWPV) 34L, deerpoxvirus (DPV) 042, sheeppoxvirus (SPPV) 034, MPXV F3, MYXV M029 and CePV-TA 20 and 21) were used to perform a schematic alignment using COBALT program from the NCBI platform. Length of each E3L homologue as well as their identity to VACV E3 proteins are shown in the column to the right. **(C)** Amino acid sequence comparison of 11 different members of the E3L family including VACV E3, TATV 060, YMTV 034, SWPV 34L, DPV 042, SSPPV 034, MPXV F3 and MYXV M029 and CePV-TA 20 and 21. Conserved areas known to bind to zNA are shown in grey boxes.

mechanism against pathogen infection, facilitating the elimination of virus-infected cells before the production of progeny virions (1–3). Given the crucial role of necroptosis in the innate immune response of humans and mice (43, 44), it was broadly accepted that this pathway was ubiquitous in mammals. Our results from 67 species across nine mammalian lineages provides the first comprehensive picture of the molecular evolution of necroptosis in mammals. We show that while RIPK1 is under evolutionary conservation, RIPK3 and MLKL are poorly conserved in lineages that evolved separately over the course of evolution.

A detailed analysis of RIPK3 and MLKL in mammals reveals a complex pattern where lagomorphs, cetaceans, carnivores and species from rodent and afrotheria lineages separately lost key components of the necroptotic pathway (**Figure 2A**). The order lagomorpha includes two big families, Ochotonidae and Leporidae (33). The presence of the same frameshift mutation in Leporid species (+G, **Figure 3A**) suggests that the disruption of the necroptotic pathway occurred early, but only after the bifurcation between Ochotona and Leporid given that the plateau Pika presents intact RIPK3 and MLKL proteins (**Figure 2A**). Despite all efforts, and despite the complete genome assembly surrounding the MLKL flanking genes in the European rabbit genome, no partial or complete MLKL gene was found, indicating that this gene is deleted in the European rabbit genome and possibly in the remaining Leporid species (**Figure 4**). The core genes of the necroptotic pathway also presented premature stop codons in cetaceans. In the studied cetaceans, the MLKL gene presented a common stop codon in the first exon, resulting in the inactivation of this gene (**Figure 3B**). Again, the presence of similar patterns of pseudogenization in RIPK3 or MLKL genes within species of the same order infer that disruption of these genes occurred before their diversification and was maintained throughout evolution. On the other hand, *RIPK3* disruption in cetaceans appears to be the result of insertions or deletions that are not shared between closely related species, but rather specific to each species (**Supplementary File 6**), suggesting that these disrupting mutations occurred later in evolution when compared to MLKL. The addition of some rodents as well as afrotheria species to the list of mammals that have disrupted necroptotic pathways, raises the possibility that other closely related species might have lost this pathway after the diversification of these lineages. It is currently believed that activation of the RIPK3 and recruitment of MLKL are critical steps during necroptosis. For example, deleting either RIPK3 or MLKL can lead to the suppression of skin and liver inflammation in mice (45, 46). Moreover, when mice are treated intravenously with a high-doses of TNF, there appears to be no differences between RIPK3-deficient and MLKL-deficient mice (5), substantiating the premise that MLKL follows RIPK3 in the necroptotic signalling. However, this appears not to be the case for all

necroptotic cell death responses, as different studies revealed alternative pathways for MLKL and RIPK3-dependent programmed necrosis that are executed in the absence of RIPK3 or MLKL, respectively (47, 48). For example, Inositol Phosphate metabolites have been previously shown to be required for human MLKL activation by directly acting on the N-terminal 4HB to dislodge an inhibitory region, revealing that MLKL-mediated necroptosis can also be controlled by a metabolite (49). To date, there are no studies suggesting that RIPK3/MLKL double-knockout mice are still able to induce necroptosis, which indicates that species that have disrupted RIPK3 and MLKL lost the necroptotic pathway throughout evolution.

The loss of function of *RIPK3* and *MLKL* in independently evolving lineages (convergent evolution) indicates that gene loss is an important evolutionary mechanism for phenotypic change in these animals and may contribute to similar adaptations. Even though it would be expected that loss of genes is maladaptive, gene loss can be beneficial by providing an evolutionary mechanism for adaptations (50). In fact, if the loss of an existing gene would increase fitness by making a species better adapted to the environment that surrounds it, then gene loss would be an easy solution to an evolutionary problem (50). Necroptosis contributes to innate immunity as a pathogen clearance mechanism (1). However, contrary to apoptosis, in which several highly immunogenic intracellular proteins are sequestered in the dead cell, necroptosis releases DAMPs in the surrounding tissue that promote strong inflammatory responses and result in the attraction of different types of immune cells to the site of infection (51). Studies in mouse models have provided strong evidence that necroptosis is implicated in several inflammatory neurodegenerative diseases, including multiple sclerosis and amyotrophic lateral sclerosis (52, 53). Mouse-model experiments identified keratinocyte necroptosis as a trigger of skin inflammation (54) and a correlation between necroptosis and intestinal inflammation has also been established (55, 56). Thus, while necroptosis might mediate host defense, its inhibition in certain contexts may lessen disease severity. It is known that excessive inflammation can promote cancer cell growth and metastasis (57). Thus, it is possible that a pro-inflammatory cell death like necroptosis might promote metastasis and thus, inhibition of this pathway might represent an advantage for regulation of cancer cell growth. Intriguingly, some of the species that are lacking the core necroptotic machinery are known to resist cancer. That is the case for cetaceans, the naked mole-rat and african elephants (58–60). It is also possible that selection against necroptosis in different mammalian lineages could have been driven by different factors depending on the environment or conditions. For example, it was previously suggested that the absence of MLKL in Carnivores reflected a microbe-rich and virus-containing diet of raw meat, causing evolutionary counter-selection against necroptosis (12). Nevertheless, the absence of the necroptotic pathway in independently evolving lineages suggest that the deregulation of this pathway was detrimental for the host organism, which ultimately drove selection against the presence of RIPK3 and MLKL.

As many other viruses, poxviruses express immunomodulatory and host-range factors important for the suppression and evasion of the host innate and adaptive antiviral responses (61, 62). VACV protein E3 not only sequester dsRNA through their dsRNA-BD limiting the activation of the innate immune system against the virus infection, but also inhibit the IFN-induced dsRNA dependent protein kinase (PKR), known to be a crucial component of the host innate immunity against viral infection, replication, and spread (63, 64). Our results show that the dsRNA-BD of distant E3L proteins present high levels of sequence similarity (**Figures 5B, C**), which is consistent with the ability of this domain to target conserved pathways present in different hosts. Although the dsRNA sequestration functions of the E3 C-terminal have been clear for decades (20, 65), the IFN sensitivity of VACV E3 N-terminal deletion mutants remained unresolved for a long time. Recently, strong evidence showed that the E3 N-terminal domain competes with ZBP1 to prevent ZBP1-dependent activation of RIPK3 and consequent necroptosis (38, 66). The model proposed by the authors suggests that during WT-VACV infection, the zNA-BD of E3 binds to VACV-induced Z-form nucleic acid and masks it, preventing sensing by ZBP1 and further RIPK3 necroptosis induction (38, 66). However, it is interesting that poxviruses like MPXV, MYXV and CePV-TA have E3L homologs that present a complete dsRNA-BD but not zNA-BD (**Figure 5B**). In VACV-E3 Δ83N–infected cells (mutant lacking the first 83 aa corresponding to the zNA-BD), the absence of the zNA-BD facilitates ZBP1 to sense VACV-induced PAMPs and initiate necroptosis induction (38). Therefore, it is expected that E3L homologues that lack N-terminal zNA domains, like CePV-TA-21 and M029, cannot prevent Z-form nucleic acid sensing triggering necroptosis induction and early abortion of viral replication. Like CePV-TA-20, F3 protein is also missing several amino acids in the N-terminal region and presents high conservation in areas that are known to bind to zNA (**Figure 5C**). Nevertheless, F3 protein seems to have lost the ability to compete with ZBP1 and inhibit sensing (unpublished data). It was previously shown that the N-terminus of VACV E3 is necessary for IFN resistance in JC cells since VACV-E3Δ37N (mutant mimicking MPXV E3 zNA-BD) did not initiate DNA replication (67). However, MPXV was able to replicate efficiently in the same cells, despite having a partial N-terminal zNA-BD, suggesting that the predicted binding to z-form nucleic acid was intragenic and downstream of z-NA sensing, rather than related to the ability of F3 zNA-BD to mask z-form nucleic acid (67).

Interestingly, inactivation of necroptosis in Lagomorphs and Cetaceans seems to correlate with the absence of the E3L zNA-BD in their naturally infecting poxviruses, namely leporipoxviruses (MYXV and SFV) and cetaceanpoxviruses (CePV-TA), respectively. Monkeypox is a viral zoonosis endemic to central and western Africa areas where African rope squirrels and other rodents are likely reservoir hosts (68). Interestingly, the absence of a functional N-terminus in MPXV F3 protein also seems to correlate with the fact that some rodents

appear incapable of undergoing necroptosis. Like MPXV, leporipoxviruses and CePV-TA pathogenesis are restricted to only certain species and have little or no pathogenesis capability in all others (41, 67, 69, 70). Infection of the same host over hundreds of years or even millennia may drive the evolution of each virus to rapidly evolve to a fitness peak in a given host environment. Previous niche-filling models (71–73) emphasize the role of host interactions in shaping virus evolution. According to these models, as hosts diversify and speciate over longer evolutionary periods, viral host factors that aim to counter the host antiviral functions are subject to continuous changes. Indeed, it is known that genes associated with host antiviral mechanisms present high evolutionary rates and are often under positive selection (15, 74, 75). Here, we suggest that during the evolution of these poxviruses, the loss of the zNA-BD did not present a disadvantage in the host organism; therefore, this trait was maintained, which reflects how these viruses adapt as their niche changed.

CONCLUSION

The disruption of necroptosis in independently evolving lineages suggests a convergent evolutionary loss of this pathway, probably reflecting an important selective mechanism for phenotypic change. Interestingly, we also found a strong correlation between the disruption of necroptosis in leporids and cetaceans and the absence of the E3L zNA-BD (responsible for necroptosis inhibition) in their naturally infecting poxviruses as in the case of MYXV and CePV-TA, respectively. Overall, our study provides the first comprehensive picture of the molecular evolution of necroptosis in mammals, highlighting the importance of gene/pathway loss for the process of species adaptation and suggesting that it is a true pathogen-response pathway that is not required for normal mammalian development. Moreover, this study sheds some light on a co-evolutionary relationship between poxviruses and their hosts, emphasizing the role of host adaptation in shaping virus evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

AÁ-P, BJ, LC, GM, MR and PE contributed to conception and design of the study. AÁ-P, FN and LA were responsible for the acquisition and statistical analysis of the data. AÁ-P, LC, LA, BJ and PE contributed for the interpretation of data. AÁ-P wrote the first draft of manuscript. MR, GM and PE acquired funding to support the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Foundation for Science and Technology (FCT), which supported the doctoral fellowship of AÁ-P (ref.SFRH/BD/128752/2017), the investigator grant of PJE (IF/00376/2015) and the project UIDB/50027/2020 (Base). This article is also a result of the project NORTE-01-0145-FEDER-000007, supported by Norte Portugal Regional Operational Programme (NORTE2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). GM and MMR's research is supported by the National Institute of Health (NIH) grants R01AI080607 and R01AI148302.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.747737/full#supplementary-material

REFERENCES

1. Xia X, Lei L, Wang S, Hu J, Zhang G. Necroptosis and its Role in Infectious Diseases. *Apoptosis* (2020) 25:169–78. doi: 10.1007/s10495-019-01589-x
2. Nogusa S, Thapa RJ, Dillon CP, Liedmann S, Oguin TH3rd, Ingram JP, et al. RIPK3 Activates Parallel Pathways of MLKL-Driven Necroptosis and FADD-Mediated Apoptosis to Protect Against Influenza A Virus. *Cell Host Microbe* (2016) 20:13–24. doi: 10.1016/j.chom.2016.05.011
3. Upton JW, Chan FK-M. Staying Alive: Cell Death in Antiviral Immunity. *Mol Cell* (2014) 54:273–80. doi: 10.1016/j.molcel.2014.01.027
4. Sun L, Wang H, Wang Z, He S, Chen S, Liao D, et al. Mixed Lineage Kinase Domain-Like Protein Mediates Necrosis Signaling Downstream of RIP3 Kinase. *Cell* (2012) 148:213–27. doi: 10.1016/j.cell.2011.11.031
5. Moerke C, Bleibaum F, Kunzendorf U, Krautwald S. Combined Knockout of RIPK3 and MLKL Reveals Unexpected Outcome in Tissue Injury and Inflammation. *Front Cell Dev Biol* (2019) 7:19. doi: 10.3389/fcell.2019.00019
6. He S, Wang L, Miao L, Wang T, Du F, Zhao L, et al. Receptor Interacting Protein Kinase-3 Determines Cellular Necrotic Response to TNF-Alpha. *Cell* (2009) 137:1100–11. doi: 10.1016/j.cell.2009.05.021
7. Zhang D-W, Shao J, Lin J, Zhang N, Lu B-J, Lin S-C, et al. RIP3, an Energy Metabolism Regulator That Switches TNF-Induced Cell Death From Apoptosis to Necrosis. *Science* (2009) 325:332–6. doi: 10.1126/science.1172308
8. Upton JW, Kaiser WJ, Mocarski ES. Virus Inhibition of RIP3-Dependent Necrosis. *Cell Host Microbe* (2010) 7:302–13. doi: 10.1016/j.chom.2010.03.006
9. Upton JW, Kaiser WJ, Mocarski ES. DAI/ZBP1/DLM-1 Complexes With RIP3 to Mediate Virus-Induced Programmed Necrosis That Is Targeted by Murine Cytomegalovirus vIRA. *Cell Host Microbe* (2012) 11:290–7. doi: 10.1016/j.chom.2012.01.016
10. He S, Liang Y, Shao F, Wang X. Toll-Like Receptors Activate Programmed Necrosis in Macrophages Through a Receptor-Interacting Kinase-3-Mediated Pathway. *Proc Natl Acad Sci USA* (2011) 108:20054–9. doi: 10.1073/pnas.1116302108
11. Kaiser WJ, Upton JW, Mocarski ES. Viral Modulation of Programmed Necrosis. *Curr Opin Virol* (2013) 3:296–306. doi: 10.1016/j.coviro.2013.05.019
12. Dondelinger Y, Hulpiau P, Saeys Y, Bertrand MJM, Vandenabeele P. An Evolutionary Perspective on the Necroptotic Pathway. *Trends Cell Biol* (2016) 26:721–32. doi: 10.1016/j.tcb.2016.06.004
13. Sharma V, Hecker N, Roscito JG, Foerster L, Langer BE, Hiller M. A Genomics Approach Reveals Insights Into the Importance of Gene Losses

for Mammalian Adaptations. *Nat Commun* (2018) 9:1215. doi: 10.1038/s41467-018-03667-1
14. Águeda-Pinto A, Castro LFC, Esteves PJ. The Evolution of S100A7: An Unusual Gene Expansion in Myotis Bats. *BMC Evol Biol* (2019) 19:102. doi: 10.1186/s12862-019-1433-0
15. Águeda-Pinto A, Lemos de Matos A, Pinheiro A, Neves F, de Sousa-Pereira P, Esteves PJ. Not So Unique to Primates: The Independent Adaptive Evolution of TRIM5 in Lagomorpha Lineage. *PloS One* (2019) 14:e0226202. doi: 10.1371/journal.pone.0226202
16. de Matos AL, van der Loo W, Areal H, Lanning DK, Esteves PJ. Study of Sylvilagus Rabbit TRIM5α Species-Specific Domain: How Ancient Endoviruses Could Have Shaped the Antiviral Repertoire in Lagomorpha. *BMC Evol Biol* (2011) 11:294. doi: 10.1186/1471-2148-11-294
17. Alves LQ, Ruivo R, Fonseca MM, Lopes-Marques M, Ribeiro P, Castro LFC. PseudoChecker: An Integrated Online Platform for Gene Inactivation Inference. *Nucleic Acids Res* (2020) 48:W321–31. doi: 10.1093/nar/gkaa408
18. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE V2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol Biol Evol* (2018) 35:2582–4. doi: 10.1093/molbev/msy159
19. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics* (2012) 28:1647–9. doi: 10.1093/bioinformatics/bts199
20. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res* (1994) 22:4673–80. doi: 10.1093/nar/22.22.4673
21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis Across Computing Platforms. *Mol Biol Evol* (2018) 35:1547–9. doi: 10.1093/molbev/msy096
22. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: Fast Selection of Best-Fit Models of Protein Evolution. *Bioinformatics* (2011) 27:1164–5. doi: 10.1093/bioinformatics/btr088
23. Pond SLK, Frost SDW. Datamonkey: Rapid Detection of Selective Pressure on Individual Sites of Codon Alignments. *Bioinformatics* (2005) 21:2531–3. doi: 10.1093/bioinformatics/bti320
24. Kosakovsky Pond SL, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol* (2005) 22:1208–22. doi: 10.1093/molbev/msi105
25. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PloS Genet* (2012) 8:e1002764. doi: 10.1371/journal.pgen.1002764
26. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, et al. FUBAR: A Fast, Unconstrained Bayesian Approximation for Inferring Selection. *Mol Biol Evol* (2013) 30:1196–205. doi: 10.1093/molbev/mst030
27. He S, Wang X. RIP Kinases as Modulators of Inflammation and Immunity. *Nat Immunol* (2018) 19:912–22. doi: 10.1038/s41590-018-0188-x
28. Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, et al. Patterns of Positive Selection in Six Mammalian Genomes. *PloS Genet* (2008) 4:e1000144. doi: 10.1371/journal.pgen.1000144
29. Mifflin L, Ofengeim D, Yuan J. Receptor-Interacting Protein Kinase 1 (RIPK1) as a Therapeutic Target. *Nat Rev Drug Discovery* (2020) 19:553–71. doi: 10.1038/s41573-020-0071-y
30. Sun X, Yin J, Starovasnik MA, Fairbrother WJ, Dixit VM. Identification of a Novel Homotypic Interaction Motif Required for the Phosphorylation of Receptor-Interacting Protein (RIP) by RIP3. *J Biol Chem* (2002) 277:9505–11. doi: 10.1074/jbc.M109488200
31. Li J, McQuade T, Siemer AB, Napetschnig J, Moriwaki K, Hsiao Y-S, et al. The RIP1/RIP3 Necrosome Forms a Functional Amyloid Signaling Complex Required for Programmed Necrosis. *Cell* (2012) 150:339–50. doi: 10.1016/j.cell.2012.06.019
32. Meng H, Liu Z, Li X, Wang H, Jin T, Wu G, et al. Death-Domain Dimerization-Mediated Activation of RIPK1 Controls Necroptosis and RIPK1-Dependent Apoptosis. *Proc Natl Acad Sci USA* (2018) 115:E2001–9. doi: 10.1073/pnas.1722013115
33. Melo-Ferreira J, Lemos de Matos A, Areal H, Lissovsky AA, Carneiro M, Esteves PJ. The Phylogeny of Pikas (Ochotona) Inferred From a Multilocus Coalescent Approach. *Mol Phylogenet Evol* (2015) 84:240–4. doi: 10.1016/j.ympev.2015.01.004
34. Ge D, Wen Z, Xia L, Zhang Z, Erbajeva M, Huang C, et al. Evolutionary History of Lagomorphs in Response to Global Environmental Change. *PloS One* (2013) 8:e59668. doi: 10.1371/journal.pone.0059668
35. McGowen MR, Tsagkogeorga G, Álvarez-Carretero S, Dos Reis M, Struebig M, Deaville R, et al. Phylogenomic Resolution of the Cetacean Tree of Life Using Target Sequence Capture. *Syst Biol* (2020) 69:479–501. doi: 10.1093/sysbio/syz068
36. Huelsmann M, Hecker N, Springer MS, Gatesy J, Sharma V, Hiller M. Genes Lost During the Transition From Land to Water in Cetaceans Highlight Genomic Changes Associated With Aquatic Adaptations. *Sci Adv* (2019) 5:eaaw6671. doi: 10.1126/sciadv.aaw6671
37. Kawasaki K, Mikami M, Goto M, Shindo J, Amano M, Ishiyama M. The Evolution of Unusually Small Amelogenin Genes in Cetaceans; Pseudogenization, X-Y Gene Conversion, and Feeding Strategy. *J Mol Evol* (2020) 88:122–35. doi: 10.1007/s00239-019-09917-0
38. Koehler H, Cotsmire S, Langland J, Kibler KV, Kalman D, Upton JW, et al. Inhibition of DAI-Dependent Necroptosis by the Z-DNA Binding Domain of the Vaccinia Virus Innate Immune Evasion Protein, E3. *Proc Natl Acad Sci USA* (2017) 114:11506–11. doi: 10.1073/pnas.1700999114
39. Kim Y-G, Lowenhaupt K, Oh D-B, Kim KK, Rich A. Evidence That Vaccinia Virulence Factor E3L Binds to Z-DNA *In Vivo*: Implications for Development of a Therapy for Poxvirus Infection. *Proc Natl Acad Sci USA* (2004) 101:1514–8. doi: 10.1073/pnas.0308260100
40. Kim Y-G, Muralinath M, Brandt T, Pearcy M, Hauns K, Lowenhaupt K, et al. A Role for Z-DNA Binding in Vaccinia Virus Pathogenesis. *Proc Natl Acad Sci USA* (2003) 100:6974–9. doi: 10.1073/pnas.0431131100
41. Rodrigues TCS, Subramaniam K, Varsani A, McFadden G, Schaefer AM, Bossart GD, et al. Genome Characterization of Cetaceanpox Virus From a Managed Indo-Pacific Bottlenose Dolphin (Tursiops Aduncus). *Virus Res* (2020) 278:197861. doi: 10.1016/j.virusres.2020.197861
42. Rahman MM, McFadden G. Myxoma Virus-Encoded Host Range Protein M029: A Multifunctional Antagonist Targeting Multiple Host Antiviral and Innate Immune Pathways. *Vaccines (Basel)* (2020) 8:2–3. doi: 10.3390/vaccines8020244
43. Nailwal H, Chan FK-M. Necroptosis in Anti-Viral Inflammation. *Cell Death Differ* (2019) 26:4–13. doi: 10.1038/s41418-018-0172-x
44. Orzalli MH, Kagan JC. Apoptosis and Necroptosis as Host Defense Strategies to Prevent Viral Infection. *Trends Cell Biol* (2017) 27:800–9. doi: 10.1016/j.tcb.2017.05.007
45. Dannappel M, Vlantis K, Kumari S, Polykratis A, Kim C, Wachsmuth L, et al. RIPK1 Maintains Epithelial Homeostasis by Inhibiting Apoptosis and Necroptosis. *Nature* (2014) 513:90–4. doi: 10.1038/nature13608
46. Rickard JA, Anderton H, Etemadi N, Nachbur U, Darding M, Peltzer N, et al. TNFR1-Dependent Cell Death Drives Inflammation in Sharpin-Deficient Mice. *Elife* (2014) 3:7–8. doi: 10.7554/eLife.03464
47. Zhang T, Zhang Y, Cui M, Jin L, Wang Y, Lv F, et al. CaMKII is a RIP3 Substrate Mediating Ischemia- and Oxidative Stress-Induced Myocardial Necroptosis. *Nat Med* (2016) 22:175–82. doi: 10.1038/nm.4017
48. Günther C, He G-W, Kremer AE, Murphy JM, Petrie EJ, Amann K, et al. The Pseudokinase MLKL Mediates Programmed Hepatocellular Necrosis Independently of RIPK3 During Hepatitis. *J Clin Invest* (2016) 126:4346–60. doi: 10.1172/JCI87545
49. Dovey CM, Diep J, Clarke BP, Hal AT, McNamara DE, Guo H, et al. MLKL Requires the Inositol Phosphate Code to Execute Necroptosis. *Mol Cell* (2018) 70:936–48.e7. doi: 10.1016/j.molcel.2018.05.010
50. Albalat R, Cañestro C. Evolution by Gene Loss. *Nat Rev Genet* (2016) 17:379–91. doi: 10.1038/nrg.2016.39
51. Kaczmarek A, Vandenabeele P, Krysko DV. Necroptosis: The Release of Damage-Associated Molecular Patterns and Its Physiological Relevance. *Immunity* (2013) 38:209–23. doi: 10.1016/j.immuni.2013.02.003
52. Ito Y, Ofengeim D, Najafov A, Das S, Saberi S, Li Y, et al. RIPK1 Mediates Axonal Degeneration by Promoting Inflammation and Necroptosis in ALS. *Science* (2016) 353:603–8. doi: 10.1126/science.aaf6803
53. Ofengeim D, Ito Y, Najafov A, Zhang Y, Shan B, DeWitt JP, et al. Activation of Necroptosis in Multiple Sclerosis. *Cell Rep* (2015) 10:1836–49. doi: 10.1016/j.celrep.2015.02.051

54. Bonnet MC, Preukschat D, Welz P-S, van Loo G, Ermolaeva MA, Bloch W, et al. The Adaptor Protein FADD Protects Epidermal Keratinocytes From Necroptosis *In Vivo* and Prevents Skin Inflammation. *Immunity* (2011) 35:572–82. doi: 10.1016/j.immuni.2011.08.014
55. Welz P-S, Wullaert A, Vlantis K, Kondylis V, Fernández-Majada V, Ermolaeva M, et al. FADD Prevents RIP3-Mediated Epithelial Cell Necrosis and Chronic Intestinal Inflammation. *Nature* (2011) 477:330–4. doi: 10.1038/nature10273
56. Pierdomenico M, Negroni A, Stronati L, Vitali R, Prete E, Bertin J, et al. Necroptosis Is Active in Children With Inflammatory Bowel Disease and Contributes to Heighten Intestinal Inflammation. *Am J Gastroenterol* (2014) 109:279–87. doi: 10.1038/ajg.2013.403
57. Najafov A, Chen H, Yuan J. Necroptosis and Cancer. *Trends Cancer Res* (2017) 3:294–301. doi: 10.1016/j.trecan.2017.03.002
58. Tejada-Martinez D, de Magalhães JP, Opazo JC. Positive Selection and Gene Duplications in Tumour Suppressor Genes Reveal Clues About How Cetaceans Resist Cancer. *Proc Biol Sci* (2021) 288:20202592. doi: 10.1098/rspb.2020.2592
59. Liang S, Mele J, Wu Y, Buffenstein R, Hornsby PJ. Resistance to Experimental Tumorigenesis in Cells of a Long-Lived Mammal, the Naked Mole-Rat (Heterocephalus Glaber). *Aging Cell* (2010) 9:626–35. doi: 10.1111/j.1474-9726.2010.00588.x
60. Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, et al. Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. *JAMA* (2015) 314:1850–60. doi: 10.1001/jama.2015.13134
61. Oliveira GP, Rodrigues RAL, Lima MT, Drumond BP, Abrahão JS. Poxvirus Host Range Genes and Virus-Host Spectrum: A Critical Review. *Viruses* (2017) 9:9–12. doi: 10.3390/v9110331
62. Werden SJ, Rahman MM, McFadden G. Poxvirus Host Range Genes. *Adv Virus Res* (2008) 71:135–71. doi: 10.1016/S0065-3527(08)00003-1
63. Davies MV, Chang HW, Jacobs BL, Kaufman RJ. The E3L and K3L Vaccinia Virus Gene Products Stimulate Translation Through Inhibition of the Double-Stranded RNA-Dependent Protein Kinase by Different Mechanisms. *J Virol* (1993) 67:1688–92. doi: 10.1128/jvi.67.3.1688-1692.1993
64. Sharp TV, Moonan F, Romashko A, Joshi B, Barber GN, Jagus R. The Vaccinia Virus E3L Gene Product Interacts With Both the Regulatory and the Substrate Binding Regions of PKR: Implications for PKR Autoregulation. *Virology* (1998) 250:302–15. doi: 10.1006/viro.1998.9365
65. Chang HW, Jacobs BL. Identification of a Conserved Motif That is Necessary for Binding of the Vaccinia Virus E3L Gene Products to Double-Stranded RNA. *Virology* (1993) 194:537–47. doi: 10.1006/viro.1993.1292
66. Koehler HS, Cotsmire S, Zhang T, Balachandran S, Upton JW, Langland J, et al. Competition Between E3 and ZBP1 for Z-RNA Dictates Susceptibility to Vaccinia Virus-Induced Necroptosis. *SSRN Electron J* (2020) 114(43):11506–11. doi: 10.2139/ssrn.3717768
67. Arndt WD, Cotsmire S, Trainor K, Harrington H, Hauns K, Kibler KV, et al. Evasion of the Innate Immune Type I Interferon System by Monkeypox Virus. *J Virol* (2015) 89:10489–99. doi: 10.1128/JVI.00304-15
68. Essbauer S, Pfeffer M, Meyer H. Zoonotic Poxviruses. *Vet Microbiol* (2010) 140:229–36. doi: 10.1016/j.vetmic.2009.08.026
69. Rahman MM, Liu J, Chan WM, Rothenburg S, McFadden G. Myxoma Virus Protein M029 is a Dual Function Immunomodulator That Inhibits PKR and Also Conscripts RHA/DHX9 to Promote Expanded Host Tropism and Viral Replication. *PloS Pathog* (2013) 9:e1003465. doi: 10.1371/journal.ppat.1003465
70. Águeda-Pinto A, Lemos de Matos A, Abrantes M, Kraberger S, Risalde MA, Gortázar C, et al. Genetic Characterization of a Recombinant Myxoma Virus in the Iberian Hare (). *Viruses* (2019) 11:3–13. doi: 10.3390/v11060530
71. Cooper N, Jetz W, Freckleton RP. Phylogenetic Comparative Approaches for Studying Niche Conservatism. *J Evol Biol* (2010) 23:2529–39. doi: 10.1111/j.1420-9101.2010.02144.x
72. Holt RD. Bringing the Hutchinsonian Niche Into the 21st Century: Ecological and Evolutionary Perspectives. *Proc Natl Acad Sci USA* (2009) 106 Suppl 2:19659–65. doi: 10.1073/pnas.0905137106
73. Simmonds P, Aiewsakun P, Katzourakis A. Prisoners of War - Host Adaptation and Its Constraints on Virus Evolution. *Nat Rev Microbiol* (2019) 17:321–8. doi: 10.1038/s41579-018-0120-2
74. van der Lee R, Wiel L, van Dam TJP, Huynen MA. Genome-Scale Detection of Positive Selection in Nine Primates Predicts Human-Virus Evolutionary Conflicts. *Nucleic Acids Res* (2017) 45:10634–48. doi: 10.1093/nar/gkx704
75. Münk C, Willemsen A, Bravo IG. An Ancient History of Gene Duplications, Fusions and Losses in the Evolution of APOBEC3 Mutators in Mammals. *BMC Evol Biol* (2012) 12:71. doi: 10.1186/1471-2148-12-71

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fimmu.2021.733541

Bacterial Engulfment Mechanism Is Strongly Conserved in Evolution Between Earthworm and Human Immune Cells

Bohdana Kokhanyuk[1], *Kornélia Bodó*[1], *György Sétáló Jr*[2,3], *Péter Németh*[1] *and Péter Engelmann*[1]*

[1] Department of Immunology and Biotechnology, Clinical Center, Medical School, University of Pécs, Pécs, Hungary,
[2] Department of Medical Biology and Central Electron Microscope Laboratory, Medical School, University of Pécs, Pécs, Hungary,
[3] Signal Transduction Research Group, János Szentágothai Research Centre, Pécs, Hungary

Invertebrates, including earthworms, are applied to study the evolutionarily conserved cellular immune processes. Earthworm immunocytes (so-called coelomocytes) are functionally similar to vertebrate myeloid cells and form the first line of defense against invading pathogens. Hereby, we compared the engulfment mechanisms of THP-1 human monocytic cells, differentiated THP-1 (macrophage-like) cells, and *Eisenia andrei* coelomocytes towards *Escherichia coli* and *Staphylococcus aureus* bacteria applying various endocytosis inhibitors [amantadine, 5-(*N*-ethyl-*N*-isopropyl) amiloride, colchicine, cytochalasin B, cytochalasin D, methyl-ß-cyclodextrin, and nystatin]. Subsequently, we investigated the messenger RNA (mRNA) expressions of immune receptor-related molecules (*TLR*, *MyD88*, *BPI*) and the colocalization of lysosomes with engulfed bacteria following uptake inhibition in every cell type. Actin depolymerization by cytochalasin B and D has strongly inhibited the endocytosis of both bacterial strains in the studied cell types, suggesting the conserved role of actin-dependent phagocytosis. Decreased numbers of colocalized lysosomes/bacteria supported these findings. In THP-1 cells *TLR* expression was increased upon cytochalasin D pretreatment, while this inhibitor caused a dropped *LBP/BPI* expression in differentiated THP-1 cells and coelomocytes. The obtained data reveal further insights into the evolution of phagocytes in eukaryotes. Earthworm and human phagocytes possess analogous mechanisms for bacterial internalization.

Keywords: macrophages, coelomocytes, evolution, endocytosis, phagocytosis, inhibitors

OPEN ACCESS

Edited by:
Bostjan Kobe,
The University of Queensland, Australia

Reviewed by:
James L. Stafford,
University of Alberta, Canada
Lage Cerenius,
Uppsala University, Sweden

***Correspondence:**
Péter Engelmann
engelmann.peter@pte.hu

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *30 June 2021*
Accepted: *12 August 2021*
Published: *01 September 2021*

Citation:
Kokhanyuk B, Bodó K, Sétáló Jr G, Németh P and Engelmann P (2021) Bacterial Engulfment Mechanism Is Strongly Conserved in Evolution Between Earthworm and Human Immune Cells. Front. Immunol. 12:733541. doi: 10.3389/fimmu.2021.733541

1 INTRODUCTION

The cellular uptake mechanism of macromolecules, nutrients, and various particles is generally considered as endocytosis. Its action is strongly related to the plasma membrane, which, by invaginating, forms vesicles containing transported molecules. Endocytic pathways govern essential functions such as cell–cell communication, signal transduction, immune response, and cellular

homeostasis. Generally, two major types of endocytosis can be distinguished in the cells: phagocytosis and pinocytosis (1). Phagocytosis is carried out against particles larger than >0.5 μm by specialized cells to protect the organism from microbes and eliminate damaged cells (2).

At least four types of pinocytosis mechanisms exist in the cells, among which are macropinocytosis, clathrin-mediated endocytosis (CME), caveolae-mediated endocytosis (CvME), and clathrin/caveolin-independent endocytosis. All these pathways differ from the point of the origin of the particle engulfed, size, and mechanism of endocytic vesicle formation (1).

Due to its importance in cellular homeostasis, endocytosis (phagocytosis and pinocytosis) is a phylogenetically conserved trafficking process that evolved from amoeba to human cells (3). It seems that phagocytosis is lost from fungi and plants. Interestingly, it is suggested that the primary forms of uptake mechanisms coevolved with nutrition intake and only later opted as immune/defense mechanism during metazoan phylogenesis (4). To eliminate pathogen structures, distinct macrophage types (e.g., archeocytes, hemocytes, coelomocytes, microglia, Kupffer cells) have been developed in the diverse animal groups (5).

In this regard, well-conserved cellular and humoral innate immune components have emerged in invertebrates to keep their self-integrity similarly to those in vertebrates. For instance, segmented worms (earthworms) are "up and coming" research animal models to study the phylogenesis of immunity (6, 7). Earthworms lack adaptive immunity that allows to solely study the mechanisms of innate immunity. Earthworm's body cavity (coelomic cavity) is an immunologically competent compartment containing motile immune cells (so-called coelomocytes) and a fluid abundant in multifunctional proteins (coelomic fluid) (7). Coelomocytes can be subdivided into two major subpopulations (amoebocytes and eleocytes). These cells possess a variety of immune functions that resembles vertebrate leukocytes. In particular, amoebocytes are capable of phagocytosis and encapsulation, while eleocytes produce antimicrobial factors and maintain homeostasis (8).

To this end, we hypothesized that earthworm amoebocytes resemble vertebrate myeloid lineage, not only on functional but also molecular levels. To test this theory, applying various endocytosis inhibitors, we compared the bacterial uptake mechanisms, intracellular localization of engulfed bacteria, and mRNA expression of pattern recognition receptors (PRRs) between earthworm coelomocytes and the vertebrate counterpart (THP-1 human monocytic cell line).

2 MATERIALS AND METHODS

2.1 Cell Culture Conditions

Human monocytic leukemia cell line THP-1 (ATCC® TIB-202™) was cultured in Roswell Park Memorial Institute (RPMI) medium supplemented with 10% heat-inactivated fetal bovine serum (FBS, Euroclone, Milan, Italy) and 1% penicillin/streptomycin (100 U/ml penicillin and 100 μg/ml streptomycin, Lonza, Basel, Switzerland) at 37°C in a humidified 5% CO_2 atmosphere. THP-1 cells were differentiated (diff. THP-1 cells) using 5 ng/ml phorbol 12-myristate-13-acetate (PMA) for 48 h applying the aforementioned culture conditions to obtain macrophage-like cells (9). Differentiation of monocytes was verified with light microscopy.

2.2 Earthworm Husbandry and Coelomocyte Isolation

Adult *Eisenia andrei* (Oligochaeta, Lumbricidae) were maintained at standard laboratory conditions at room temperature (21°C) and fed with manure soil (10). Before coelomocyte isolation, earthworms were placed overnight onto moist paper towel to empty their digestive tract. Coelomocytes were collected by applying an extrusion buffer and washed in *Lumbricus* balanced salt solution (LBSS) as described earlier (10). Density of coelomocytes was enumerated (1×10^6 cells/ml in each assay) by a dead-cell exclusion method applying trypan blue dye.

2.3 Endocytosis Inhibitor Treatments

THP-1 cells, diff. THP-1 cells, and coelomocytes were placed onto 24-well plates and pretreated with different pharmacological pathway inhibitors: cytochalasin D, cytochalasin B, colchicine, 5-(*N*-ethyl-*N*-isopropyl) amiloride (EIPA), amantadine, methyl-ß-cyclodextrin, and nystatin. All inhibitors were purchased from Sigma-Aldrich (St. Louis, MO, USA). Applied inhibitor concentrations and incubation time were selected based on the previous literature [(11–17), please see **Table 1**]. For energy-dependent uptake inhibition, the target cells were preincubated at 4°C for 30 min to block any active uptake mechanism (18). Prior to flow cytometry measurements, samples were stained with 7-aminoactinomycin D (7-AAD, 1 μg/ml, Biotium, Fremont, CA, USA) for the detection of cell viability.

2.4 *In Vitro* Bacterial Challenge

Initially, THP-1 cells, diff. THP-1 cells, and coelomocytes were incubated with fluorescein isothiocyanate (FITC)-coupled,

TABLE 1 | List of the pharmacological inhibitors and their applied conditions in the endocytosis experiments.

Inhibitor	Target pathway	Mode of action	Applied conditions	References
Cytochalasin B	Phagocytosis	F-actin depolymerization	5 μM, 1 h	(11)
Cytochalasin D			5 μM, 1 h	(12)
Colchicine	Pinocytosis	Inhibits microtubule polymerization	100 μM, 2 h	(13)
5-(*N*-ethyl-*N*-isopropyl) amiloride (EIPA)	Macropinocytosis	Blocking the Na^+/H^+ exchanger	5 μM, 30 min	(14)
Amantadine	CME	Blocking the budding of clathrin-coated pits	500 μM, 30 min	(15)
Methyl-ß-cyclodextrin (MßCD)	CvME	Removing cholesterol out of the plasma membrane	1 mM, 30 min	(16)
Nystatin	CvME	Cholesterol sequestration	54 μM, 30 min	(17)

heat-inactivated *Escherichia coli* (ATCC 25922) and *Staphylococcus aureus* (OKI 112001) for 1, 2, 4, 16, and 24 h (THP-1 and diff. THP-1 at 37°C) or 1, 2, 4, 16, 24, and 48 h (coelomocytes at room temperature) to assess bacterial uptake kinetics. Immune cells (10^6) and bacteria cells (10^7) were incubated together in 1 ml of cell culture media (8). Prior to flow cytometry, trypan blue was added to differentiate bound/ingested bacteria by quenching. For uptake inhibition analysis, inhibitor-treated target cells were incubated with bacterial strains at appropriate temperatures (except during energy-dependent uptake inhibition target cells were kept at 4°C) for 24 h. Besides, vehicle controls were also applied. Samples have been washed and resuspended in phosphate-buffered saline (PBS) (THP-1 cells and diff. THP-1 cells) or LBSS (coelomocytes). Fluorescence signals were measured in FL1 (530/30 filter) gated on living cells, as detected by 7-AAD in FL3 (670 LP filter).

2.5 Flow Cytometry

Flow cytometry analysis was carried out with a FACSCalibur flow cytometer (Beckton Dickinson, Franklin Lakes, NJ, USA). During experiments, 30,000 events per sample were measured based on their forward scatter (FSC) and sideward scatter (SSC) characteristics. Results were analyzed with FCS Express software (De Novo Software, Glendale, CA, USA).

2.6 Labeling F-Actin With Phalloidin and Lysosome Staining

After inhibitor treatment and incubation with FITC-coupled bacterial strains, cells (80 µl of 1×10^5/ml) were spread onto slides applying Cytospin 3 (SHANDON, Thermo Scientific, Waltham, MA, USA). Cells were fixed in 4% paraformaldehyde, washed, and permeabilized with PBS/0.1% Triton X-100 with 5% bovine serum albumin (BSA) for 20 min. F-Actin was stained with CF568 Phalloidin (Biotium, Fremont, CA, USA) or Alexa Fluor 488 Phalloidin (Invitrogen Molecular Probes, Eugene, OR, USA) for 45 min. For lysosome staining, samples were incubated with LysoTracker® Red DND-99 (Molecular Probes, Eugene, OR, USA) at a final concentration of 100 nM for 30 min, room temperature, in the dark followed by nuclear counterstaining with 4′,6-diamidino-2-phenylindole (DAPI, Sigma-Aldrich).

2.7 Confocal Laser Scanning Microscopy and Quantitative Colocalization Analysis

Confocal laser scanning microscopy (CLSM) images were generated using an Olympus FV-1000 laser scanning system. Single optical sections were taken with a 40×, long working distance phase objective at zoom settings 3, 4, 5, or 6 for fluorescent captures. For immersion images, the magnification was 60×. Images were then processed and merged with ImageJ1.x software.

CLSM images were used to quantify the colocalization of FITC-conjugated *E. coli* and *S. aureus* with LysoTracker labeled lysosomes. At least 14 lysosomes were evaluated from each condition. The Coloc2 plug-in for Fiji was used to determine the Manders' coefficient after applying region of interest (ROI) for each microscopic picture (19, 20). Manders' coefficient measures the fraction of one dye (green) that colocalizes with a second dye (red); M1 measures the fraction of bacteria colocalizing with lysosomes.

2.8 Statistical Analysis

All the experiments were performed at least three independent times. Results were analyzed with GraphPad Prism 8.0 software (La Jolla, CA, USA). Data were expressed as mean ± SD of the values received in independent experiments. One-way ANOVA or two-way ANOVA with multiple comparison *post-hoc* test was carried, out and the significance of data was evaluated. Differences were considered significant if $p < 0.05$.

3 RESULTS

3.1 Actin Polymerization Inhibitors Attenuate the Bacterial Engulfment by THP-1 Cells

First, THP-1 cells were exposed to FITC-coupled heat-inactivated *E. coli* and *S. aureus* at different time points to estimate their uptake capacity. Kinetic analysis revealed the prompt bacteria engulfment starting from the first hour of incubation and reaching approximately 90% of cellular uptake at 24 h in THP-1 cells (**Figure 1A**). Next, prior to adding bacteria, we pretreated THP-1 cells with different endocytosis pathway inhibitors (**Table 1**) to obtain information about the uptake mechanisms. Cytochalasin D and B were applied to inhibit phagocytosis, colchicine for pinocytosis blocking, EIPA to disrupt macropinocytosis, amantadine to inhibit CME and MßCD, and nystatin to block CvME.

Among all applied inhibitors, cytochalasin B significantly inhibited only the engulfment of *S. aureus* (**Figures 1B, C**), regardless of its weak toxicity against THP-1 cells (**Supplementary Figure 1A**). In contrast, cytochalasin D significantly blocked the uptake of both bacterial strains (**Figures 1B, C**) compared to Ø inhibitor controls (**Figures 1B, C**). There is also a slight inhibitory effect of amantadine for *S. aureus* engulfment (**Figure 1C**), while other tested inhibitors did not reduce significantly the bacterial uptake in THP-1 cells (**Figure 1C**). The highest inhibition rate in bacterial engulfment was detected when THP-1 cells were incubated with bacteria at 4°C to assess the energy dependence of endocytosis (**Figure 1C**).

CLSM studies using phalloidin showed colocalization of engulfed FITC-coupled bacteria and labeled actin, which supports the involvement of actin-dependent endocytosis in the bacterial uptake (**Figure 1D**). The preincubation of cells with cytochalasin D resulted in disruption of actin polymerization and visible change of its structure (**Supplementary Figure 2**) that subsequently leads to a biased internalization of bacteria (**Figure 1E**).

3.2 Differentiated THP-1 Cells Resemble Normal THP-1 Monocytes During the Bacterial Engulfment

The macrophage-like phenotype of THP-1 cells (diff. THP-1) was achieved by applying a PMA stimulation for 48 h. Kinetic analysis of bacterial engulfment by diff. THP-1 cells (**Figure 2A**) revealed similar trends to those of normal THP-1 cells (**Figure 1A**).

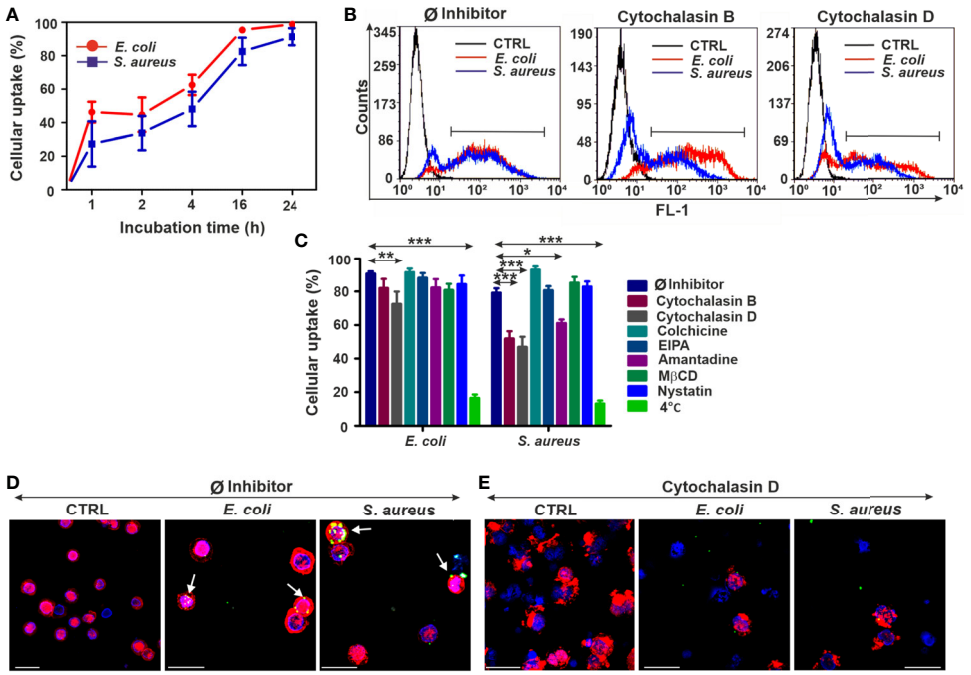


FIGURE 1 | The upshots of endocytosis inhibition on the bacterial uptake in THP-1 cells. **(A)** Kinetics of FITC-*E. coli* and FITC-*S. aureus* bacteria uptake by THP-1 cells after 1, 2, 4, 16, and 24 h of incubation. **(B)** Representative flow cytometry histograms of THP-1 cells after 24 h of bacteria engulfment without or with inhibitors (5 μM cytochalasin B/cytochalasin D). **(C)** Efficiency of different endocytosis inhibitors on the bacterial uptake of THP-1 cells. The results are presented as mean ± SD of four replicates. Asterisks denote statistical significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$) between the Ø inhibitor control and different treatments. **(D)** Merged CLSM images of THP-1 cells incubated in the absence of inhibitors and **(E)** in the presence of 5 μM cytochalasin D that demonstrate the changes in the actin cytoskeleton (red, CF568 Phalloidin) and inhibition of bacterial (green, FITC-*E. coli* and FITC-*S. aureus*) uptake. Nuclear counterstaining was performed with DAPI (blue). Note the overlapping signal between engulfed bacteria and actin filaments (arrows, **D**). Scale bars: 50 μm.

Until 24 h of incubation, *E. coli* internalization reached approximately 90% and somewhat less for *S. aureus* (**Figure 2A**). Subsequently, we analyzed the effects of various inhibitors on the bacterial uptake by diff. THP-1 cells (survival rate of diff. THP-1 cells was only affected by colchicine, **Supplementary Figure 1B**). While cytochalasin B only decreased the *S. aureus* engulfment rate, the strongest inhibition was evoked by cytochalasin D in diff. THP-1 cells (**Figures 2B, C**) and in normal THP-1. We did not observe any effects in the case of other inhibitors (**Figure 2C**). The energy dependence of bacterial uptake was monitored at 4°C and evidenced strong inhibition similarly to normal THP-1 cells (**Figure 2C**).

By means of CLSM, an extensive bacterial uptake in untreated diff. THP-1 cells was observed (**Figure 2D**). Imaging of the diff. THP-1 immunocytes exposed to cytochalasin D (**Figure 2E**, **Supplementary Figure 2**) has supported the previous results of bacterial uptake inhibition obtained by flow cytometry.

3.3 Coelomocyte-Mediated Bacterial Engulfment Is Blocked by Actin Depolymerization

According to physical parameters such as size and granularity, the two major coelomocyte subpopulations (e.g., amoebocytes, and eleocytes) can be distinguished by flow cytometry (8). In our uptake experiments, we have focused on amoebocytes, since this population is involved in the cellular antibacterial defense among coelomocyte subtypes (8). In addition, eleocytes possess a strong autofluorescence (derived from riboflavin) that hampers their fluorescence dye-based analysis (21).

In the case of coelomocytes (amoebocytes), the bacterial uptake starts to gradually increase at 4 h and reaches about 70% at the end of the observation period (48 h) (**Figure 3A**). Cytochalasin D showed slight cytotoxicity for earthworm coelomocytes (**Supplementary Figure 1C**); nevertheless, it has efficiently blocked their endocytic activity similarly to human immunocytes (**Figures 3B, C**). Interestingly, Gram-negative bacterial uptake was also blocked by cytochalasin B (**Figures 3B, C**) on the contrary to normal and diff. THP-1 cells (**Figures 1C**, **2C**). Similarly, to the human cells, other endocytosis inhibitors did not cause any noticeable blocking effects in coelomocytes; however, in the case of nystatin, we observed a non-significant trend of decreased engulfment rate (**Figure 3C**).

Inhibition of energy-dependent endocytosis (coelomocytes at 4°C) did not rescue *S. aureus* bacteria from engulfment. Even the blockade of *E. coli* internalization was less extended compared to the human counterparts (**Figure 3C**). CLSM analysis of phalloidin costained coelomocytes (Ø inhibitor control, **Figure 3D**) has revealed the immense uptake of target bacteria, especially in the case of *S. aureus*. In line with flow cytometry measurements, CLSM analysis has supported that cytochalasin D pretreatment

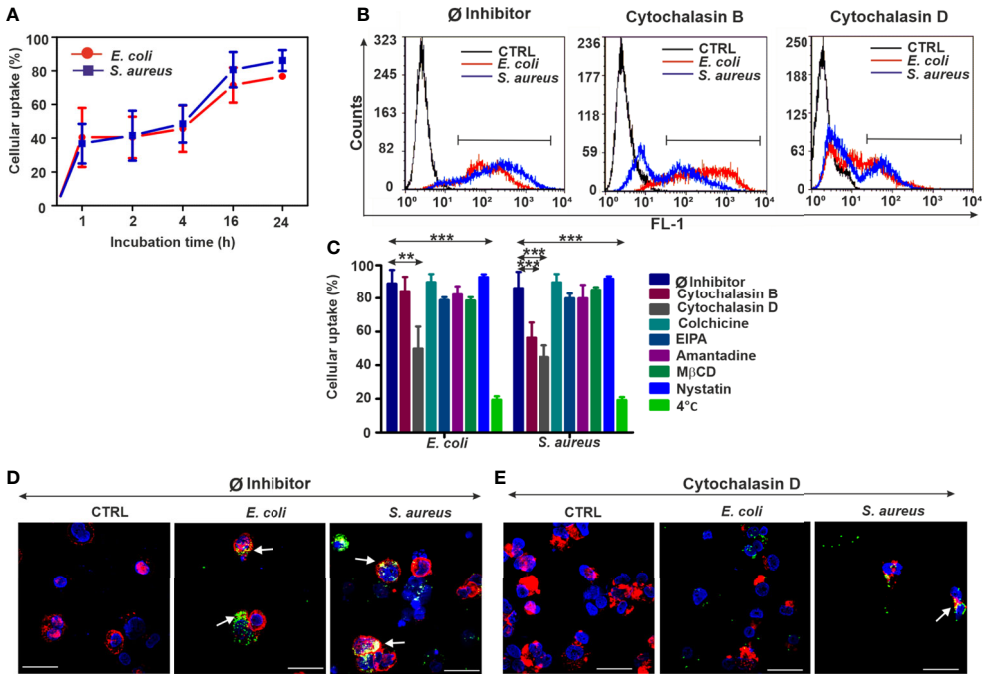


FIGURE 2 | Biased bacterial engulfment in diff. THP-1 cells upon endocytosis inhibition. **(A)** Kinetics of FITC-*E. coli* and FITC-*S. aureus* uptake by diff. THP-1 cells after 1, 2, 4, 16, and 24 h of incubation. **(B)** Representative flow cytometry histograms of diff. THP-1 cells after 24 h of bacteria uptake, without or with inhibitors (5 μM cytochalasin B/cytochalasin D). **(C)** Efficiency of various endocytosis blockers on the bacterial engulfment in diff. THP-1 cells. The results are presented as mean ± SD of four replicates. Asterisks mark statistical significance (**$p < 0.01$, ***$p < 0.001$) between the Ø inhibitor control and various treatments. **(D)** Merged CLSM images of diff. THP-1 cells in the absence of endocytosis inhibitors or **(E)** in the presence of 5 μM cytochalasin D that demonstrate the changes in the actin cytoskeleton (red, CF568 Phalloidin) and inhibition of bacterial (green, FITC-*E. coli* and FITC-*S. aureus*) uptake. Nuclear counterstaining was performed with DAPI (blue). Note the overlapping signal between engulfed bacteria and actin filaments (arrows, **D**). Scale bars: 50 μm.

efficiently impairs actin polymerization (**Supplementary Figure 2**) and bacterial uptake by coelomocytes (**Figure 3E**).

3.4 Decreased Colocalization of Lysosomes and Engulfed Bacteria by Actin Polymerization Blockers

To track the intracellular fate of fluorescent bacterial particles, earthworm and human immune cells were stained with LysoTracker Red (**Figure 4** and **Supplementary Figures 3–5**). In the course of CLSM analysis, we observed stronger fluorescent signals from lysosomes in bacteria-treated cells compared to the unexposed, control THP-1 cells (**Supplementary Figure 3A**). Application of cytochalasin D has lowered the bacterial uptake by THP-1 cells; hence, it significantly decreased the amount of colocalized signal (**Supplementary Figure 3B**), which was statistically confirmed by Manders' colocalization coefficient value analysis (**Figure 4A**). The fraction of colocalized *E. coli*/lysosomes and *S. aureus*/lysosomes upon the cytochalasin D treatment was nearly two times lower compared to the Ø inhibitor control values (**Figure 4A**).

Similarly, to normal THP-1 cells, in diff. THP-1 cells, the Manders' coefficient analysis revealed that the fluorescent signals from colocalized bacteria and lysosomes significantly decreased upon cytochalasin D treatment (**Figure 4B** and **Supplementary Figures 4A, B**).

Regarding the role of lysosomes in the elimination of ingested bacteria, the uptake inhibition by cytochalasin D was demonstrated by CLSM in coelomocytes (**Figure 4C**, **Supplementary Figures 5A, B**). We observed more than 10-fold decrease in colocalized *E. coli*/lysosomes and more than 3-fold decrease in colocalized *S. aureus*/lysosomes (**Figure 4C**) in the case of coelomocytes.

3.5 Pattern Recognition Receptor mRNA Levels Vary Upon Cytochalasins Exposure

Elimination of microbial pathogens is facilitated by evolutionarily conserved pattern recognition receptors (PRRs). Certain homologs of PRRs have been described from earthworms (22). To assess their possible involvement during the endocytosis in earthworm and human cells, we tested the mRNA expression levels of *toll-like receptor (TLR)*, *myeloid differentiation factor 88 (MyD88)*, and *lipopolysaccharide-binding protein/bactericidal permeability-increasing protein (LBP/BPI)* [for detailed methodological description of RNA isolation, complementary DNA (cDNA) synthesis, and the conditions of qPCR measurements, please see the **Supplementary Material**] upon bacterial challenge with or without endocytosis inhibitor pretreatments (the characteristics of applied primer sequences are detailed in **Supplementary Table 1**).

No significant elevations of *TLR* and *MyD88* levels by bacterial stimulation were revealed in THP-1 cells, although

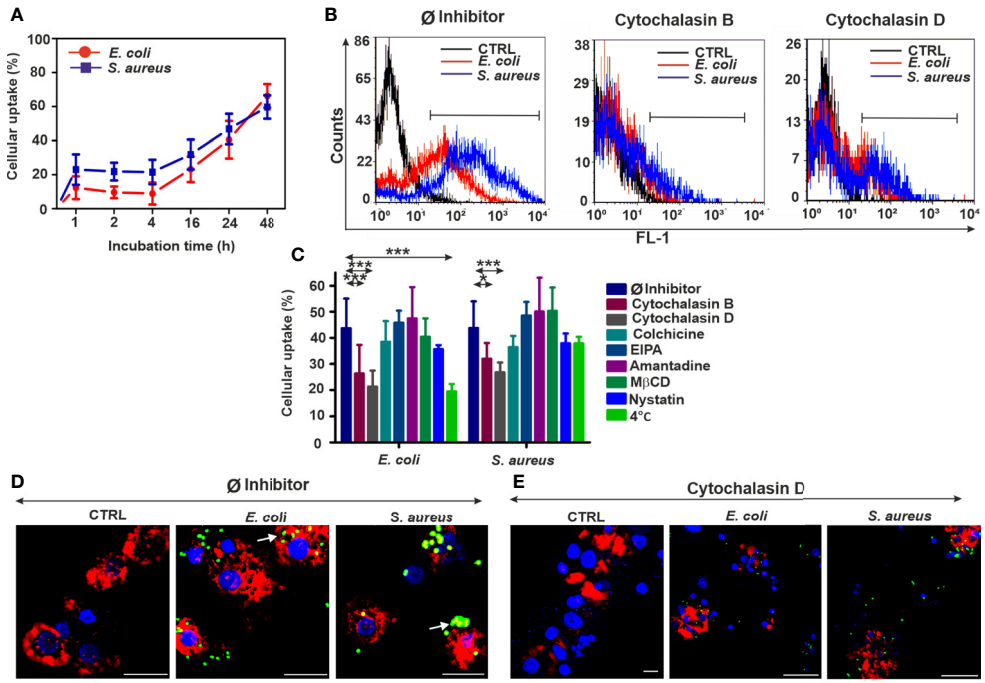


FIGURE 3 | Reduced bacterial uptake in the coelomocytes following endocytosis inhibition. **(A)** Uptake kinetics of FITC-*E. coli* and FITC-*S. aureus* bacteria by coelomocytes after 1, 2, 4, 16, 24, and 48 h of incubation. **(B)** Representative flow cytometry histograms after 24 h of bacteria engulfment without or with inhibitors (5 μM cytochalasin B/cytochalasin D). **(C)** Efficiency of numerous endocytosis inhibitors on the bacterial uptake in coelomocytes. The results are exhibited as mean ± SD of four replicates. Asterisks indicate statistical significance (*$p < 0.05$, ***$p < 0.001$) between the Ø inhibitor control and different treatments. **(D)** Merged CLSM images of coelomocytes in the absence or **(E)** presence of 5 μM cytochalasin D that display the changes in the actin cytoskeleton (red: CF568 Phalloidin) and inhibition of bacterial (green, FITC-*E. coli* and FITC- *S. aureus*) uptake. Nuclear counterstaining was performed with DAPI (blue). Note the overlapping signal between engulfed bacteria and actin filaments (arrows, **D**). Scale bars: 20 μm.

cytochalasin D increased their expression regardless of bacterial exposure (**Figures 5A, B**). *BPI* mRNA expression pattern was not affected by either condition (**Figure 5C**). Incubation with *E. coli* has downregulated *TLR* mRNA expression compared to the corresponding control samples in diff. THP-1 cells (**Figure 5D**). Cytochalasin B treatment evoked a slight decrease in *TLR* mRNA level (**Figure 5D**). In the case of *MyD88*, there were no significant changes (**Figure 5E**), while *BPI* mRNA level was elevated following *S. aureus* exposure and decreased upon actin disruption by cytochalasin D (**Figure 5F**).

In the case of coelomocytes, *TLR* mRNA expression was profoundly elevated only in the response towards *S. aureus*; however, neither of cytochalasins have changed this parameter (**Figure 5G**). No alteration in *MyD88* mRNA level was observed during any exposure conditions (**Figure 5H**). Nevertheless, *LBP/BPI* expression is significantly peaked in response to Gram-negative *E. coli* challenge and attenuated by cytochalasin D inhibition in every condition (**Figure 5I**).

4 DISCUSSION

In vertebrates, myeloid and lymphoid cells are responsible for maintaining the immune response, while in invertebrates, immune functions are carried out by mesoderm-derived, specialized immune cells (e.g., coelomocytes, hemocytes). Earthworms, living in a microbe-rich environment, are constantly exposed to bacteria; their immune homeostasis is governed by several cellular and humoral components (23–25). Recent publications are more focused on the variations of immune-related genes upon microbial challenge in earthworms (26, 27). However, the different cellular engulfment routes to eliminate microbes by coelomocytes are rather unknown in earthworms, similarly to other invertebrate models.

In this study, we aimed to comparatively analyze the various endocytic pathways of bacterial internalization mediated by earthworm and human immune (THP-1) cells. THP-1 cell line proved to be a reliable model in our previous comparative *in vitro* study to observe nanoparticle toxicity against earthworm and human cells (28), and we favored this cell line over other monocyte/macrophage cell lines (J774, RAW264, etc.) to keep our comparative approach standard. Another benefit is that THP-1 monocytes can be differentiated into macrophage-like cells, which proved to resemble human macrophages at functional and phenotype levels (29). Hence, usage of the THP-1 cells line allowed us to investigate and compare bacterial uptake pathways in two human myeloid cell types. Another advantage of cell lines is their uniformity and simplicity

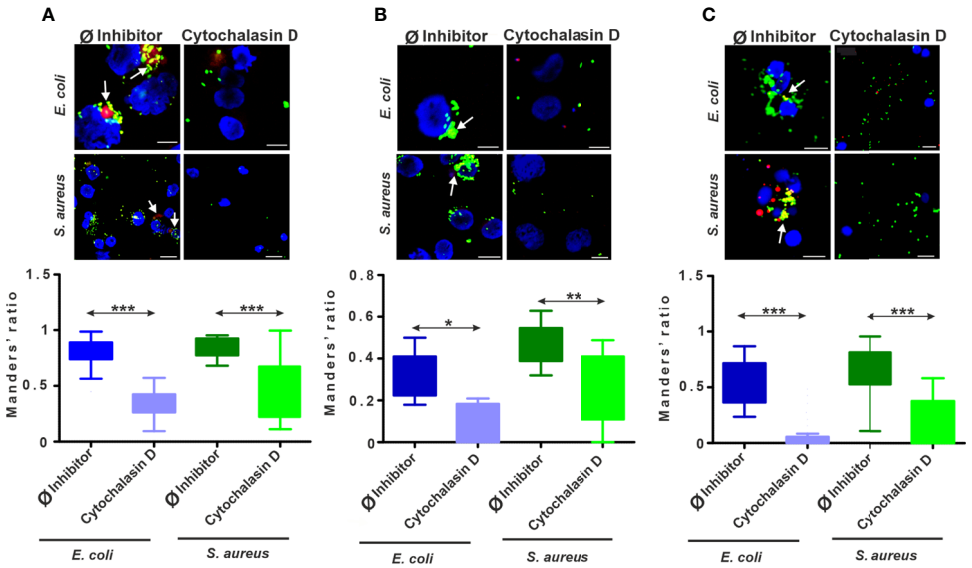


FIGURE 4 | The intracellular colocalization of lysosomes and bacteria is biased upon endocytosis inhibition. Representative CLSM images and Manders' colocalization coefficients ratios for **(A)** THP-1 cells, **(B)** diff. THP-1 cells, and **(C)** coelomocytes after 24 h incubation with bacteria in the absence or the presence of 5 μM cytochalasin D. Representative images demonstrate the colocalization of FITC-coupled bacteria (green) with the LysoTracker labeled (red) lysosomes (arrows). Nuclear counterstaining was performed with DAPI (blue). Scale bars: 10 μm; 50 μm (THP-1 with *S. aureus*). The corresponding graphs show the fraction of colocalized *E. coli*/lysosomes and *S. aureus*/lysosomes. Results are presented as mean ± SD from three independent experiments. Asterisks point out statistical significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$) between the Ø inhibitor control and cytochalasin D exposed cells.

in maintaining under standard laboratory conditions compared to primary cells. It is a valid concern that primary macrophages behave differently compared to a monocyte/macrophage cell line; however, a recent publication (30) has compared the bacterial uptake and inflammatory consequences (chemokine/cytokine mRNA expression and cytokine secretion) between human monocyte-derived macrophages and THP-1 cells and found no significant differences. In addition, another comparison (31) has revealed that THP-1 cell line can be applied as simple substitute for human macrophages in basic functional assays but should be avoided in fine-screening assays of certain drug candidates.

Earlier, we have applied (8, 26) heat-inactivated *E. coli* and *S. aureus* bacteria strains for *in vitro* phagocytosis/stimulation assays. We aimed to keep the reproducibility and comparability with these previous experiments applying the same bacterial strains. Indeed, earthworms tolerate well the microbe-rich environment. Now, there is an increasing body of information about earthworm bacterial community. Recent article by Pass et al. (32) has identified that *Proteobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Acidobacteria* are dominant in the whole body of *Lumbricus rubellus* earthworms. These data was supported by the findings (33) from the gut of *Eisenia andrei*, with only the difference of the occurrence of *Firmicutes* phyla. Furthermore, *L. rubellus* microbiome is dominated by the genus *Serratia*, while in *E. andrei*, it is dominated by *Acinetobacter* genus. According to Dales and Kalaç (23), *Aeromonas hydrophila* and *Serratia marcescens* Gram-negative bacteria are potential natural pathogen bacteria in *E. fetida* earthworms; however, it is widely accepted that human pathogen bacteria are introduced into earthworms (or other invertebrates) (34, 35).

Initially, the kinetic analysis revealed that bacterial engulfment in THP-1, diff. THP-1 cells, and coelomocytes began at the first hour of incubation but to a different extent. According to the literature (29, 36), diff. THP-1 cells take up the bacteria to a greater extent than the undifferentiated THP-1 monocytes. Yet, in our results, the rates of uptake by THP-1 cells and diff. THP-1 cells were rather similar during kinetical analysis.

Compared to the human counterpart, we observed a slower rate of uptake by coelomocytes. This correlates with previous results (37), which also showed a maximum increase in antibacterial (lysozyme-like) activity in coelomic fluid upon 2 days of *in vivo* bacterial challenge. An earthworm-specific PRR, the coelomic cytolytic factor (CCF) that can recognize lipopolysaccharide (LPS), peptidoglycan, and β-1,3-glucan, is also evidenced to show an increased tendency after 24 h of bacterial challenge (6). Hence, this is a reliable representation of the humoral innate response of earthworms that is in line with our data focusing on bacterial uptake by coelomocytes.

We have monitored the passive transfer of bacteria into the cells, by incubation at 4°C, since at this temperature, the active transport, including phagocytosis, is blocked. The uptake was significantly reduced in the tested cell types (**Figures 1**–**3**), except during *S. aureus* uptake by coelomocytes (**Figure 3**). These results indicate that the internalization of *E. coli* and *S. aureus* is an energy-dependent process. Generally, earthworms flourish at 15–20°C, and their cells can tolerate better the lower

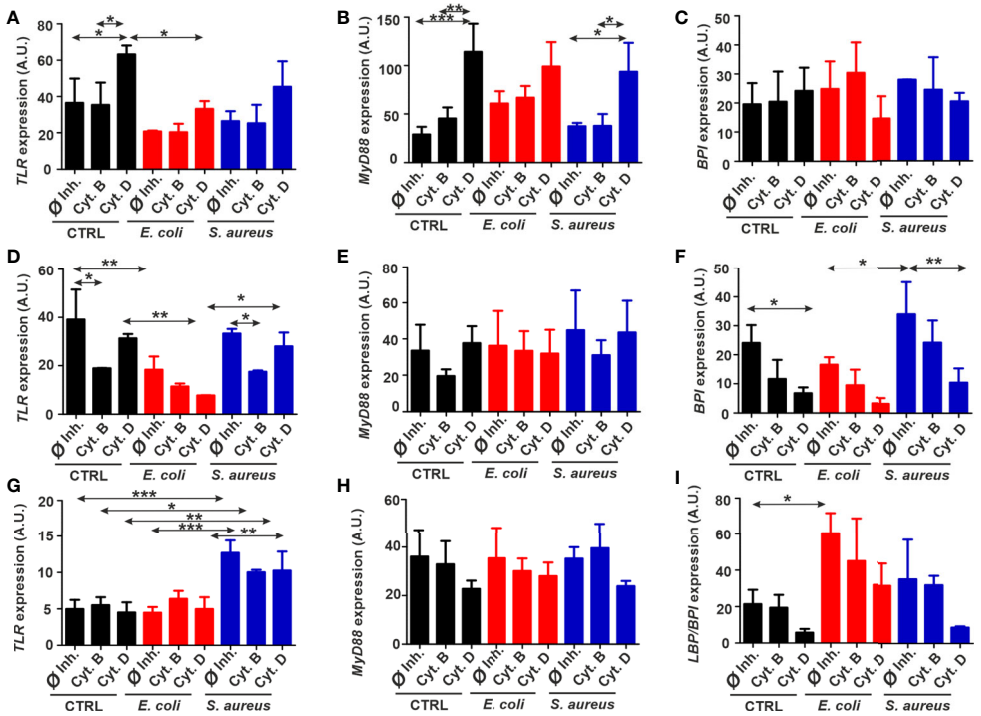


FIGURE 5 | Fluctuations of immune-related gene expressions in target cells exposed to bacteria and phagocytosis inhibitors. Normalized expression of *TLR*, *MyD88*, and *BPI* in **(A–C)** THP-1, **(D–F)** diff. THP-1 and normalized expression of *TLR*, *MyD88*, and *LBP/BPI* **(G–I)** in coelomocytes after bacterial challenge in the absence or presence of the uptake inhibitors. The results are presented as mean ± SD of three replicates. Asterisks display statistical significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). The significance of data was evaluated by one-way ANOVA with Tukey's multiple comparison test. A.U., arbitrary units.

temperature compared to mammalian cells. This may explain the maintained engulfment activity in coelomocytes at 4°C (38).

It is generally established that the particles larger than 500 nm, including most of the bacteria, are taken up by phagocytosis. In this regard, Veiga and Cossart (39) demonstrated that bacterial internalization might not be restricted only to phagocytic uptake, but other endocytosis pathways can be also encompassed. For example, some bacteria, including *E. coli*, manage to avoid degradation in lysosomes by entering the cell through caveolae-mediated endocytosis (40). Therefore, we started our experiments by testing different groups of inhibitors to target certain uptake pathways. By inhibiting the core components of the different endocytic pathways, it is possible to evaluate which of them participates in the endocytosis process. Albeit some of the pharmacological inhibitors have a weak specificity, they remain widely applied (14). They were chosen to target the major known cellular uptake pathways: macropinocytosis, clathrin- and caveolin-mediated endocytosis, and phagocytosis (1).

Amiloride and its derivative 5-(*N*-ethyl-*N*-isopropyl) amiloride (EIPA) was reported to inhibit macropinocytosis by blocking the Na^+/H^+ exchanger (14) and colchicine by disturbing microtubules in pinocytosis (41). In addition, since CvME involves cell membrane cholesterol, this endocytic pathway can be disrupted by cholesterol depletion (42). This is a mechanism of action for methyl-ß-cyclodextrin (MßCD), which possesses hydrophobicity resulting in a high affinity to membrane cholesterol and its depletion (43, 44). Antibiotic nystatin can induce changes in the shape of the caveolae and make aggregates of cholesterol leading to its sequestration (45). As for CME, amantadine was reported to interfere by blocking the budding of clathrin-coated pits (46). The aforementioned endocytosis inhibitors were ineffective for bacterial engulfment in the studied cell types; however, we observed a slight amantadine action on THP-1-mediated *S. aureus* internalization, which might imply that some endocytosis pathways are activated when the other is blocked.

Due to the core role of actin in phagosome formation, such F-actin depolymerizing agents, as cytochalasin B and D, can inhibit phagocytosis (47). Applying cytochalasin D, it turned out to be a universal inhibitor for the uptake of both bacterial strains in the studied cell types, suggesting that both *E. coli* and *S. aureus* are internalized *via* phagocytosis (**Figures 1**–**3**). Another inhibitor of phagocytosis, cytochalasin B also showed effectiveness in blocking the uptake of both bacterial types in coelomocytes (**Figure 3**). Unlike for earthworm immunocytes, in THP-1 and diff. THP-1 cells, cytochalasin B inhibited only the engulfment of *S. aureus* (**Figures 1**, **2**). These results correlate with previous data of *E. coli* engulfment inhibition by cytochalasin D in human monocyte-derived macrophages (48). Cytochalasin D also reduced the *Streptococcus pneumoniae* uptake by THP-1 cells (49). These results underline the crucial interactions of

cytoskeleton rearrangement and phagocytosis in bacterial uptake by in fact different but evolutionarily and functionally similar earthworm and human immunocytes.

To this end, it is worth to point out the role of the actin filaments in various types of endocytosis. Our CLSM studies evidenced that fluorochrome-conjugated phalloidin binds to actin filaments of the cytoskeleton, resulting in the high imaging contrast of the filaments bundles, which was especially valuable in depicting the cytoskeleton changes after application of cytochalasins. As expected, cytochalasin D caused actin depolymerization, which disturbed actin-dependent phagocytic uptake of bacteria, keeping them out of the cells or on the surface of the cells (**Figures 1E**, **3E**). It is well known for mammalian cells that membrane extensions created by actin polymerization are necessary for phagocytosis and macropinocytosis, while the other uptake processes are considered not to involve actin cytoskeleton rearrangements (39, 50, 51). Here, our data underscore the role of actin-dependent phagocytosis of bacteria not only in human but also in invertebrate immunocytes. On the other hand, in this set of experiments, we have only concentrated on the inhibitors of actin polymerization, but phagocytosis is a complex process mediated by several other target molecules. For instance, it is known that inhibition of phosphoinositide 3-kinase (PI 3-kinase), protein kinase C, and Rho family of GTPases also results in reduced phagocytosis by vertebrate macrophages (52). In this regard, such inhibitors (e.g., wortmannin, LY290042, genistein, bisindolymaleimide, ML141) should be tested during future engulfment studies in earthworm coelomocytes, since limited data are available about those from invertebrate models (53, 54).

Applying LysoTracker dye to selectively stain the acidic organelles and the quantitative colocalization approach enables us to more precisely monitor bacterial fate in the cells. Microbial infections can increase the lysosome numbers in earthworm coelomocytes and elevate the activity of phagolysosomal markers (acid and alkaline phosphatase) in amoebocytes (8, 10). These data correlate with our recent CLSM images, where we observed colocalization of engulfed bacteria with lysosomes, suggesting acidification of these organelles and production of phagosomal enzymes. Manders' coefficient values corresponding to bacteria/lysosome colocalization greatly decreased for cytochalasin D treated cells of all types, suggesting that lysosomes assist *E. coli* and *S. aureus* phagocytosis in both invertebrate and vertebrate immunocytes (**Figure 4**).

Innate immunity operates with a set of highly conserved pattern recognition receptors (PRRs). For instance, TLRs were first discovered in *Drosophila melanogaster* as a molecule important in the response to fungal and Gram-positive bacterial infections (55). TLR signaling in mammals shares similarities with *Drosophila*'s TLR-induced mechanisms. It is worth highlighting the role of the most conserved region, Toll/IL-1R (TIR) domain, a cytoplasmic part of TLR, which mediates the recruitment of adaptor proteins and starts the intracellular signaling cascade then leads to inflammatory cytokine production (56).

We have very limited information about the function of PRRs in earthworms (22), and we have relatively restricted research tools to test their involvement during earthworm immune response (7). Downstream inflammatory events occur following pathogen ligand and PRR engagements (27); however, cytokine homologs have not been yet identified in earthworms. Potential inflammatory candidates would be lysenin (8) and the pattern recognition receptor CCF (57); however, those molecules have no known human (vertebrate) homologs. To keep the comparative approach, we decided to choose only those immune-related genes (*TLR*, *MyD88*, and *LBP/BPI*) that we can test in both models.

TLR was recently identified in *E. andrei* earthworms (*Ea*TLR), and it is modulated upon bacteria stimulation (58). Hence, since it resembles structural and functional similarity to its human counterpart, we aimed to identify *TLR* expression upon the application of cytochalasin B and D and incubation with bacteria.

Bacterial stimulation did not elevate the *TLR* mRNA expression levels in THP-1, and in the case of diff. THP-1, *TLR* was even downregulated following the incubation with *E. coli*. Since cytochalasin B was reported to show anti-inflammatory responses, its impact is noticeable in decreased immune-related genes' expression (11), which coincides with *TLR* mRNA downregulation in diff. THP-1 cells of our experiments (**Figure 5D**). In contrast, while cytochalasin D treatment could stop bacteria from entering, it did not prevent the TLR signaling. Since cytochalasin D is also able to upregulate nuclear factor kappa B (NF-κB) levels (59), this might correlate with TLR expression. Nevertheless, it just confirms that the relationships between TLR levels, actin cytoskeleton, and intracellular signaling are still poorly understood.

TLR expression in *S. aureus*-exposed coelomocytes showed a significant increase (**Figure 5G**), which supports earlier results (58) where a fourfold change in *EaTLR* expression was reported after *Bacillus subtilis* stimulation. In contrast, incubation with *E. coli* did not significantly amplify the *EaTLR* expression that is in line with our observations. Since cytochalasins could not markedly reduce *EaTLR* levels, we affirm that actin depolymerization hinders bacterial engulfment in coelomocytes but does not change the capacity of TLR for the engagements with bacterial ligands.

After ligand binding of TLR, MyD88 adaptor triggers downstream signaling cascades, leading to NF-κB and mitogen-activated protein kinase (MAPK) activation and inflammatory cytokine production (60). However, a MyD88-independent pathway was reported by Kawai et al. (61), where they demonstrated that despite impaired cytokine production, MyD88-deficient macrophages had normal NF-κB signaling in response to LPS. In our experiments, bacterial stimulation did not provoke the elevation of *MyD88* expression compared to Ø inhibitor controls (**Figure 5**), which might suggest the involvement of MyD88-independent mechanism in these phagocytic processes.

Furthermore, there is a connection between activation of one or another pathway and the state of the actin cytoskeleton in human macrophages. TLR induces phagocytic gene expression through the MyD88 pathway without the involvement of actin. In turn, a second mechanism is MyD88-independent but requires actin for activation of this pathway (62). This provides the evidence of MyD88-independent pathway involvement in the process of bacterial uptake. Upon the impairment of actin polymerization by cytochalasins, an alternative activation of

actin-independent MyD88 is probable. Besides, we cannot rule out that *MyD88* expression variations occurred at the earlier time points. In this case, our results might be connected to negative feedback regulatory processes to control excessive TLR-mediated mechanisms in the later stages of the infections (63–65).

Even though the MyD88-dependent/independent TLR pathway is well conserved across the evolution (66), limited data for *E. andrei* are available regarding the MyD88 involvement in earthworm immune response. Whole transcriptome analysis of *E. andrei* has identified several immune-related sequences, including MyD88 as an adaptive molecule for MAPK signaling activated by TLR (67). As our results suggest, there is no significant difference between *MyD88* expression levels in unstimulated and bacteria stimulated coelomocytes. Yet, in oppose to THP-1 and diff. THP-1 cells data, cytochalasins show a tendency to downregulate *MyD88* mRNA expression in coelomocytes, which is in line with decreased phagocytosis of bacteria (**Figure 5H**). This could propose that just like in *D. melanogaster*, MyD88 in coelomocytes is located in the plasma membrane (68) and can be disrupted by cytochalasins. Therefore, in earthworms, actin cytoskeleton rearrangement may be a crucial event in the regulation of innate immune responses that control TLRs and their downstream signaling proteins.

The BPI, which, in human macrophages, is known to possess antimicrobial functions, is known to be less expressed in human monocytes (69) that was also observed in our experiments. Unexpectedly, diff. THP-1 cells stronger reacted to Gram-positive than to Gram-negative bacteria (**Figure 5F**). Nevertheless, cytochalasin D affected *BPI* mRNA levels, indicating that BPI is located in the plasma membrane and might also directly participate in the bacterial uptake.

The homolog of this well-conserved PRR was also identified in *E. andrei* (70). Later, our research group showed that *EaLBP/BPI* mRNA is specifically expressed in the amoebocyte subset (71). Previous data have revealed upregulation of *EaLBP/BPI* gene transcription after stimulation with bacteria, including *E. coli* (70), which is in line with our findings. For *S. aureus*, there might be another PRR involved in its recognition and signal transduction, for example, CCF known to be more specific for Gram-positive bacteria (57).

To conclude, we applied *E. coli* and *S. aureus* bacteria to characterize the microbial uptake routes in invertebrate and vertebrate immunocytes. Despite some differences in the intracellular signaling, the conserved actin-dependent phagocytosis suggests the functional similarities of bacterial internalization process for earthworm and human immunocytes.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Conceptualization: BK and PE. Methodology: BK, KB, GS, and PE. Investigation: BK, KB, and GS. Writing—original draft preparation: BK. Writing—review and editing: KB, GS, PN, and PE. Supervision: PE. Funding acquisition: PE and KB. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Medical School Research Foundation University of Pécs (PTE-ÁOK-KA 2017/4), GINOP-232-15-2016-00050 and EFOP-361-16-2016-00004. The purchase of the Olympus FV-1000 laser scanning confocal system was supported by grant GVOP-3.2.1-2004-04-0172/3.0 to the University of Pécs. TKP2020-IKA-08 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 funding scheme. The work was supported by the ÚNKP-19-3-I New National Excellence Program of the Ministry for Innovation and Technology and by the György Romhányi Research Scholarship of the Medical School, University of Pécs.

ACKNOWLEDGMENTS

We sincerely thank Dr. László Molnár and Dániel Dunai (University of Pécs) for providing earthworm specimens. We are grateful for the reagent contribution provided by Dr. Edina Szabó-Meleg (University of Pécs).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.733541/full#supplementary-material

REFERENCES

1. Conner SD, Schmid SL. Regulated Portals of Entry Into the Cell. *Nature* (2003) 422:37–44. doi: 10.1038/nature01451
2. Hall A, Nobes CD. Rho GTPases: Molecular Switches That Control the Organization and Dynamics of the Actin Cytoskeleton. *Philos Trans R Soc Lond B Biol Sci* (2000) 355:965–70. doi: 10.1098/rstb.2000.0632
3. DeDuve C. *Blueprint for a Cell: The Nature and Origin of Life. 1st edition*. Burlington, N.C: Carolina Biological Supply Co (1991).
4. Hartenstein V, Martinez P. Phagocytosis in Cellular Defense and Nutrition: A Food-Centered Approach to the Evolution of Macrophages. *Cell Tissue Res* (2019) 377:527–47. doi: 10.1007/s00441-019-03096-6
5. Arroyo Portilla C, Tomas J, Gorvel JP, Leoulard H. From Species to Regional and Local Specialization of Intestinal Macrophages. *Front Cell Dev Biol* (2021) 8:624213. doi: 10.3389/fcell.2020.624213
6. Bilej M, Baetselier PD, Dijck EV, Stijlemans B, Colige A, Beschin A. Distinct Carbohydrate Recognition Domains of an Invertebrate Defense Molecule Recognize Gram-Negative and Gram-Positive Bacteria. *J Biol Chem* (2001) 276:45840–7. doi: 10.1074/jbc.M107220200
7. Engelmann P, Bodó K, Najbauer J, Németh P. Annelida: Oligochaetes (Segmented Worms): Earthworm Immunity, Quo Vadis? Advances and New Paradigms in the Omics Era. In: Cooper EL, editor. *Advances in Comparative Immunology*. Cham: Springer International Publishing (2018). p. 135–59. doi: 10.1007/978-3-319-76768-0_6

8. Engelmann P, Hayashi Y, Bodó K, Ernszt D, Somogyi I, Steib A, et al. Phenotypic and Functional Characterization of Earthworm Coelomocyte Subsets: Linking Light Scatter-Based Cell Typing and Imaging of the Sorted Populations. *Dev Comp Immunol* (2016) 65:41–52. doi: 10.1016/j.dci.2016.06.017
9. Park EK, Jung HS, Yang HI, Yoo MC, Kim C, Kim KS. Optimized THP-1 Differentiation Is Required for the Detection of Responses to Weak Stimuli. *Inflammation Res* (2007) 56:45–50. doi: 10.1007/s00011-007-6115-5
10. Engelmann P, Molnár L, Pálinkás L, Cooper EL, Németh P. Earthworm Leukocyte Populations Specifically Harbor Lysosomal Enzymes That may Respond to Bacterial Challenge. *Cell Tissue Res* (2004) 316:391–401. doi: 10.1007/s00441-004-0874-x
11. Kim MY, Kim JH, Cho JY. Cytochalasin B Modulates Macrophage-Mediated Inflammatory Responses. *Biomol Ther (Seoul)* (2014) 22:295–300. doi: 10.4062/biomolther.2014.055
12. DeFife KM, Jenney CR, Colton E, Anderson JM. Cytoskeletal and Adhesive Structural Polarizations Accompany IL-13-Induced Human Macrophage Fusion. *J Histochem Cytochem* (1999) 47:65–74. doi: 10.1177/002215549904700107
13. Valberg PA, Brain JD, Kane D. Effects of Colchicine or Cytochalasin B on Pulmonary Macrophage Endocytosis. *Vivo J Appl Physiol* (1981) 50:621–9. doi: 10.1152/jappl.1981.50.3.621
14. Ivanov AI. "Pharmacological Inhibition of Endocytic Pathways: Is It Specific Enough to be Useful"? In: AI Ivanov, editor. *Exocytosis and Endocytosis. Methods in Molecular Biology, vol. 440*. New York, USA: Humana Press (2008). p. 15–33. doi: 10.1007/978-1-59745-178-9_2
15. Van Hamme E, Dewerchin HL, Cornelissen E, Verhasselt B, Nauwynck HJ. Clathrin- and Caveolae-Independent Entry of Feline Infectious Peritonitis Virus in Monocytes Depends on Dynamin. *J Gen Virol* (2008) 89:2147– 56. doi: 10.1099/vir.0.2008/001602-0
16. Atger VM, de la Llera Moya M, Stoudt GW, Rodrigueza WV, Phillips MC, Rothblat GH. Cyclodextrins as Catalysts for the Removal of Cholesterol From Macrophage Foam Cells. *J Clin Invest* (1997) 99:773–80. doi: 10.1172/JCI119223
17. Florian P, Macovei A, Sima L, Nichita N, Mattsby-Baltzer I, Roseanu A. Endocytosis and Trafficking of Human Lactoferrin in Macrophage-Like Human THP-1 Cells. *Biochem Cell Biol* (2012) 90:449–55. doi: 10.1139/o11-090
18. Gliga AR, Skoglund S, Wallinder IO, Fadeel B, Karlsson HL. Size-Dependent Cytotoxicity of Silver Nanoparticles in Human Lung Cells: The Role of Cellular Uptake, Agglomeration and Ag Release. *Part Fibre Toxicol* (2014) 11:11. doi: 10.1186/1743-8977-11-11
19. Bolte S, Cordelières FP. A Guided Tour Into Subcellular Colocalization Analysis in Light Microscopy. *J Microsc* (2006) 224:213–32. doi: 10.1111/j.1365-2818.2006.01706.x
20. Manders EMM, Verbeek FJ, Aten JA. Measurement of Co Localization of Objects in Dual Colour Confocal Images. *J Microsc* (1993) 169:375–82. doi: 10.1111/j.1365-2818.1993.tb03313.x
21. Rorat A, Kachamakova-Trojanowska N, Jozkowicz A, Kruk J, Cocquerelle C, Vandenbulcke F, et al. Coelomocyte-Derived Fluorescence and DNA Markers of Composting Earthworm Species. *J Exp Zool* (2014) 321A:28–40. doi: 10.1002/jez.1834
22. Prochazková P, Roubalova R, Dvorák J, Navarro Pacheco NI, Bilej M. Pattern Recognition Receptors in Annelids. *Dev Comp Immunol* (2020) 102:103493. doi: 10.1016/j.dci.2019.103493
23. Dales RP, Kalaç Y. Phagocytic Defence by the Earthworm *Eisenia foetida* Against Certain Pathogenic Bacteria. *Comp Biochem Physiol* (1992) 101A:487–90. doi: 10.1016/0300-9629(92)90499-G
24. Cooper EL, Kauschke E, Cossarizza A. Digging for Innate Immunity Since Darwin and Metchnikoff. *BioEssays* (2002) 24:319–33. doi: 10.1002/bies.10077
25. Engelmann P, Cooper EL, Németh P. Anticipating Innate Immunity Without a Toll. *Mol Immunol* (2005) 42:931–42. doi: 10.1016/j.molimm.2004.09.038
26. Bodó K, Boros Á, Rumpler É, Molnár L, Böröcz K, Németh P, et al. Identification of Novel Lumbricin Homologues in *Eisenia andrei* Earthworms. *Dev Comp Immunol* (2019) 90:41–6. doi: 10.1016/j.dci.2018.09.001
27. Dvořák J, Roubalová R, Procházková P, Rossmann P, Škanta F, Bilej M. Sensing Microorganisms in the Gut Triggers the Immune Response in *Eisenia andrei* Earthworms. *Dev Comp Immunol* (2016) 57:67–74. doi: 10.1016/j.dci.2015.12.001
28. Hayashi Y, Engelmann P, Foldbjerg R, Szabó M, Somogyi I, Pollák E, et al. Earthworms and Humans *In Vitro*: Characterizing Evolutionarily Conserved Stress and Immune Responses to Silver Nanoparticles. *Environ Sci Technol* (2012) 46:4166–73. doi: 10.1021/es3000905
29. Daigneault M, Preston JA, Marriott HM, Whyte MKB, Dockrell DH. The Identification of Markers of Macrophage Differentiation in PMA-Stimulated THP-1 Cells and Monocyte-Derived Macrophages. *PLoS One* (2010) 5:e8668. doi: 10.1371/journal.pone.0008668
30. Madhvi A, Mishra H, Leisching GR, Mahlobo PZ, Baker B. Comparison of Human Monocyte Derived Macrophages and THP-1 Like Macrophages as *In Vitro* Models for *M. tuberculosis* Infection. *Comp Immunol Microbiol Infect* (2019) 67:101355. doi: 10.1016/j.cimid.2019.101355
31. Tedesco S, De Majo F, Kim J, Trenti A, Trevisi L, Fadini GP, et al. Convenience Versus Biological Significance: Are PMA-Differentiated THP-1 Cells Are Reliable Substitute for Blood-Derived Macrophages When Studying *In Vitro* Polarization? *Front Pharmacol* (2018) 9:71. doi: 10.3389/fphar.2018.00071
32. Pass DA, Morgan AJ, Read DS, Filed D, Weightman AJ, Kille P. The Effect of Anthropogenic Arsenic Contamination on the Earthworm Microbiome. *Environ Microbiol* (2015) 17:1884–96. doi: 10.1111/1462-2920.12712
33. Aira M, Bybee S, Pérez-Losada M, Dominguez J. Feeding on Microbiomes: Effects of Detritivory on the Taxonomic and Phylogenetic Bacterial Composition of Animal Manures. *FEMS Microbiol Ecol* (2015) 91:fiv117. doi: 10.1093/femsec/fiv117
34. Roubalová R, Procházková P, Hanč A, Dvořák J, Bilej M. Mutual Interactions of *E. andrei* Earthworms and Pathogens During the Process of Vermicomposting. *Envrion Sci Pollut Res Int* (2020) 27:33429–37. doi: 10.1007/s11356-019-04329-5
35. Kaito C, Murakami K, Imai L, Furuta K. Animal Infection Models Using non-Mammals. *Microbiol Immunol* (2020) 64:585–92. doi: 10.1111/1348-0421.12834
36. Schwende H, Fitzke E, Ambs P, Dieter P. Differences in the State of Differentiation of THP-1 Cells Induced by Phorbol Ester and 1,25-Dihydroxyvitamin D3. *J Leuk Biol* (1996) 59:555–61. doi: 10.1002/jlb.59.4.555
37. Köhlerová P, Beschin A, Šilerová M, De Baetselier P, Bilej M. Effect of Experimental Microbial Challenge on the Expression of Defense Molecules in *Eisenia foetida* Earthworm. *Dev Comp Immunol* (2004) 28:701–11. doi: 10.1016/j.dci.2004.01.001
38. Molnár L, Engelmann P, Somogyi I, Mácsik LL, Pollák E. Cold-Stress Induced Formation of Calcium and Phophorous Rich Chloragocyte Granules (Chloragosomes) in the Earthworm *Eisenia foetida*. *Comp Biochem Physiol* (2012) 163A:199–209. doi: 10.1016/j.cbpa.2012.06.005
39. Veiga E, Cossart P. The Role of Clathrin-Dependent Endocytosis in Bacterial Internalization. *Trends Cell Biol* (2006) 16:499–504. doi: 10.1016/j.tcb.2006.08.005
40. Shin JS, Abraham SN. Caveolae as Portals of Entry for Microbes. *Microb Infect* (2001) 3:755–61. doi: 10.1016/S1286-4579(01)01423-X
41. Starling D, Duncan R, Lloyd JB. The Role of Microtubules in Pinocytosis. Inhibition of Fluid-Phase Pinocytosis in the Rat Visceral Yolk Sac by Mitoclasic and Related Agents. *Cell Biol Int Rep* (1983) 7:593–602. doi: 10.1016/0309-1651(83)90113-3
42. Razani B, Woodman SE, Lisanti MP. Caveolae: From Cell Biology to Animal Physiology. *Pharmacol Rev* (2002) 54:431–67. doi: 10.1124/pr.54.3.431
43. Irie T, Fukunaga K, Pitha J. Hydroxypropylcyclodextrins in Parenteral Use. I: Lipid Dissolution and Effects on Lipid Transfers *In Vitro*. *J Pharm Sci* (1992) 81:521–3. doi: 10.1002/jps.2600810609
44. Kilsdonk EP, Yancey PG, Stoudt GW, Bangerter FW, Johnson WJ, Phillips MC, et al. Cellular Cholesterol Efflux Mediated by Cyclodextrins. *J Biol Chem* (1995) 270:17250–6. doi: 10.1074/jbc.270.29.17250
45. Ros-Baro A, Lopez-Iglesias C, Peiro S, Bellido D, Palacin M, Zorzano A, et al. Lipid Rafts Are Required for GLUT4 Internalization in Adipose Cells. *Proc Natl Acad Sci USA* (2001) 98:12050–5. doi: 10.1073/pnas.211341698
46. Perry DG, Daugherty GL, Martin WJ. Clathrin-Coated Pit-Associated Proteins Are Required for Alveolar Macrophage Phagocytosis. *J Immunol* (1999) 162:380–6.
47. Peterson JR, Mitchison TJ. Small Molecules, Big Impact: A History of Chemical Inhibitors and the Cytoskeleton. *Chem Biol* (2002) 9:1275–85. doi: 10.1016/s1074-5521(02)00284-3
48. Peiser L, Gough PJ, Kodama T, Gordon S. Macrophage Class A Scavenger Receptor-Mediated Phagocytosis of *Escherichia coli*: Role of Cell

Heterogeneity, Microbial Strain, and Culture Conditions *In Vitro*. *Infect Immun* (2000) 68:1953–63. doi: 10.1128/iai.68.4.1953-1963.2000
49. Kohler TP, Scholz A, Kiachludis D, Hammerschmidt S. Induction of Central Host Signaling Kinases During Pneumococcal Infection of Human THP-1 Cells. *Front Cell Infect Microbiol* (2016) 6:48. doi: 10.3389/fcimb.2016.00048
50. Ayscough KR. Endocytosis and the Development of Cell Polarity in Yeast Require a Dynamic F-Actin Cytoskeleton. *Curr Biol* (2000) 10:1587–90. doi: 10.1016/s0960-9822(00)00859-9
51. Fujimoto LM, Roth R, Heuser JE, Schmid SL. Actin Assembly Plays a Variable, But Not Obligatory Role in Receptor-Mediated Endocytosis in Mammalian Cells. *Traffic* (2000) 1:161–71. doi: 10.1034/j.1600-0854.2000.010208.x
52. Aderem A, Underhill DM. Mechanisms of Phagocytosis in Macrophages. *Ann Rev Immunol* (1999) 17:593–623. doi: 10.1146/annurev.immunol.17.1.593
53. Plows LD, Cook RT, Davies AJ, Walker AJ. Phagocytosis of *Lymnae stagnalis* Haemocytes: A Potential Role for Phosphatidylinositol 3-Kinase But Not Protein Kinase A. *J Inverteb Pathol* (2006) 91:74–7. doi: 10.1016/j.jip.2005.10.011
54. De Winter P, Rayne RC, Coast GM. The Effects of Intracellular Signaling Pathway Inhibitors on Phagocytosis by Haemocytes of *Manduca sexta*. *J Insect Physiol* (2007) 53:975–82. doi: 10.1016/j.jinsphys.2007.04.001
55. Michel T, Reichhart JM, Hoffmann JA, Royet J. *Drosophila* Toll Is Activated by Gram-Positive Bacteria Through a Circulating Peptidoglycan Recognition Protein. *Nature* (2001) 414:756–9. doi: 10.1038/414756a
56. Takeda K, Akira S. Toll Receptors and Pathogen Resistance. *Cell Microbiol* (2003) 5:143–53. doi: 10.1046/j.1462-5822.2003.00264.x
57. Beschin A, Bilej M, Hanssens F, Raymakers J, Van Dyck E, Revets H, et al. Identification and Cloning of a Glucan- and Ipopolysaccharide-Binding Protein From *Eisenia foetida* Earthworm Involved in the Activation of Prophenoloxidase Cascade. *J Biol Chem* (1998) 273:24948–54. doi: 10.1074/jbc.273.38.24948
58. Škanta F, Roubalová R, Dvořák J, Procházková P, Bilej M. Molecular Cloning and Expression of TLR in the *Eisenia andrei* Earthworm. *Dev Comp Immunol* (2013) 41:694–702. doi: 10.1016/j.dci.2013.08.009
59. Rosette C, Karin M. Cytoskeletal Control of Gene Expression: Depolymerization of Microtubules Activates NF-Kappa B. *J Cell Biol* (1995) 128:1111–9. doi: 10.1083/jcb.128.6.1111
60. Akira S, Takeda K, Kaisho T. Toll-Like Receptors: Critical Proteins Linking Innate and Acquired Immunity. *Nat Immunol* (2001) 2:675–80. doi: 10.1038/90609
61. Kawai T, Adachi O, Ogawa T, Takeda K, Akira S. Unresponsiveness of MyD88-Deficient Mice to Endotoxin. *Immunity* (1999) 11:115–22. doi: 10.1016/S1074-7613(00)80086-2
62. Kong L, Ge BX. MyD88-Independent Activation of a Novel Actin-Cdc42/Rac Pathway Is Required for Toll-Like Receptor-Stimulated Phagocytosis. *Cell Res* (2008) 18:745–55. doi: 10.1038/cr.2008.65
63. Brint EK, Xu D, Liu H, Dunne A, McKenzie ANJ, O'Neill LAJ, et al. ST2 Is an Inhibitor of Interleukin 1 Receptor and Toll-Like Receptor 4 Signaling and Maintains Endotoxin Tolerance. *Nat Immunol* (2004) 5:373–9. doi: 10.1038/ni1050
64. Janssens S, Burns K, Tschopp J, Beyaert R. Regulation of Interleukin-1- and Lipopolysaccharide-Induced NF-KappaB Activation by Alternative Splicing of Myd88. *Curr Biol* (2002) 12:467–71. doi: 10.1016/s0960-9822(02)00712-1
65. Liew FY, Xu D, Brint EK, O'Neill LAJ. Negative Regulation of Toll-Like Receptor-Mediated Immune Responses. *Nat Rev Immunol* (2005) 5:446–58. doi: 10.1038/nri1630
66. Yang M, Yuan S, Huang S, Li J, Xu L, Huang H, et al. Characterization of bbtTICAM From Amphioxus Suggests the Emergence of a MyD88-Independent Pathway in Basal Chordates. *Cell Res* (2011) 21:1410–23. doi: 10.1038/cr.2011.156
67. Mikami Y, Fukushima A, Kuwada-Kusunose T, Sakurai T, Kitano T, Komiyama Y, et al. Whole Transcriptome Analysis Using Next-Generation Sequencing of Sterile-Cultured *Eisenia andrei* for Immune System Research. *PLoS One* (2015) 10:e0118587. doi: 10.1371/journal.pone.0118587
68. Marek LR, Kagan JC. Phosphoinositide Binding by the Toll Adaptor Dmyd88 Controls Antibacterial Responses in Drosophila. *Immunity* (2012) 36:612–22. doi: 10.1016/j.immuni.2012.01.019
69. Balakrishnan A, Schnare M, Chakravortty D. Of Men Not Mice: Bactericidal/Permeability-Increasing Protein Expressed in Human Macrophages Acts as a Phagocytic Receptor and Modulates Entry and Replication of Gram-Negative Bacteria. *Front Immunol* (2016) 7:455. doi: 10.3389/fimmu.2016.00455
70. Škanta F, Procházková P, Roubalová R, Dvořák J, Bilej M. LBP/BPI Homologue in *Eisenia andrei* Earthworms. *Dev Comp Immunol* (2016) 54:1–6. doi: 10.1016/j.dci.2015.08.008
71. Bodó K, Ernszt D, Németh P, Engelmann P. Distinct Immune-and Defense-Related Molecular Fingerprints in Sepatated Coelomocyte Subsets of *Eisenia andrei* Earthworms. *Inv Surv J* (2018) 15:338–45. doi: 10.25431/1824-307X/isj.v15i1.338-345

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 08 September 2021
doi: 10.3389/fpls.2021.719549

High-Throughput Sequencing Reveals the Regulatory Networks of Transcriptome and Small RNAs During the Defense Against *Marssonina brunnea* in Poplar

*Yangwenke Liao, Qingyue Zhang, Rongrong Cui, Xin Xu, Fuyuan Zhu, Qiang Cheng and Xiaogang Li**

Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing, China

OPEN ACCESS

Edited by:
Bostjan Kobe, The University of Queensland, Australia

Reviewed by:
Hikmet Budak, Montana Bioagriculture, Inc., United States
Suresh Pokhrel, Donald Danforth Plant Science Center, United States

***Correspondence:**
Xiaogang Li
xgli@njfu.edu.cn

Specialty section:
This article was submitted to Plant Pathogen Interactions, a section of the journal Frontiers in Plant Science

Received: *02 June 2021*
Accepted: *05 August 2021*
Published: *08 September 2021*

Citation:
Liao Y, Zhang Q, Cui R, Xu X, Zhu F, Cheng Q and Li X (2021) High-Throughput Sequencing Reveals the Regulatory Networks of Transcriptome and Small RNAs During the Defense Against Marssonina brunnea in Poplar. Front. Plant Sci. 12:719549. doi: 10.3389/fpls.2021.719549

MicroRNAs are implicated in the adjustment of gene expression in plant response to biotic stresses. However, the regulatory networks of transcriptome and miRNAs are still poorly understood. In the present study, we ascertained the induction of genes for small RNA biosynthesis in poplar defense to a hemibiotrophic fungus *Marssonina brunnea* and afterward investigated the molecular regulatory networks by performing comprehensive sequencing analysis of mRNAs and small RNAs in *M. brunnea*-inoculated leaves. Differentially expressed genes in *M. brunnea*-infected poplar are mainly involved in secondary metabolisms, phytohormone pathways, the recognition of pathogens, and MAPK pathway in the plant, with real-time quantitative PCR (qPCR) validating the mRNA-seq results. Furthermore, differentially expressed miRNAs, such as MIR167_1-6, MIR167_1-12, MIR171_2-3, MIR395-13, MIR396-3, MIR396-16, MIR398-8, and MIR477-6, were identified. Through psRobot and TargetFinder programs, MIR167-1-6, MIR395-13, MIR396-3, MIR396-16, and MIR398-8 were annotated to modulate the expression of genes implicated in transportation, signaling, and biological responses of phytohormones and activation of antioxidants for plant immunity. Besides, validated differentially expressed genes involved in lignin generation, which were *phenylalanine ammonia-lyase*, *ferulate-5-hydroxylase*, *cinnamyl alcohol dehydrogenase,* and *peroxidase 11*, were selected as targets for the identification of novel miRNAs. Correspondingly, novel miRNAs, such as Novel MIR8567, Novel MIR3228, Novel MIR5913, and Novel MIR6493, were identified using the Mireap online program, which functions in the transcriptional regulation of lignin biosynthesis for poplar anti-fungal response. The present study underlines the roles of miRNAs in the regulation of transcriptome in the anti-fungal response of poplar and provides a new idea for molecular breeding of woody plants.

Keywords: poplar, microRNA, transcriptome, regulatory networks, plant-fungi interaction, *Marssonina brunnea*

INTRODUCTION

The plants confront variable biotic stresses on account of their immobility, such as the challenge of viruses, actinomyces, bacteria, fungi, nematodes, and pests (Chisholm et al., 2006; Eyles et al., 2010; Cagirici et al., 2017). During the response to pathogen infection, the plant can initiate the recognition of pathogen-associated molecular patterns (PAMPs)-triggered immunity and then induce alternations in proteins, protein kinases, and transcription factors, leading to the regulations of a great number of functional genes (Torres et al., 2006; Sturrock et al., 2011; Bilir et al., 2019). Simultaneously, plant own nucleotide-binding domain and leucine-rich repeat-containing proteins (NB-LRRs), which specialize in perceiving pathogen effectors and booting up effector-triggered immunity, resulting in hypersensitive reactions and activation of *R* genes (Torres et al., 2006). Hence, the expression of quantitative genes altered in plant cells in response to pathogen infection, while exploring the relationships between these genes and regulation of them is still challenging.

MicroRNAs (miRNAs) 21-24 nucleotides (nt) in size, which are small endogenous non-coding single-stranded RNAs synthesized by an RNA silencing system (Bhogireddy et al., 2021), play momentous roles in modulating the gene expression in the plant at the post-transcriptional level (Waterhouse and Hellens, 2015). An increasing number of reports have demonstrated that miRNAs regulate the growth, development, and stress responses in the plant by binding reverse complementary sequences and afterward bringing about the cleavage and/or translational inhibition of target mRNAs (Brant and Budak, 2018). The previous studies showed a bunch of differentially expressed miRNAs in plant species, such as cotton, cucumber, and tomato, during the interaction with pathogens, *via* high-throughput sequencing of small RNAs (Jin et al., 2012; Wang et al., 2018; Bilir et al., 2019). However, the regulatory mechanisms of miRNAs-adjusted genes during plant-pathogen interaction remain unclear.

Poplar (*Populus* sp.), an important forestry species growing fast and possessing high biomass, has become a model organism for tree research in recent years because of its broad distribution, genotypic diversity, and suitability for molecular analysis (Jansson and Douglas, 2007). The black spot disease caused by a hemibiotrophic fungus *Marssonina brunnea*, resulting in early defoliation, is a common poplar disease (Han et al., 2000). *M. brunnea* f. sp. *monogermtubi* can compatibly interact with certain plants from the *Leuce* section in the *Populus* genus and *M. brunnea* f. sp. *multigermtubi* can compatibly interact with some species from the *Aigeiros* section, which are two different specialized strains of *M. brunnea* identified in China (Zhang et al., 2018). Only several miRNAs have been mined involved in poplar defense to fungi pathogens (Chen et al., 2012; Chen and Cao, 2015; Su et al., 2018). In the present study, we performed high-throughput sequencing to analyze the changes in transcriptome profiles and small RNAs in *M. brunnea*-inoculated poplar leaves *in vitro*. In the current study, the experiments uncovered the regulatory networks of transcriptome and miRNAs in poplar defense to fungal infection and provided deeper insight into the interaction between trees and fungal pathogen.

MATERIALS AND METHODS

Plant Material and Fungus

Stem cuttings of Nanlin895 (*Populus deltoides* × *Populus euramericana* "Nanlin 895"), part resistant to *M. brunnea* f. sp. *multigermtubi* strain (Chen et al., 2015), were subcultured *in vitro* to sprout, in Murashige and Skoog (MS) medium supplemented with naphthalene acetic acid (NAA, 0.4 mg L^{-1}), 6-benzylaminopurine (6-BA, 4.0 mg L^{-1}), thidiazuron (TDZ, 0.04 mg L^{-1}), sucrose (25 g L^{-1}), and agar (5.5 g L^{-1}); and then transferred to MS medium added with NAA (0.2 mg L^{-1}), 6-BA (4.0 mg L^{-1}), sucrose (25 g L^{-1}), and agar (5.5 g L^{-1}), to grow into seedlings with 2–3 cm height. The culture condition was as followed: day/night temperature of 23/23°C, a 16/8 h light/dark period, and illumination intensity of 150 μmol m^{-2} s^{-1}. Poplar seedlings subsequently grew in sterile nutrient soil-vermiculite (3:1) mix under day/night temperature of 26/20°C, with a 16/8 h light/dark period, and a 68–73% relative humidity. Pots were put in plastic trays used for sub-irrigation, and plants were supplied with water one time per week. Then, 7–10 days later, leaves were detached for inoculation experiments.

The fungus *M. brunnea* f. sp. *multigermtubi* strain was offered by Dr. Qiang Cheng and Dr. Qin Xiong (Nanjing Forestry University) and preserved on potato dextrose agar medium (PDA) at 4°C using slant tubes, which was isolated according to Cheng et al. (2010) from *P. canadensis* in the campus of Nanjing Forestry University and identified morphologically and molecularly according to Xiong et al. (2019). We smeared the preserved *M. brunnea* on PDA in a sterile Petri dish and then maintained the dishes at 25–28°C in the dark for 20 days until conidia elution and inoculation.

Inoculation of Poplar Leaves With Fungus and Sample Collection

We eluted the conidia incubated on PDA with sterile ddH_2O and adjusted their density to 6×10^6 conidia L^{-1} for the inoculum, referring to the method of Liao et al. (2020). Afterward, we collected fully expanded healthy poplar leaves from seedlings, put them on a moist filter paper in Petri dishes, and dropped the spore suspension (30 μL) onto the abaxial leaf surface. The Petri dishes containing inoculated leaves were sealed and incubated in the dark at 25–28°C and relative humidity of 100%, and the disease symptoms on leaves were surveilled from 0 to 18 days postinoculation (dpi) with *M. brunnea*. As a negative control, mock inoculations were performed using sterile ddH_2O alone.

The inoculated leaves were harvested at 3, 6, 12, 24, and 48 h post inoculation (hpi) to analyze the expression of genes related to the miRNA biosynthesis, at 6 and 24 hpi to sequence mRNAs and small RNAs, and at 3, 6, 10, and 18 dpi to examine the disease symptoms of *M. brunnea* on poplar leaf. One biological sample was obtained by pooling three to four leaves, and three biological replicates were analyzed for each treatment. The experiments were independently performed four times.

Detection of Fungal Infection

Marssonina brunnea infection status was evaluated at 3, 6, 10, and 18 dpi by aniline staining and microscopy, according to the study of Liao et al. (2020). Briefly, the inoculated leaf section was cut into 1 × 1 cm piece, immersed in 0.15% saturated trichloroacetic acid bleaching liquid (solvent, 3:1, alcohol:chloroform solution) for 12 h for decoloration, and then stained in a mixed solution (saturated chloral hydrate and aniline blue, 5:1) at 58°C for 4 h. The stained leaf pieces were detected and photographed using a light microscope (BH200, SDPTOP, Sunny Instruments, Ningbo, China).

RNA Extraction and mRNA-Seq Analysis

The total RNA of leaf samples pooled at 6 and 24 hpi was extracted using HiPure HP Plant RNA Mini Kit (Magen, Guangzhou, China) for mRNA-seq analysis, as described in the instructions from the manufacturer. Next, mRNA was purified by the interaction of the poly (A) tails and magnetic oligo (dT) beads, and double-stranded cDNA was synthesized to construct the library. RNA sequencing services were provided by Personal Biotechnology Co., Ltd. (Shanghai, China). The cDNA libraries were examined by an Agilent High Sensitivity DNA Kit on Agilent 2100 Bioanalyzer (Agilent, St. Clara, CA, USA) with an average fragment length of 200–300 bp. The libraries were sequenced by Illumina NextSeq500 (Illumina Inc., San Diego, CA, USA) to generate paired-end reads with 150 bp length.

The transcriptome was analyzed following the procedure of Guo et al. (2017). The raw data were transformed into FASTQ format by the software of the sequencing platform. To get high-quality clean data, connectors and low-quality reads were filtered by Cutadapt (v1.16). Clean reads were used for mapping analysis by Bowtie2 (2.2.6) and Tophat2 (2.0.14), and the reference genome for gene annotation was *P. deltoides* WV94 V2.1 (https://phytozome-next.jgi.doe.gov/). HTSeq (0.9.1) and DESeq (1.30.0) were used for gene expression analysis at different expression levels (**Supplementary Table 2**). The *R* package, edgeR, was used to identify the differentially expressed genes (DEGs) (Robinson et al., 2010). The expression level of each unigene was calculated and normalized to generate FPKM. In the present study, the selection of DEGs was based on $|\log_2 \text{fold change}| \geq 1$ with p-value < 0.05. The threshold of the p-value was determined using a false discovery rate (FDR) in multiple tests.

Small RNA-Seq and Bioinformatics Analysis

The same RNA samples used for the mRNA-seq were used for the miRNA-seq. Small RNA libraries were constructed using NEB Next Multiplex Small RNA Library Prep Kit (New England Biolabs, Inc., Hitchin, UK), following the instructions of the manufacturer on an Illumina NextSeq500. The ligation of total RNA (10 μg) from 12 independent samples belonging to 4 groups, respectively, with 3′ adapters and 5′ adapters was performed using the T4 RNA ligase. Double-stranded cDNA was synthesized *via* RNA reverse-transcription using Superscript II Reverse Transcriptase. DNA fragments, which were enriched by PCR, were segregated by polyacrylamide electrophoresis (PAGE) gel. Next, the verification of fragment size and DNA library distribution was performed using Agilent 2100 for quality control of fragments enriched by PCR. After detecting the total concentration by Picogreen, the constructed libraries were single-end sequenced on the HiSeq 2500 platform at Personal Biotechnology Co., Ltd. (Shanghai, China).

The number of clean reads, with sequence lengths more than 18 nt and less than 36 nt, was counted, which were applied for small RNA analysis. The identical sequence in one single sample was deduplicated and the calculation of sequence abundance was carried out to obtain the unique reads, which were subsequently compared with the Rfam (14.0) database by BLAST. Four types of known non-coding RNAs (rRNA, tRNA, snRNA, and snoRNA) were screened, with a screening criterion of 0–2 mismatches. Unique reads which were not annotated to the above four types of non-coding RNAs were compared to mature miRNA sequences of the known miRNA in miRBase22 by BLAST, with screening criteria of 0–2 mismatches. Unique reads, not aligned with the Rfam and miBase databases, were compared to the genome to predict the novel miRNAs through the Mireap online program (http://sourceforge.net/projects/mireap/), using the default criteria of the program. Their secondary structure maps were drawn by RNA fold. Inter-sample correction of total reads was required for the standardization of expression amount, and then, the gene expression pattern of the sample was comprehensively investigated by count per million (CPM). The DEM (differentially expressed miRNA) selection was based on $|\log2 \text{ fold change}| \geq 1$ and p-value < 0.05 using DESeq. Finally, sequences for target genes of DEMs were predicted by psRobot_tar and TargetFinder (Bo and Wang, 2005; Zhao et al., 2020). The analysis script of these two programs was shown in **Supplementary Table 7**. Target genes that met the criteria of both programs were sorted out for the subsequent discussion.

Real-Time Quantitative PCR Assay

For the real-time quantitative PCR (qPCR) of miRNA biosynthesis-associated gene expression in poplar response to *M. brunnea* and the validation of RNA-seq results, total RNA of leaves sampled at different time points was extracted using a plant RNA Extraction Kit (Tiangen, Beijing, China) and reverse-transcribed using a ReverTra Ace qPCR RT Kit (Toyobo, Osaka, Japan). Sequences of genes associated with miRNA biosynthesis and for verification of mRNA sequencing were mined from *P. deltoides* WV94 V2.1 (https://phytozome-next.jgi.doe.gov/). All primers used were shown in **Supplementary Table 6**. The qPCR was performed using the SYBR Green PCR MasterMix (Vazyme, Nanjing, China), and the cycling conditions were as follows: 95°C for 5 min, followed by 40 cycles of denaturation at 95°C for 15 s, and annealing at 60°C for 30 s. Fluorescent signals were collected during the 60°C step. To validate the amplification of a single product, a melt curve, with conditions as 95°C for 15 s, 60°C for 1 min, and 95°C for 15 s, was generated at the end of the PCR cycles using software offered by the Step One Real-Time PCR Detection System (ThermoFisher, Waltham, MA, USA). The cDNA used for the validation of small RNA-seq was reverse-transcribed from 1 μg of total RNA for each sample using

a miR-X miRNA First-Strand Synthesis Kit (Takara, Beijing, China), and the qPCR analysis was performed on the CFX Connect Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) at Personal Biotechnology Co., Ltd. (Shanghai, China), using the TB Green Premix Ex TaqII (Takara, Beijing, China). The *actin* gene and 5S were used as internal genes. The threshold cycle (Ct) value of the internal gene was subtracted from that of the gene of interest to obtain a ΔCt value. The Ct value of the mock-inoculated control sample was subtracted from the ΔCt value to obtain a ΔΔCt value. The fold changes in expression level relative to the control were expressed as $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001).

RESULTS

Phenotype After *M. brunnea* Inoculation and miRNA Biosynthetic Gene Expression in Local Poplar Leaf

We first inoculated the abaxial surface of Nanlin895 leaf with *M. brunnea* and monitored the disease-developing process. Aniline blue staining revealed that *M. brunnea* successfully infected poplar, with secondary hyphae, emerged 3 dpi and then spread, which then significantly grew and developed to the whole leaf at 18 dpi (**Figure 1A**). Further, disease symptoms were photographed at corresponding time points. No distinct phenotype appeared at 3 dpi. The black spots merged and formed a dark brown imprint at 18 dpi (**Figure 1B**). This suggested that *M. brunnea* can infect Nanlin895 plants and compatibly interacts with this species.

To estimate the initiation of miRNA expression and determine appropriate sampling time points for mRNA and miRNA expression profiling, we tested the dynamic responses of critical genes involved in small RNA biosynthesis at 3, 6, 12, 24, and 48 hpi (**Figure 1C**). *RDR2* and *RDR6* encoding RNA-dependent RNA polymerases, play central roles in 24-nt-long repeat-associated small RNAs (rasiRNAs) and small interfering RNAs (siRNAs) biosynthesis; *DCL1* encoding Dicer-like (DCL) enzymes is critical in processing miRNAs, and *AGO1d* encoding the Argonaute (AGO) protein is responsible for the RNA-induced silencing complex (RISC) formation, which regulates expression of target mRNAs in aspects of plant biology; *DCL4* and *DCL2* function in producing siRNAs; *AGO4* serves in the loading of rasiRNAs; and *DRB1* encoding dsRNA-binding protein is a cofactor of DCL (Bhogireddy et al., 2021). All the above genes were significantly triggered by *M. brunnea* inoculation before 48 hpi. Compared with mock, the transcript level of *RDR2* increased to 1.85-, 1.98-, and 2.14-fold, respectively, at 3, 6, and 24 hpi, and *RDR6* expression was, respectively, upregulated to 5.30- and 6.88-fold at 6 and 48 hpi; the transcript level of *DCL1* increased to 1.98-fold at 12 hpi and the expression of *DCL2* elevated to 1.78- and 2.67-fold, respectively, at 3 and 12 hpi, while the transcript level of *DCL4* reached the peaks of 5.99- and 2.89-fold, respectively, at 3 and 24 hpi; *DRB1* showed apparent increases to 6.71-fold at 3 hpi; and *M. brunnea* inoculation strongly enhanced *AGO1d* transcription to 7.31-fold and *AGO4* to 3.52-fold, respectively, at 3 and 12 hpi (**Figure 1C**). These results indicate that *M. brunnea* infection can rapidly activate the RNA silencing system to synthesize small RNAs in poplar leaves before 24 hpi.

Functional Genes Expression of *M. brunnea*-Inoculated Poplar Leaves Based on High-Throughput Sequencing

To examine the response of functional genes expression to *M. brunnea* infection, we performed high-throughput sequencing of mRNAs in poplar leaves at 6 and 24 hpi. After the removal of adapter sequences and low-quality reads, a total of 40893478/41398298/38410968, 39250688/41064414/39597552, 40176520/39938864/41793126, and 46429992/44517684/46841894 mRNA sequences were obtained for Mk_6h_1/2/3 (three replicates of mock treatment at 6 hpi), Mb_6h_1/2/3 (*M. brunnea* inoculation at 6 hpi), Mk_24h_1/2/3 (Mock at 24 hpi), and Mb_24h_1/2/3 (*M. brunnea* inoculation at 24 hpi), respectively. The proportion of clean reads mapped to the *P. deltoides* genome in each library was above 88.5%, and the proportion of uniquely mapped reads ranged from 91.41 to 92.35% (**Supplementary Table 1**), with normalized gene expression (FPKM) of all clean reads presented in **Supplementary Table 2**. A principal component analysis (PCA) was performed to compare the transcriptome characteristics of all samples intuitionally (**Supplementary Figure 1**). The samples from two-time points were significantly divided and the replicates were closely spaced, while the samples at 6 hpi were not isolated between mock and *M. brunnea* inoculation, indicating that more genes were differentially expressed at 24 hpi rather than 6 hpi with fungal pathogen.

Furthermore, we found 4,976 (2,396 up, 2,580 down) DEGs in the group of Mk_24h vs. Mk_6h (Mk_24h/Mk_6h), 996 (646 up, 350 down) DEGs in the Mb_6h/Mk_6h group, 3,936 (1,758 up, 2,178 down) DEGs in the Mb_24h/Mb_6h group, and 1,923 (561 up, 1,362 down) DEGs in the Mb_24h/Mk_24h group (**Supplementary Figure 2A**), with the clustering analysis of DEGs expression pattern in a heatmap (**Supplementary Figure 2C**). Despite the most (4,976) DEGs exhibited in Mk_24h/Mk_6h group, only 490 of them were differentially expressed in the Mb_6h/Mk_6h comparison and the rest were due to the sampling time point. Besides, 1,914 DEGs were overlapped in Mk_24h/Mk_6h and Mb_24h/Mb_6h comparisons, attributed to the time point (**Supplementary Figure 2B**). Both the Venn diagram and the heatmap showed intuitional time-specific expression patterns in this study. Further, there were 412 (41%) DEGs in Mb_6h/Mk_6h group excluded from Mk_24h/Mk_6h and Mb_24h/Mk_24h comparison and 1,073 (56%) DEGs in Mb_24h/Mk_24h group excluded from Mb_24h/Mb_6h and Mb_6h/Mk_6h comparison (**Supplementary Figure 2B**). These data certified the effect of *M. brunnea* inoculation on mRNAs and the reliability of transcriptome analysis in the present study.

To better study the highlighted pathway in response to *M. brunnea* infection in poplar, we performed a Kyoto Encyclopedia

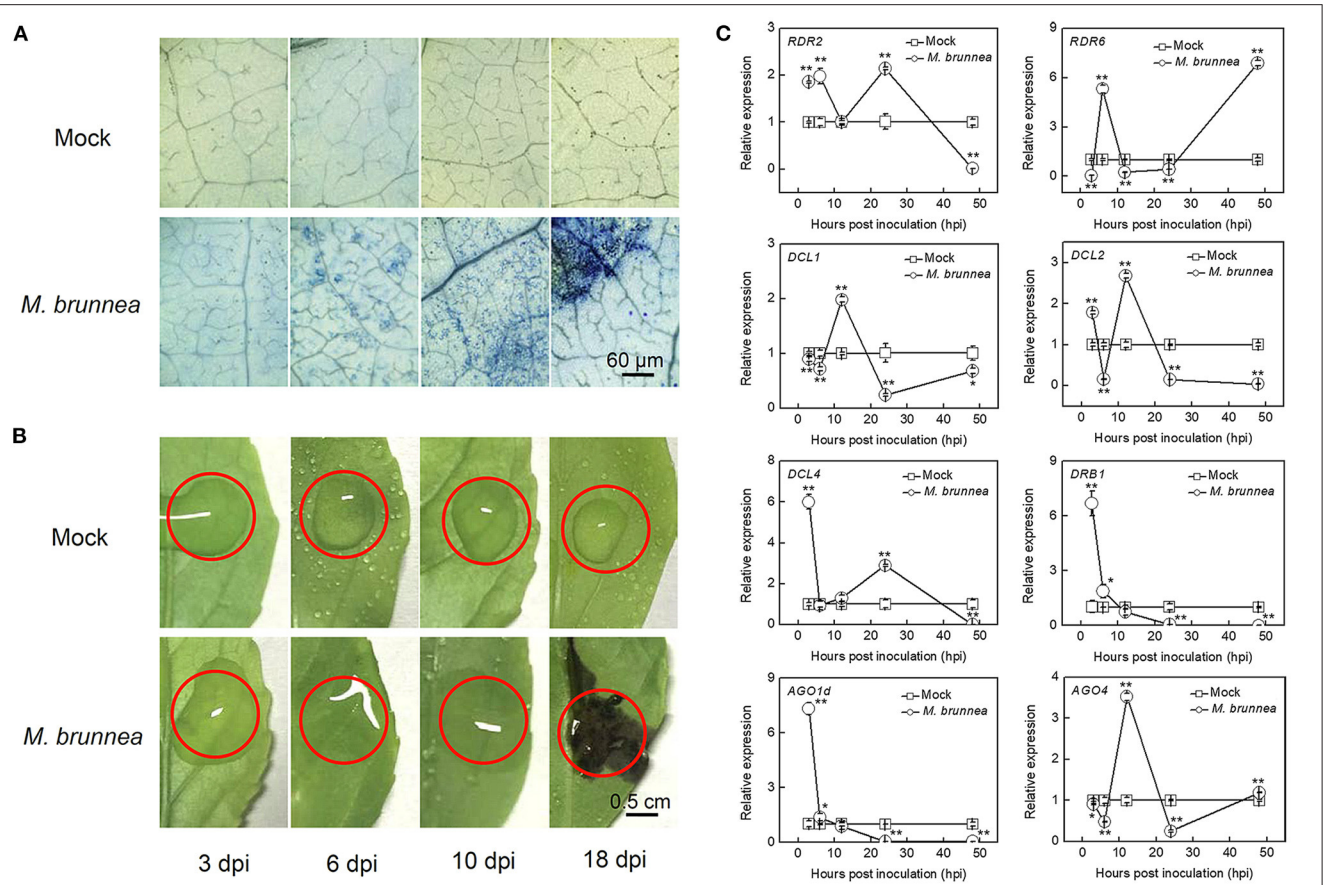


FIGURE 1 | Disease symptoms and gene expression of poplar leaves at different time points post *Marssonina brunnea* inoculation. **(A)** Light micrographs of leaf pieces stained with aniline blue and **(B)** photographs at 3, 6, 10, and 18 days postinoculation (dpi). Area within red cycle in **(B)** is inoculated; bar = 60 μm in **(A)** or 0.5 cm in **(B)**. **(C)** The expression of microRNA biosynthetic genes at 3, 6, 12, 24, and 48 h postinoculation (hpi). Leaf samples were collected from the inoculated leaves with *M. brunnea* and mock inoculations were performed with sterile ddH_2O alone as a negative control. The square indicates mock treatment and the cycle represents *M. brunnea* inoculation, with positive and negative error bars. The results are presented as mean values ± SD; $n = 3$. Asterisks represent significant differences in gene expression between treatments of mock and *M. brunnea* inoculation, according to Duncan's test. $^*P < 0.05$; $^{**}P < 0.01$. RDR, RNA-dependent RNA polymerase; DCL, dicer-like protein; DRB, double-stranded RNA-binding protein; AGO, argonate protein.

of Genes and Genomes (KEGG) annotation analysis for the DEGs. The top 20 enriched KEGG pathways of each comparison were showcased in **Figure 2**. Among the DEGs of Mk_24h/Mk_6h group, "Sulfur metabolism," (pop00920) "Carbon fixation and photosynthetic organisms," (pop00710) "Photosynthesis-antenna proteins," (pop00196) and "α-Linolenic acid metabolism" (pop00592) ranked from the first to fourth, followed by "Phenylpropanoid biosynthesis," (pop00940) "Starch and sucrose metabolism," (pop00500) "Cysteine and methionine metabolism," (pop00270) and "Amino sugar and nucleotide sugar metabolism." (pop00520) (**Figure 2A**) Among the DEGs of Mb_6h/Mk_6h, "Flavonoid biosynthesis" (pop00941) ranked the first, followed by "Circadian rhythm," (pop04712) "Alanine, aspartate, and glutamate metabolism," (pop00250) "Glycine, serine, and threonine metabolism," (pop00260) and "Ascorbate and aldarate metabolism." (pop00053) "Tryptophan metabolism" (pop00380) was also apparently enriched in the Mb_6h/Mk_6h DEGs (**Figure 2B**). Among the DEGs of Mb_24h/Mk_24h, "Plant-pathogen interaction" (pop04626) was the most significantly changed pathway, followed by "MAPK signal pathway," (pop04016) "Plant hormone signal transduction," (pop04075) "Diterpenoid biosynthesis," (pop00904) and "Starch and sucrose metabolism." (pop00500) A significant alternation also emerged in "Phenylpropanoid biosynthesis" (pop00940) (**Figure 2C**). Besides, "Plant hormone signal transduction," (pop04075) "Phenylpropanoid biosynthesis," (pop00940) and "MAPK signaling pathway" (pop04016) ranked high among the Mb_24h/Mb_6h DEGs (**Figure 2D**). More inhibited genes appeared in these pathways than induced genes. Finally, 26 genes belonging to seven pathways involved in the poplar response to fungal attack were focused. The selected pathways were "Flavonoid biosynthesis," (pop00941) "Phenylpropanoid biosynthesis," (pop00940) "Tryptophan metabolism," (pop00380) "Plant-pathogen interaction," (pop04626) "MAPK signal pathway," (pop04016) "Plant hormone signal transduction," (pop04075) and "Diterpenoid biosynthesis," (pop00904) with

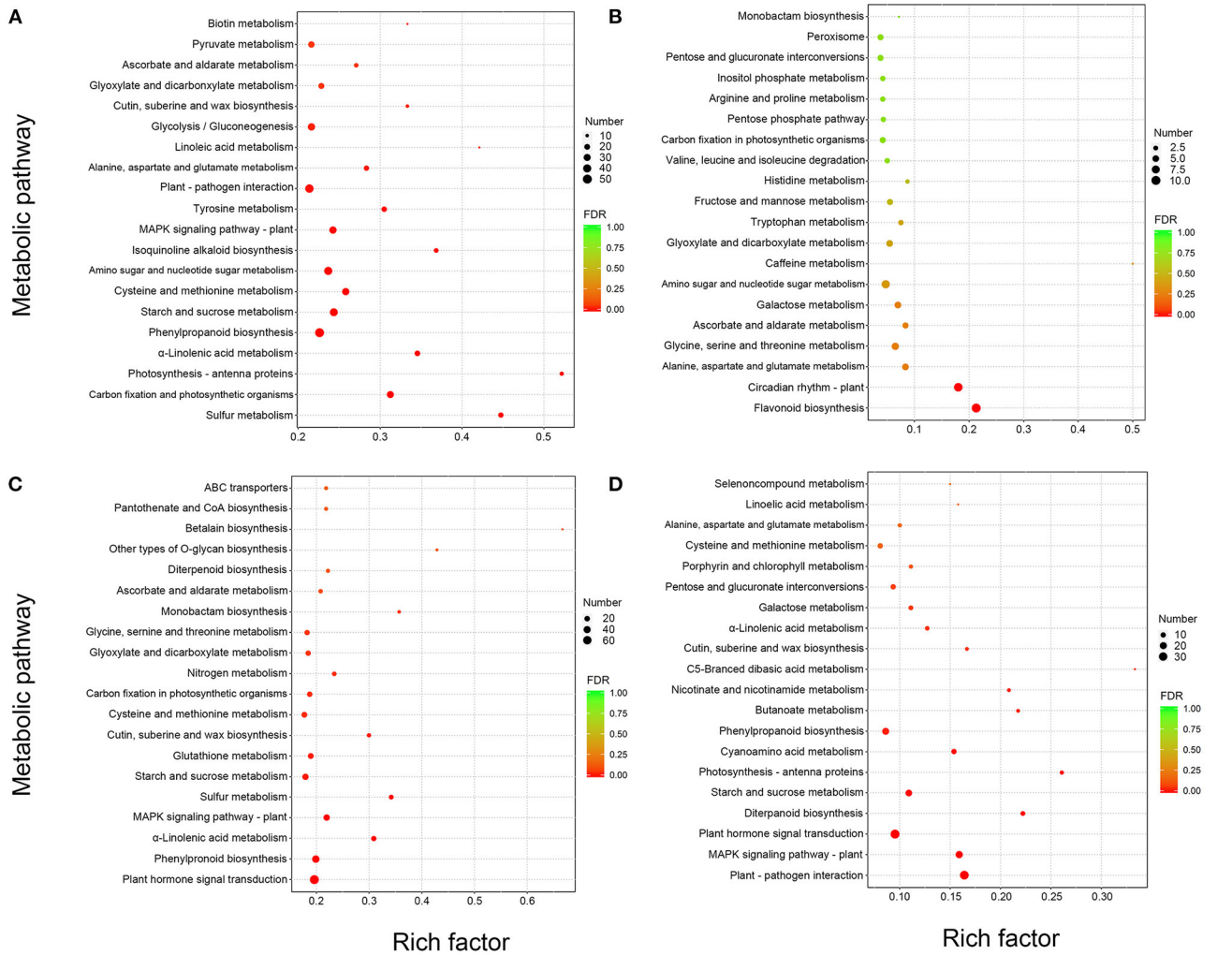


FIGURE 2 | Enrichment analysis of all DEGs. The top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with the most significant enrichment of **(A)** Mk_24h/Mk_6h, **(B)** Mb_6h/Mk_6h, **(C)** Mb_24h/Mk_24h, and **(D)** Mb_24h/Mb_6h, accordingly. The "/" implies "vs." The diameter of the cycle is positively correlated with the number of DEGs, and the color of the cycle indicates the false discovery rate (FDR) value as the scale shows.

potential genes involved in poplar response to fungal infection as shown in **Table 1**.

To validate the reliability of RNA-seq results, 18 genes were selected for qPCR validation, with data shown in **Figure 3**. The expression of genes involved in the phenylpropanoid pathway, such as *shikimate O-hydroxycinnamoyltransferase* (*HCT*, Podel.19G001400), *phenylalanine ammonia-lyase* (*PAL*, Podel.10G229600), *cinnamyl alcohol dehydrogenase-related* (*CAD*, Podel.16G069500), and *β-glucosidase* (*bGL*, Podel.01G233500), were significantly upregulated by *M. brunnea* inoculation at 6 hpi with the fungal pathogen, while the expression of *PAL*, *CAD,* and *bGL* decreased at 24 hpi. Expression of *ferulate-5-hydroxylase* (*F5H*, Podel.07G018500) and *peroxidase 11* (*POD11*, Podel.10G135900, and encoding class III POD) increased at 24 hpi (Almagro et al., 2009), together with *HCT* transcription. With respect to flavonoid biosynthetic genes, the abundance of *chalcone synthase* (*CHS*, Podel.03G190100), *leucoanthocyanidin reductase* (*LAR*, Podel.15G053500), and *anthocyanidin synthase/leucoanthocyanidin dioxygenase* (*ANS/LDOX*, Podel.03G127400) exhibited at a very high level at 6 h post *M. brunnea* inoculation. At 24 hpi, *CHS* and *ANS* were expressed to a lower level, with an obvious increase in *LAR* transcription. The *resistance to Pseudomonas syringae 2 orthology* (*RPS2*, Podel.13G009000), encoding an NB-LRR protein functioning in the recognition of pathogen, was slightly upregulated at 6 hpi but downregulated at 24 hpi. After the *M. brunnea* inoculation, *auxin-responsive protein IAA 19* (*IAA19*, Podel.06G269000) from Aux/IAA family, a repressor of auxin signaling, was inhibited at 24 hpi; and one gene from *Small Auxin Up RNA* family (*SAUR*, Podel.01G127200) mediating polar auxin transport was also downregulated. The transcript level of *DELLA* (Podel.17G162900) was reduced at 24 hpi, while *phytochrome-interacting factor 3* (*PIF3*, Podel.14G115200) was induced both at 6 and 24 hpi. The expression of *TIFY 10A* (Podel.01G177300)

TABLE 1 | Potential genes may be involved in poplar response to *Marssonina brunnea*.

Category	Gene ID	Gene description	Regulation
Phenylpropanoid	Podel.19G001400	Shikimate O-hydroxycinnamoyltransferase (HCT)	↑
biosynthesis	Podel.10G229600	Phenylalanine ammonia-lyase (PAL)	↑
	Podel.07G018500	Ferulate-5-hydroxylase (CYP84A, F5H)	↑
	Podel.14G110800	Caffeic acid 3-O-methyltransferase (COMT)	↓
	Podel.16G069500	Cinnamyl alcohol dehydrogenase GroES-like domain (CAD)	↑
	Podel.10G135900	Peroxidase 11 (class III POD)	↑
	Podel.01G233500	β-glucosidase (bGL)	↓
Flavonoid	Podel.03G190100	Chalcone synthase (CHS)	↑
biosynthesis	Podel.15G053500	Leucoanthocyanidin reductase (LAR)	↑
	Podel.03G127400	Anthocyanidin synthase/ leucoanthocyanidin dioxygenase (ANS/LDOX)	↑
	Podel.04G144800	Flavonol synthase (FLS)	↓
Plant hormone	Podel.04G034400	Transport inhibitor response 1 (TIR1)	↑
signal transduction	Podel.06G269000	Auxin-responsive protein IAA 19 (IAA19)/ (Aux/IAA family)	↓
	Podel.01G127200	Small Auxin Up RNA family protein (SAUR)	↑
	Podel.14G021600	F-box protein GID2 (GID2, SLY1)	↑
	Podel.17G162900	DELLA protein	↓
	Podel.14G115200	Phytochrome-interacting factor 3 (PIF3)	↑
	Podel.01G177300	TIFY 10A, a JAZ domain-containing protein	↓
	Podel.02G196200	Basic/helix-loop-helix transcription factor 14 (bHLH14)	↑
Tryptophan metabolism	Podel.04G173600	Amidase (AMI)	↑
Diterpenoid biosynthesis	Podel.15G136600	Gibberellin-44 dioxygenase (GDOX)	↑
	Podel.04G116600	Cytochrome P450, family 82, subfamily G, polypeptide 1 (CYP82G1)	↑
Plant-pathogen	Podel.11G016000	RPM1-interacting protein 4 (RIN4)	↑
interaction	Podel.13G009000	Disease resistance protein resistance to *Pseudomonas syringae* 2 orthology, leucine-rich repeat-containing protein, RPS2	↑
MAPK signaling	Podel.09G036500	Protein phosphatase 2C 3-related	↑
pathway	Podel.05G267900	Catalase	↑

and the transcription factor *bHLH14* (Podel.02G196200) did not change at 6 hpi, while a dramatic decrease occurred in the former gene expression and an obvious increase emerged in the transcription of the latter gene at 24 hpi. Besides, *amidase* (*AMI*, Podel.04G173600) involved in tryptophan metabolism and free auxin production, and *gibberellin-44 dioxygenase* (*GDOX*, Podel.15G136600) functioning in gibberellin (GA) biosynthesis expressed to a higher extend at 6 hpi while to a lower level at 24 hpi. Therefore, the results of qPCR assays agreed with the transcriptome data (**Figure 3**). The qPCR results suggest that genes related to the biosynthesis of phenylpropanoid pathway-derived lignin were generally activated, with flavonoid biosynthetic genes strongly induced within 24 h post *M. brunnea* inoculation. Furthermore, the plant-fungi interaction and pathways of auxin, GA, and jasmonic acid (JA) were induced by a fungal infection.

Annotations of Known miRNAs and Identification of Differentially Expressed miRNAs in *M. brunnea*-Inoculated Poplar Leaves

Clean reads (≥18 nt) were produced by small RNA sequencing ranging from 11.59 to 26.05 million from 12 small RNA libraries (**Supplementary Table 3**). Most of the non-coding RNAs in total reads were ribosomal RNA (rRNA) while the unknown group contributed to the majority in unique reads (**Supplementary Figure 3**). Clean reads were BLAST searched against known mature miRNAs and pre-miRNAs of miRBase (version 22.0) to identify the known conserved miRNAs. According to the difference in expression ($|\text{fold change}| > 2$, $p\text{-value} < 0.05$) in the multi-sample group comparison, we found six known DEMs in the Mb_6h/Mk_6h group (3 miRNAs were upregulated and 3 miRNAs were downregulated), 24 DEMs in Mk_24h/Mk_6h comparison (6 miRNAs were upregulated and 18 miRNAs were downregulated), 20 DEMs in the Mb_24h/Mb_6h group (9 miRNAs were upregulated and 11 miRNAs were downregulated), and 4 DEMs in the Mb_24h/Mk_24h group (2 miRNAs were upregulated and 2 miRNAs were downregulated) (**Figure 4**). The Venn diagram showed that two DEMs were both included in Mk_24h/Mk_6h and Mb_6h/Mk_6h groups while the Mk_24h/Mk_6h and Mb_24h/Mk_24h comparisons did not share any DEM. Most of the DEMs in Mk_24h/Mk_6h can be attributed to the change of time point. In addition, only three DEMs were shared by Mk_24h/Mk_6h and Mb_24h/Mb_6h comparisons, and 10 DEMs of Mb_6h/Mk_6h group were different from that of Mb_24h/Mk_24h comparison, suggesting that *M. brunnea*

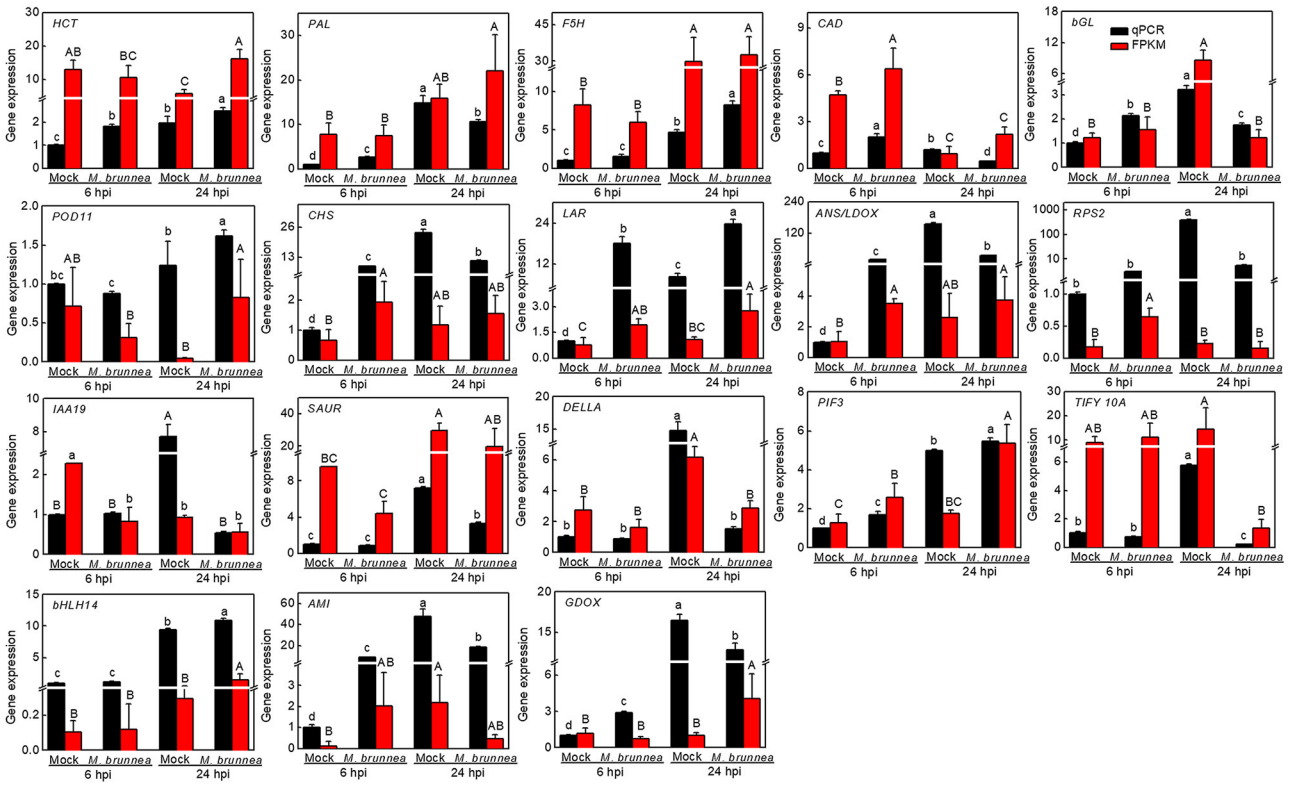


FIGURE 3 | Quantitative PCR (qPCR) verifies the quality of transcriptome sequencing using the $2^{-\Delta\Delta Ct}$ method. The black bar represents the qPCR validation result of mRNA in the sample, with positive and negative error bars; and the red bar represents the normalized result (FPKM) of the sample in the transcriptome. The results of qPCR are presented as mean values ± SD; $n = 3$. The results of FPKM are presented as mean values; $n = 3$. Lowercase letters indicate significant differences in qPCR results between treatments, with capital letters for FPKM values ($P < 0.05$).

inoculation caused diverse effect at 6 and 24 hpi on miRNAs in poplar leaves. The heatmap also presented intuitional time-specific expression patterns (**Supplementary Figure 4**).

The miRNAs complementarily paired bind target sites to regulate the transcription of target genes which we predicted using psRobot_tar and TargetFinder based on poplar genomic and transcriptomic databases. We selected the target genes that met the criteria of both programs.

Eight of these DEMs from Mb_6h/Mk_6h, Mb_24h/Mb_6h, and Mb_24h/Mk_24h groups (DEMs that we failed to predict target gene were excluded) and their target genes may be implicated in the poplar response to *M. brunnea* infection (**Table 2**), with a fold change of them shown in **Supplementary Table 5**.

Some of the selected DEMs targeted the same genes. For instance, both MIR396-3 and MIR396-16 engaged *GTP-binding* (Podel.02G196400) and *Serine protease inhibitor* (Podel.T131700) as targets. Besides, *Chaperone DnaJ-domain superfamily protein* (Podel.01G226500) and *Ribosomal RNA adenine dimethylase family protein* (Podel.14G078200) were also targeted by the same MIR396-3. *SNARE associated Golgi protein family* (Podel.19G075500), mainly involved in vesicle-associated membrane fusion, was targeted by MIR167_1-6. *ABC-2 type transporter family protein* (Podel.09G052200), primarily serving in plant hormones exportation, was the target of MIR395-13. *Tudor/PWWP/MBT domain-containing protein* (Podel.05G102600) was aimed by MIR398-8. However, Podel.18G116000, Podel.16G149200, and Podel.07G123500, which were target genes of MIR167_1-12, MIR171_2-3, and MIR477-6, respectively, were not annotated in the reference database.

Identification of Novel miRNAs and Their Targets

The MIREAP platform was hired to predict the precursors of new small miRNAs when the score was ≥ 2.2, with the RNAfold web server for a description of their secondary structures. Potential novel miRNAs were identified based on several important previously validated DEGs that we selected, and we identified four potential novel miRNAs (**Table 2**). The secondary structures predicted for the precursors of these candidate novel miRNAs are shown in **Supplementary Figure 5**, with red and blue asterisks indicating the 5′- and 3′-end positions of mature novel miRNA on the secondary structure of the precursor, respectively. Their expression patterns were analyzed along with their targets by qPCR verification (**Figure 5**). Compared to mock, the significant accumulation of Novel Pde-MIR8567 occurred at 6 hpi with no apparent change in abundance at 24 hpi, while the transcription

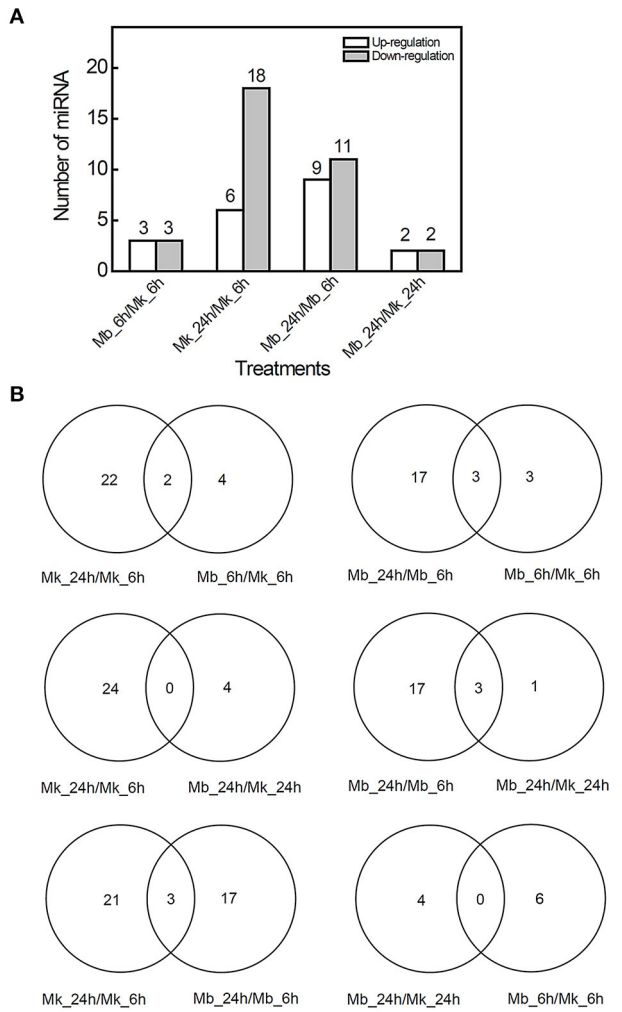


FIGURE 4 | Analysis of differential known microRNA expression in *M. brunnea*-infected poplar leaves. **(A)** The upregulated and downregulated ($|\log_2$ foldchange$| \geq 1$ and $P < 0.05$) amount and **(B)** the Venn diagram of differentially expressed known miRNAs (DEMs).

of its target gene exhibited an increase at 6 hpi but decrease at 24 hpi. Novel Pde-MIR3228 expressed to a higher extent at 6 hpi but showed no obvious alternation at 24 hpi, and the target gene was only induced at 24 hpi. With respect to Novel Pde-MIR6493, an induction emerged at 6 hpi rather than 24 hpi, while its target expression showed an increase at 24 hpi but no significant change at 6 hpi, relative to the transcript levels in mock. At 6 hpi, the expression of Novel Pde-MIR5913 was inhibited and its target was upregulated, while no obvious change occurred to the transcript level of this novel miRNA. However, no cleavage among the novel miRNAs and their targets was validated by qPCR assays (**Figure 5**).

DISCUSSION

Evidence indicated that miRNAs can be crucial in transcriptional control of the fungi pathogen-responsive genes during the infection process in various species, such as cotton, tomato, and cucumber, using high-throughput sequencing and bioinformatics tools (Jin et al., 2012; Wang et al., 2018; Bilir et al., 2019). However, miRNA-mRNA regulatory networks during poplar response to pathogenic fungi infection remain unknown. In the present study, we found the rapid activation of miRNA biosynthesis in *M. brunnea*-infected poplar leaves, implying miRNA biosynthesis was initiated to respond to the fungal infection at the early stage. Therefore, we further performed transcriptome analysis of *M. brunnea*-inoculated poplar leaf at 6 and 24 hpi. The results showed that genes involved in phenylpropanoid biosynthesis, flavonoid biosynthesis, plant hormone signal transduction, diterpenoid biosynthesis, tryptophan metabolism, plant-pathogen interaction, and MAPK signaling pathway were significantly induced in poplar response to *M. brunnea* infection. We further selected 26 candidate genes from DEGs, and the qPCR results confirmed the data of transcriptome analysis.

Genes associated with the phenylpropanoid pathway responsible for producing the three lignin monomers called monolignols, such as *PAL* (Podel.10G229600) and *CAD* (Podel.16G069500), were induced by *M. brunnea* at 6 hpi, and *HCT* (Podel.19G001400), *F5H* (Podel.07G018500), and *POD11* (Podel.10G135900), were significantly upregulated at 24 hpi; and *bGL* (Podel.01G233500) consuming the intermediate cinnamic acid was downregulated at 24 hpi (**Figure 3**). Our results were supported by previous studies (Chen et al., 2020; von Tiedemann et al., 2021). Thus, lignin biosynthesis is intensively accelerated in the poplar defense against a fungal pathogen. Notably, *caffeic acid 3-O-methyltransferase* (*COMT*, Podel.14G110800) exhibited downregulation after fungal inoculation (**Table 1** and **Supplementary Table 2**). HCT accounts for producing precursors of Guaiacyl and Sringyl unit lignin (Wagner et al., 2007), while COMT serves in the Sringyl unit lignin biosynthesis (Goujon et al., 2003). Poplar may prefer tamping cell wall with Guaiacyl unit lignin in defense to *M. brunnea*. Lignin formation can be initiated by the signal transduction of auxin and GA in plant immunity (Zhang et al., 2020). Accordingly, we found the enhancement of biosynthesis and signal transduction of auxin and GA (**Figure 3**). Although inhibition of auxin-responsive *SAUR* (Podel.01G127200) expression occurred, it may be attributed to the *AMI* (Podel.04G173600)-activated local auxin during the early responsive stage (Spartz et al., 2012).

The flavonoid biosynthesis branches at *p*-coumaroyl CoA from the phenylpropanoid pathway (Besseau et al., 2007). In the present study, *CHS* (Podel.03G190100), *LAR* (Podel.15G053500), and *ANS*/*LDOX* (Podel.03G127400) were explosively induced by fungal inoculation. Poplar may trigger flavonoids accumulation to generate antibiotic quinones, which further synthesizes polymeric compounds to form a protective barrier and scavenge reactive oxygen species to manage fungal invasion (Pourcel et al., 2007; Agati et al., 2012). On the other hand, the MAPK signal pathway was differentially expressed by *M. brunnea* infection in poplar, which also serves as a modifier of antioxidant systems (Xing et al., 2008; Meng and Zhang, 2013). Consistently, we found a higher transcript level of *protein phosphatase 2C* (Podel.09G036500) at 6 hpi, as well as

TABLE 2 | Potential known miRNAs, novel miRNAs, and the predicted targets may be involved in poplar response to *M. brunnea*.

MicroRNA	Mature sequence	Regulation	Target ID	Target description
Pde-MIR167_1-6	AGGTCATCTTGCAGCTTCAAT	↑	Podel.19G075500	SNARE associated Golgi protein family
Pde-MIR167_1-12	GGAAGCTGCCAGCATGATC	↑	Podel.18G116000	—
Pde-MIR171_2-3	TTGAGCCGTGCCAATATCACG	↑	Podel.16G149200	—
Pde-MIR395-13	GTTCCCTTGAGCACTTCA	↑	Podel.09G052200	ABC-2 type transporter family protein
Pde-MIR396-3	TTCCACAGCTTTCTTGAAC	↓	Podel.01G226500	Chaperone DnaJ-domain superfamily protein
			Podel.02G196400	GTP binding
			Podel.14G078200	Ribosomal RNA adenine dimethylase family protein
			Podel.T131700	Serine protease inhibitor, potato inhibitor I-type family protein
Pde-MIR396-16	TTCCACAGCTTTCTTGAACA	↓	Podel.02G196400	GTP binding
			Podel.T131700	Serine protease inhibitor, potato inhibitor I-type family protein
Pde-MIR398-8	GGAGCGACCTGGAATCACATG	↓	Podel.05G102600	Tudor/PWWP/MBT domain-containing protein
Pde-MIR477-6	ATCTCCCTCAAAGGCTTCCTC	↓	Podel.07G123500	—
Novel MIR3228	GCTGGGTTTATTTTTGAT	—	Podel.07G018500.1	Ferulate-5-hydroxylase (CYP84A, F5H)
Novel MIR8567	AGGGTATGGTCTGCATTGCTTTGA	—	Podel.10G229600.1	Phenylalanine ammonia-lyase (PAL)
Novel MIR5913	GATGCTGTGCCTCTGGCTAAT	—	Podel.16G069500.1	Cinnamyl alcohol dehydrogenase GroES-like domain (CAD)
Novel MIR6493	ACCGTCACACCCCAGAAGTG	—	Podel.10G135900.1	Peroxidase 11 (class III POD)

Tables with white shading exhibit the known miRNAs, while gray showing novel miRNAs.

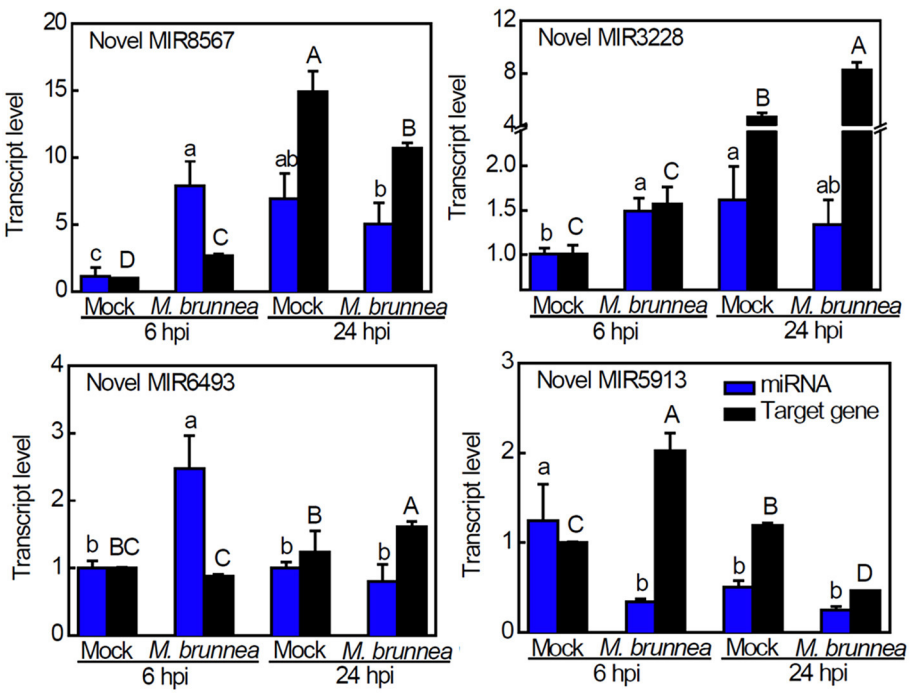


FIGURE 5 | Expression analysis of candidate novel miRNAs and targets using the $2^{-\Delta\Delta Ct}$ method. Expression analysis of novel miRNA and the predicted target gene in the sample. The blue bar represents the qPCR validation result of novel miRNA and the black bar stands for the qPCR result for the target gene. The results of qPCR are presented as mean values ± SD; $n = 3$. The results of FPKM are presented as mean values; $n = 3$. Lowercase letters indicate significant differences in qPCR results between treatments, with capital letters for FPKM values or the qPCR-validated target gene expression ($P < 0.05$).

catalase (Podel.05G267900) (**Table 1**). NB-LRRs can recognize specific effectors of pathogen and then activate MAPK cascade (Torres et al., 2006; Tsuda et al., 2013). In the present study, a slight increase in the expression of *RPS2* (Podel.13G009000) encoding an NB-LRR emerged at 6 hpi in poplar leaves. Taken together, fungal infection boots up a plant NB-LRR and MAPK

pathway and simultaneously generates flavonoids, leading to the enhancement of antioxidant systems in poplar defense.

To further reveal the miRNA-mRNA regulatory networks during the poplar defense to pathogenic fungi, the small RNA-seq was performed at the same time points post inoculation. Among 43 DEMs, we listed eight DEMs from Mb_6h/Mk_6h, Mb_24h/Mk_24h, and Mb_24h/Mk_24h comparisons and predicted four novel miRNAs, according to the DEGs we selected (**Table 2**). MIR167 can accumulate to manage the auxin signaling pathway by manipulating its target *ARF factor* after bacterial pathogen infection (Jodder et al., 2017), supporting our result that *M. brunnea* inoculation led to upregulation of MIR167_1-6 and MIR167_1-12. The target gene of MIR167_1-6, Podel.19G075500, encodes a soluble *N*-ethyl-maleimide sensitive factor attachment adaptor protein receptor (SNARE) domain associated Golgi protein which is mainly involved in vesicle-associated membrane fusion. Recent research proved that QA-SNARE SYP132, a low-abundant, secretory SNARE localizing to the plasma membrane, is tightly modulated by auxin and that abundant SYP132 caused a restriction in auxin-derived apoplast acidification and vegetative growth (Xia et al., 2019). In the present study, poplar may augment MIR167_1-6 to repress SNARE, amplifying the auxin responses to fungal infection. Accordingly, the transcriptome analysis confirmed the enhancement of auxin biosynthesis and signaling (**Table 1** and **Figure 3**).

MIR395 family is usually involved in stress-response processes. Augment in MIR395 led to suppression in WRKY26, thereby reducing the expression of some *PR* genes (Zhang et al., 2017). Our results exhibited that MIR395-13 accumulated after *M. brunnea* inoculation. The target gene *ABC-2 type transporter family protein* (Podel.09G052200) plays a key role in exporting phytohormones, such as salicylic acid (SA), jasmonic acid (JA), auxin, and strigolactone (Ye et al., 2013; Sasse et al., 2015; Dhara and Raichaudhuri, 2021). Increased MIR395-13 production might inhibit *ABC-2 type transporter* expression to block SA exportation, which thereby alleviated *PR* genes transcription in poplar response to pathogenic fungi (Dhara and Raichaudhuri,

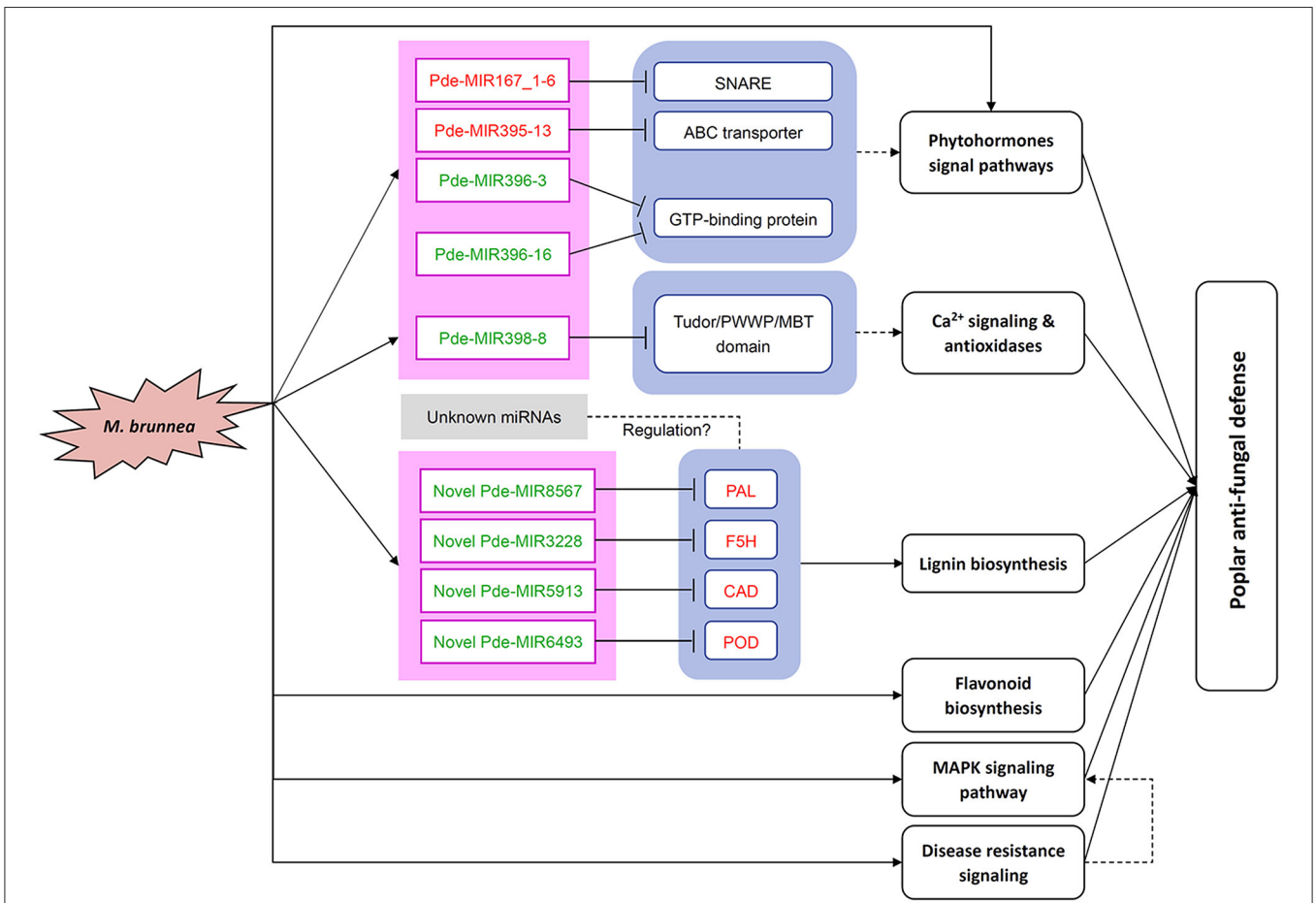


FIGURE 6 | Overview of miRNAs and genes regulation patterns in the poplar response to *M. brunnea*. The red letters are upregulated genes or miRNAs or pathways, and the green letters are downregulated genes or miRNAs or pathways. "→ " represents induction of mentioned elements or pathway, while "⊥" stands for suppression. A dashed line indicates speculated pathways. SNARE, soluble N-ethyl-maleimide sensitive factor attachment adaptor protein receptor; ABC transporter, ATP-binding cassette transporter; PAL, phenylalanine ammonia-lyase; CAD, cinnamyl alcohol dehydrogenase; POD, peroxidase; F5H, ferulate-5-hydroxylase; MAPK, mitogen-activated protein kinase.

2021). Furthermore, MIR396 can balance plant growth and disease resistance by interacting with growth-regulating factors in the plant. Blocking MIR396 expression resulted in higher grain yield while augment of MIR396 increased susceptibility to the blast fungus in rice (Chandran et al., 2019). In the present work, *M. brunnea* inoculation suppressed MIR396-3 and MIR396-16 expression. These two miRNAs coordinately regulate *GTP binding* (Podel.02G196400). During the early response to *M. brunnea*, abundant GTP-binding protein may rope JAZ domain-containing protein to provoke JA-signaling cascade in stomatal defense in poplar (Lee et al., 2018). Correspondingly, transcriptome analysis exposed the downregulation of the JA pathway inhibitor *TIFY 10A* (Podel.01G177300) expression by fungal infection (**Table 1** and **Figure 3**). The previous study demonstrating that Nanlin 895 poplar restricted the SA pathway and activated the JA pathway in defense to *M. brunnea* infection supported the findings in this study (Liao et al., 2020).

MIR398 serves to store $O_2^{\cdot-}$ *via* inhibiting antioxidant enzymes activity in plants under stress (He et al., 2021). The production of target gene *Tudor/PWWP/MBT domain-containing protein* (Podel.05G102600) can interact with histidine-rich calcium-binding protein (Yang et al., 2020). The suppression of MIR398-8 in this study may manipulate Ca^{2+} signaling to amplify antioxidant systems and restrict reactive oxygen species accumulation.

Noteworthily, four novel miRNAs were predicted, including Novel MIR8567, Novel MIR3228, Novel MIR5913, and Novel MIR6493 target to the upregulated lignin biosynthetic DEGs which are *PAL*, *F5H*, *CAD,* and *POD11*, respectively. All these target genes are critical genes to lignin biosynthesis (**Figure 6**). However, qPCR assays did not validate the cleavage among the novel miRNAs and their targets, implying the probable existence of other unknown miRNAs regulating the target genes above. Overall, miRNA-mRNA regulatory networks play pivotal roles in defense against fungal attacks in poplar. However, the detailed interactions and roles of the novel miRNAs were predicted and other unknown miRNAs in poplar defense to pathogen await further research.

CONCLUSION

In summary, global transcriptional profiles of mRNA and small RNAs were investigated in *M. brunnea*-infected poplar leaves, combined with qPCR validation. We obtained numerous DEGs and selected 26 critical fungal pathogen-responsive genes. The functional analysis demonstrated the involvement of these DEGs in lignin and flavonoid biosynthesis, phytohormone pathways, disease resistance signaling, and MAPK pathway. Further, we screen out DEMs and identified eight annotated miRNAs, four novel miRNAs, and their targets involved in poplar response to *M. brunnea* infection. By analyzing the function of the targets, we found that these miRNAs mediated plant hormones pathways, antioxidant systems enhancement, and lignin biosynthesis. According to the above analysis, we summarized these regulatory networks in **Figure 6**. The results of our research underline the role of miRNAs in poplar defense to fungal infection and provide a new idea for the molecular breeding of trees.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, PRJNA730976; https://www.ncbi.nlm.nih.gov/, PRJNA731035

AUTHOR CONTRIBUTIONS

YL and XL designed experiments. QZ, RC, and XX performed experiments. YL, QZ, RC, and FZ analyzed data. YL and XL wrote and revised the article. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China project (Grant No. 31600482) and the Natural Science Foundation of Jiangsu Province (Grant No. BK20190040). This work was supported by the Co-Innovation Center for Sustainable Forestry in Southern China of the College of Biology and Environment at Nanjing Forestry University, Nanjing, China.

ACKNOWLEDGMENTS

We thank Dr. Qiang Cheng and Dr. Qin Xiong (Nanjing Forestry University) for generously supplying us with the *M. brunnea* f. sp. *multigermtubi* strain.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.719549/full#supplementary-material

REFERENCES

Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014

Almagro, L., Ros, L. V. G., Belchi-Navarro, S., Bru, R., Barcelo, A. R., and Pedreno, M. A. (2009). Class III peroxidases in plant defence reactions. *J. Exp. Bot.* 60, 377–390. doi: 10.1093/jxb/ern277

Besseau, S., Hoffmann, L., Geoffroy, P., Lapierre, C., Pollet, B., and Legrand, M. (2007). Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. *Plant Cell* 19, 148–162. doi: 10.1105/tpc.106.044495

Bhogireddy, S., Mangrauthia, S. K., Kumar, R., Pandey, A. K., Singh, S., Jain, A., et al. (2021). Regulatory non-coding RNAs: a new frontier in regulation of plant biology. *Funct. Integr. Genomics* 18, 313–330. doi: 10.1007/s10142-021-00787-8

Bilir, O., Telli, O., Norman, C., Budak, H., Hong, Y. G., and Tor, M. (2019). Small RNA inhibits infection by downy mildew pathogen *Hyaloperonospora arabidopsidis*. *Mol. Plant Pathol.* 20, 1523–1534. doi: 10.1111/mpp.12863

Bo, X. C., and Wang, S. Q. (2005). TargetFinder: a software for antisense oligonucleotide target site selection based on MAST and secondary structures of target mRNA. *Bioinformatics* 21, 1401–1402. doi: 10.1093/bioinformatics/bti211

Brant, E. J., and Budak, H. (2018). Plant small non-coding RNAs and their roles in biotic stresses. *Front. Plant Sci.* 9:1038. doi: 10.3389/fpls.2018.01038

Cagirici, H. B., Biyiklioglu, S., and Budak, H. (2017). Assembly and annotation of transcriptome provided evidence of miRNA mobility between wheat and wheat stem sawfly. *Front. Plant Sci.* 8:1653. doi: 10.3389/fpls.2017.01653

Chandran, V., Wang, H., Gao, F., Cao, X. L., Chen, Y. P., Li, G. B., et al. (2019). MiR396-OsGRFs module balances growth and rice blast disease-resistance. *Front. Plant Sci.* 9:1999. doi: 10.3389/fpls.2018.01999

Chen, C. W., Yao, Y., Zhang, L., Xu, M. J., Jiang, J. P., Dou, T. H., et al. (2015). A comprehensive analysis of the transcriptomes of *Marssonina brunnea* and infected poplar leaves to capture vital events in host-pathogen interactions. *PLoS ONE* 10:e0134246. doi: 10.1371/journal.pone.0134246

Chen, L., Ren, Y. Y., Zhang, Y. Y., Xu, J. C., Zhang, Z. Y., and Wang, Y. W. (2012). Genome-wide profiling of novel and conserved *Populus* microRNAs involved in pathogen stress response by deep sequencing. *Planta* 235, 873–883. doi: 10.1007/s00425-011-1548-z

Chen, M., and Cao, Z. M. (2015). Genome-wide expression profiling of microRNAs in poplar upon infection with the foliar rust fungus *Melampsora larici-populina*. *BMC Genomics* 16:696. doi: 10.1186/s12864-015-1891-8

Chen, S., Zhang, Y., Zhao, Y., Xu, W., Li, Y., Xie, J., et al. (2020). Key genes and genetic interactions of plant-pathogen functional modules in poplar infected by *Marssonina brunnea*. *Mol. Plant Microbe Int.* 33, 1080–1090. doi: 10.1094/mpmi-11-19-0325-r

Cheng, Q. A., Cao, Y. Z., Jiang, C., Xu, L. A., Wang, M. X., Zhang, S. G., et al. (2010). Identifying secreted proteins of *Marssonina brunnea* by degenerate PCR. *Proteomics* 10, 2406–2417. doi: 10.1002/pmic.200900844

Chisholm, S. T., Coaker, G., Day, B., and Staskawicz, B. J. (2006). Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124, 803–814. doi: 10.1016/j.cell.2006.02.008

Dhara, A., and Raichaudhuri, A. (2021). ABCG transporter proteins with beneficial activity on plants. *Phytochemistry*. 184:112663. doi: 10.1016/j.phytochem.2021.112663

Eyles, A., Bonello, P., Ganley, R., and Mohammed, C. (2010). Induced resistance to pests and pathogens in trees. *New Phytol.* 185, 893–908. doi: 10.1111/j.1469-8137.2009.03127.x

Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., et al. (2003). A new *Arabidopsis thaliana* mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. *Plant Mol. Biol.* 51, 973–989. doi: 10.1023/a:1023022825098

Guo, D., Song, X. M., Yuan, M., Wang, Z. Y., Ge, W. N., Wang, L., et al. (2017). RNA-seq profiling shows divergent gene expression patterns in *Arabidopsis* grown under different densities. *Front. Plant Sci.* 8:2001. doi: 10.3389/fpls.2017.02001

Han, Z., Yin, T., Li, C., Huang, M., and Wu, R. (2000). Host effect on genetic variation of Marssonina brunnea pathogenic to poplars. *Theor. Appl. Genet.* 100, 614–620. doi: 10.1007/s001220050081

He, Y., Zhou, J. X., Hu, Y. F., Fang, C. Y., Yu, Y. J., Yang, J., et al. (2021). Overexpression of sly-miR398b increased salt sensitivity likely via regulating antioxidant system and photosynthesis in tomato. *Environ. Exp. Bot.* 181:104273. doi: 10.1016/j.envexpbot.2020.104273

Jansson, S., and Douglas, C. J. (2007). *Populus*: a model system for plant biology. *Annu. Rev. Plant Biol.* 58, 435–458. doi: 10.1146/annurev.arplant.58.032806.103956

Jin, W. B., Wu, F. L., Xiao, L., Liang, G. W., Zhen, Y. X., Guo, Z. K., et al. (2012). Microarray-based analysis of tomato miRNA regulated by *Botrytis cinerea*. *J. Plant Growth Regul.* 31, 38–46. doi: 10.1007/s00344-011-9217-9

Jodder, J., Basak, S., Das, R., and Kundu, P. (2017). Coherent regulation of miR167a biogenesis and expression of auxin signaling pathway genes during bacterial stress in tomato. *Physiol. Mol. Plant Pathol.* 100, 97–105. doi: 10.1016/j.pmpp.2017.08.001

Lee, S., Rojas, C. M., Oh, S., Kang, M., Choudhury, S. R., Lee, H. K., et al. (2018). Nucleolar GTP-binding protein 1-2 (NOG1-2) interacts with Jasmonate-ZIMDomain protein 9 (JAZ9) to regulate stomatal aperture during plant immunity. *Int. J. Mol. Sci.* 19:14. doi: 10.3390/ijms19071922

Liao, Y., Cui, R., Xu, X., Cheng, Q., and Li, X. (2020). Jasmonic acid- and ethylene-induced mitochondrial alternative oxidase stimulates anti-*Marssonina brunnea* defense in poplar. *Plant Cell Physiol.* 61, 2031–2042. doi: 10.1093/pcp/pcaa117

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Meng, X. Z., and Zhang, S. Q. (2013). MAPK cascades in Plant disease resistance signaling. *Annu. Rev. Phytopathol.* 51, 245–266. doi: 10.1146/annurev-phyto-082712-102314

Pourcel, L., Routaboul, J. M., Cheynier, V., Lepiniec, L., and Debeaujon, I. (2007). Flavonoid oxidation in plants: from biochemical properties to physiological functions. *Trends Plant Sci.* 12, 29–36. doi: 10.1016/j.tplants.2006.11.006

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

Sasse, J., Simon, S., Gubeli, C., Liu, G. W., Cheng, X., Frimi, J., et al. (2015). Asymmetric localizations of the ABC transporter PaPDR1 trace paths of directional strigolactone transport. *Curr. Biol.* 25, 647–655. doi: 10.1016/j.cub.2015.01.015

Spartz, A. K., Lee, S. H., Wenger, J. P., Gonzalez, N., Itoh, H., Inze, D., et al. (2012). The *SAUR19* subfamily of *SMALL AUXIN UP RNA* genes promote cell expansion. *Plant J.* 70, 978–990. doi: 10.1111/j.1365-313X.2012.04946.x

Sturrock, R. N., Frankel, S. J., Brown, A. V., Hennon, P. E., Kliejunas, J. T., Lewis, K. J., et al. (2011). Climate change and forest diseases. *Plant Pathol.* 60, 133–149. doi: 10.1111/j.1365-3059.2010.02406.x

Su, Y. Y., Li, H. G., Wang, Y. L., Li, S., Wang, H. L., Yu, L., et al. (2018). Poplar miR472a targeting NBS-LRRs is involved in effective defence against the necrotrophic fungus *Cytospora chrysosperma*. *J. Exp. Bot.* 69, 5519–5530. doi: 10.1093/jxb/ery304

Torres, M. A., Jones, J. D. G., and Dangl, J. L. (2006). Reactive oxygen species signaling in response to pathogens. *Plant Physiol.* 141, 373–378. doi: 10.1104/pp.106.079467

Tsuda, K., Mine, A., Bethke, G., Igarashi, D., Botanga, C. J., Tsuda, Y., et al. (2013). Dual regulation of gene expression mediated by extended MAPK activation and salicylic acid contributes to robust innate immunity in *Arabidopsis thaliana*. *PLoS Genet.* 9:e1004015. doi: 10.1371/journal.pgen.1004015

von Tiedemann, A., Koopmann, B., and Hoech, K. (2021). Lignin composition and timing of cell wall lignification are involved in *Brassica napus* resistance to *Sclerotinia sclerotiorum* stem rot. *Phytopathology*. doi: 10.1094/PHYTO-09-20-0425-R. [Epub ahead of print].

Wagner, A., Ralph, J., Akiyama, T., Flint, H., Phillips, L., Torr, K., et al. (2007). Exploring lignification in conifers by silencing hydroxycinnamoyl-CoA: shikimate hydroxycinnamoyltransferase in *Pinus radiata*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11856–11861. doi: 10.1073/pnas.0701428104

Wang, X. Y., Zhang, D., Cui, N., Yu, Y., Yu, G. C., and Fan, H. Y. (2018). Transcriptome and miRNA analyses of the response to *Corynespora cassiicola* in cucumber. *Sci. Rep.* 8:7798. doi: 10.1038/s41598-018-26080-6

Waterhouse, P. M., and Hellens, R. P. (2015). Coding in non-coding RNAs. *Nature* 520, 41–42. doi: 10.1038/nature14378

Xia, L. F., Marques-Bueno, M. M., Bruce, C. G., and Karnik, R. (2019). Unusual roles of secretory SNARE SYP132 in plasma membrane H^+-ATPase traffic and vegetative plant growth. *Plant Physiol.* 180, 837–858. doi: 10.1104/pp.19.00266

Xing, Y., Jia, W. S., and Zhang, J. H. (2008). AtMKK1 mediates ABA-induced *CAT1* expression and H_2O_2 production via AtMPK6-coupled signaling in *Arabidopsis*. *Plant J.* 54, 440–451. doi: 10.1111/j.1365-313X.2008.03433.x

Xiong, Q., Qian, Y., Zhang, C., Shi, N., and Zheng, X. (2019). First report of phytophthora hydropathica causing wilting and shoot blight on *Bixa orellana* in China. *Plant Dis.* 103, 163–164. doi: 10.1094/pdis-06-18-1013-pdn

Yang, J., Zhang, T., Mao, H. P., Jin, H. Q., Sun, Y. W., and Qi, Z. (2020). A *Leymus chinensis* histidine-rich Ca^{2+}-binding protein binds Ca^{2+}/Zn^{2+} and suppresses abscisic acid signaling in *Arabidopsis*. *J. Plant Physiol.* 252:11. doi: 10.1016/j.jplph.2020.153209

Ye, L. F., Liu, L., Xing, A. Q., and Kang, D. M. (2013). Characterization of a dwarf mutant allele of Arabidopsis MDR-like ABC transporter AtPGP1

gene. *Biochem. Biophys. Res. Commun.* 441, 782–786. doi: 10.1016/j.bbrc.2013.10.136

Zhang, Q. L., Li, Y., Zhang, Y., Wu, C. B., Wang, S. N., Hao, L., et al. (2017). Md-miR156ab and Md-miR395 target WRKY transcription factors to influence apple resistance to leaf spot disease. *Front. Plant Sci.* 8:52. doi: 10.3389/fpls.2017.0052

Zhang, Y. F., Tian, L. Y., Yan, D. H., and He, W. (2018). Genome-wide transcriptome analysis reveals the comprehensive response of two susceptible poplar sections to *Marssonina brunnea* infection. *Genes* 9:25. doi: 10.3390/genes9030154

Zhang, Y. S., Wang, Y. B., Ye, D. L., Xing, J. P., Duan, L. S., Li, Z. H., et al. (2020). Ethephon-regulated maize internode elongation associated with modulating auxin and gibberellin signal to alter cell wall biosynthesis and modification. *Plant Sci.* 290:16. doi: 10.1016/j.plantsci.2019.110196

Zhao, X., Yang, G., Liu, X., Yu, Z., and Peng, S. (2020). Integrated analysis of seed microRNA and mRNA transcriptome reveals important functional genes and microRNA-targets in the process of walnut (*Juglans regia*) seed oil accumulation. *Int. J. Mol. Sci.* 21:9093. doi: 10.3390/ijms21239093

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

MINI REVIEW
published: 23 September 2021
doi: 10.3389/fimmu.2021.752898

The SARM1 TIR NADase: Mechanistic Similarities to Bacterial Phage Defense and Toxin-Antitoxin Systems

Aaron DiAntonio[1,2*], *Jeffrey Milbrandt*[2,3] *and Matthew D. Figley*[1]

[1] Department of Developmental Biology, Washington University School of Medicine in Saint Louis, St. Louis, MO, United States, [2] Needleman Center for Neurometabolism and Axonal Therapeutics, Washington University School of Medicine in Saint Louis, St. Louis, MO, United States, [3] Department of Genetics, Washington University School of Medicine in Saint Louis, St. Louis, MO, United States

OPEN ACCESS

Edited by:
Bostjan Kobe, The University of Queensland, Australia

Reviewed by:
Hyun Ho Park, Chung-Ang University, South Korea
Jonathan Gilley, University of Cambridge, United Kingdom
Michael Carty, Trinity College Dublin, Ireland

***Correspondence:**
Aaron DiAntonio
diantonio@wustl.edu

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *03 August 2021*
Accepted: *26 August 2021*
Published: *23 September 2021*

Citation:
DiAntonio A, Milbrandt J and Figley MD (2021) The SARM1 TIR NADase: Mechanistic Similarities to Bacterial Phage Defense and Toxin-Antitoxin Systems. Front. Immunol. 12:752898. doi: 10.3389/fimmu.2021.752898

The Toll/interleukin-1 receptor (TIR) domain is the signature signalling motif of innate immunity, with essential roles in innate immune signalling in bacteria, plants, and animals. TIR domains canonically function as scaffolds, with stimulus-dependent multimerization generating binding sites for signalling molecules such as kinases and ligases that activate downstream immune mechanisms. Recent studies have dramatically expanded our understanding of the TIR domain, demonstrating that the primordial function of the TIR domain is to metabolize NAD^+. Mammalian SARM1, the central executioner of pathological axon degeneration, is the founding member of the TIR-domain class of NAD^+ hydrolases. This unexpected NADase activity of TIR domains is evolutionarily conserved, with archaeal, bacterial, and plant TIR domains all sharing this catalytic function. Moreover, this enzymatic activity is essential for the innate immune function of these proteins. These evolutionary relationships suggest a link between SARM1 and ancient self-defense mechanisms that has only been strengthened by the recent discovery of the SARM1 activation mechanism which, we will argue, is strikingly similar to bacterial toxin-antitoxin systems. In this brief review we will describe the regulation and function of SARM1 in programmed axon self-destruction, and highlight the parallels between the SARM1 axon degeneration pathway and bacterial innate immune mechanisms.

Keywords: NAD^+, innate immunity, NMNAT2, axon degeneration, plant, metabolism, TIR domain, abortive infection

INTRODUCTION

Injured or diseased axons initiate a self-destruction program known as Wallerian degeneration. SARM1 triggers this pathological axon degeneration (1), and is a key driver of pathology in models of chemotherapy-induced peripheral neuropathy (2–5), traumatic brain injury (6–10), glaucoma (11), and retinal degeneration (12, 13). SARM1 also participates in antiviral defense. SARM1 triggers axon degeneration following rabies infection (14), presumably to halt the spread of the virus as it travels retrogradely down the axon to the neuronal cell body, and induces neuronal cell death in

response to bunyavirus infection (15), killing infected cells and thereby reducing viral spread. Hence, the role of SARM1 in pathological axon degeneration is likely closely linked to its function in antiviral innate immunity.

SARM1 is a multi-domain protein comprised of an autoinhibitory ARM domain, tandem SAM domains mediating multimerization, and a C-terminal TIR domain NAD^+ hydrolase (16, 17). In healthy neurons, SARM1 autoinhibition is maintained by multiple intra- and intermolecular interactions (18), including binding of the N-terminal ARM domain to the C-terminal TIR domain (19) (**Figure 1A**). SARM1 autoinhibition is regulated by an allosteric binding site within the autoinhibitory ARM domain that can bind either nicotinamide adenine dinucleotide (NAD^+) (21, 22) or its precursor, nicotinamide mononucleotide (NMN) (23). NMN promotes SARM1-dependent axon degeneration (24–27). Axon injury leads to loss of the NAD^+ biosynthetic enzyme NMNAT2 (28), resulting in an increased NMN/NAD^+ ratio that promotes NMN binding to the allosteric site (23). The switch from NAD^+ to NMN binding alters the conformation of the autoinhibitory ARM domain, thereby promoting TIR-TIR interactions and enzymatic activity (23) (**Figure 1B**). Below we will highlight commonalities between SARM1 activation and effector mechanisms with similar bacterial innate immune mechanisms.

SARM1-NMNAT2 Is a Candidate Mammalian Toxin-Antitoxin Pair

Just as SARM1 can trigger axon self-destruction in response to rabies infection, so too can a bacterial population acquire immunity to bacteriophage infection through an altruistic suicide mechanism known as abortive infection. When infected, the bacterial cells activate a toxin-antitoxin (TA) system prior to phage replication, killing the infected cells and thereby protecting the community by preventing further phage expansion (29). In a TA system, bacteria express both a lethal toxin and its antagonist, the antitoxin. Upon infection, the antitoxin is degraded, unleashing the degenerative activity of the toxin. Genetic deletion of the toxin-encoding gene yields no phenotype in the absence of the inciting stimuli, whereas deletion of the antitoxin-encoding gene results in cell death due to unchecked toxin activity. This lethality can be rescued by concurrent deletion of the toxin-encoding gene. This TA relationship is strikingly reminiscent of the relationship between NMNAT2 and SARM1, with NMNAT2 serving as the antitoxin and SARM1 as the toxin (**Figure 2**). First, NMNAT2 inhibits the prodegenerative activity of SARM1, as antitoxins inhibit the functions of toxins. Second, the classic TA genetic relationship holds for SARM1 and NMNAT2. Loss of SARM1 has no obvious phenotype in mice until an appropriate stimulus, such as neuronal injury, occurs (17, 30). By contrast, genetic loss of NMNAT2 (the antitoxin) is embryonic lethal in mice (31, 32). Double mutants lacking both NMNAT2 and SARM1 fully rescue this lethality (33). Hence, the essential function of NMNAT2 is to inhibit SARM1. Third, similar to a type II TA system in bacteria (29), the antitoxin NMNAT2 is highly labile (28) and the levels of NMNAT2 are a key determinant of SARM1 activation (34). However, in contrast to the bacterial type II TA system, in which

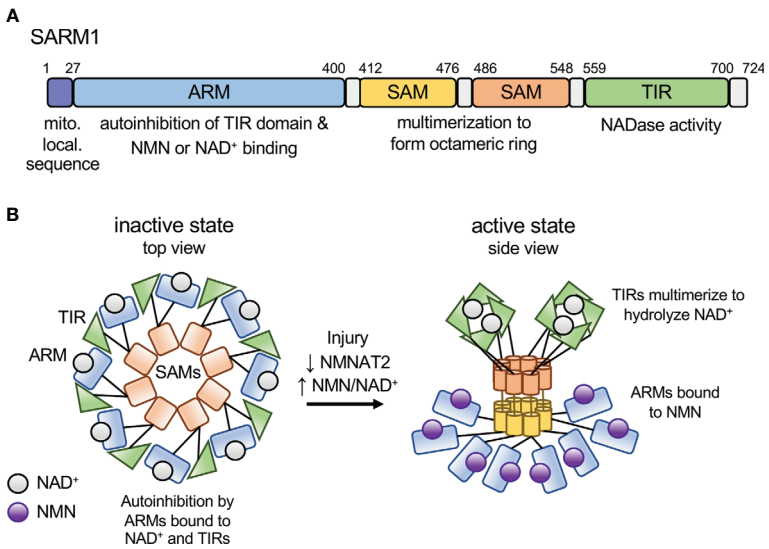


FIGURE 1 | Model of SARM1 domain structure and activation mechanism. **(A)** Domain structure of the human SARM1 protein. SARM1 contains an N-terminal mitochondrial localization sequence and Armadillo-repeat containing domain (ARM), two tandem sterile alpha motif (SAM) domains, and a C-terminal Toll/interleukin-1 receptor (TIR) domain. Numbers denote the amino acid position of the domain boundaries. **(B)** Schematic depicting the activation mechanism of SARM1. In the inactive state SARM1's ARM domains are bound to NAD^+ at the allosteric site and bound to adjacent TIR domains both intra- and inter-molecularly, mediating autoinhibition of the TIR's NADase activity. In response to an increase in the NMN/NAD^+ ratio, NMN binds to the ARM domain allosteric site, resulting in a conformational change in the ARM domain, disengagement of the ARM-TIR interactions, multimerization of the TIR domains and NADase activity. Based on recent structural data from the RPP1 TIR domain, we depict active TIRs as a tetramer forming two active sites for NAD^+ binding (20).

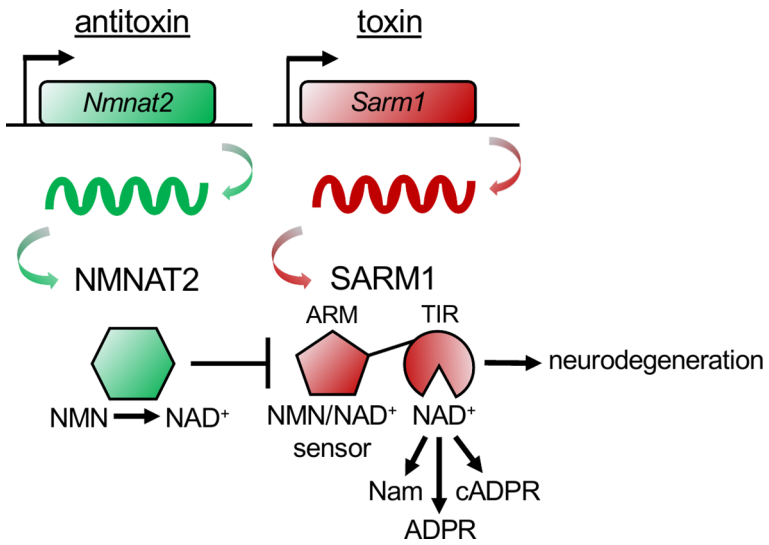


FIGURE 2 | The NMNAT2 antitoxin inhibits the SARM1 toxin to prevent axon degeneration. The antitoxin NMNAT2 converts nicotinamide mononucleotide (NMN) to nicotinamide adenine dinucleotide (NAD+) and thus maintains a healthy NMN/NAD+ ratio in axons. SARM1's ARM domain senses the ratio of NMN/NAD+ by binding to either metabolite. When the antitoxin NMNAT2 is lost, the NMN/NAD+ ratio rises, NMN binds to the toxin SARM1's ARM domain, activating SARM1's TIR domains to hydrolyze NAD+, producing nicotinamide (Nam) and adenosine diphosphate ribose (ADPR), or cyclizing ADPR into cyclic ADPR (cADPR). Activation of the toxin SARM1 drives pathological axon degeneration.

the antitoxin inactivates the toxin through direct binding, NMNAT2 inhibits SARM1 indirectly, by modulating the flow of metabolites that bind and regulate SARM1 activity.

To our knowledge, all known TA systems are found in bacteria. Here we posit that *via* convergent evolution SARM1 and NMNAT2 have developed into an analogous TA system to control axonal fate in mammalian neurons. In an injured or unhealthy axon the transport of NMNAT2 is disrupted (3, 28), and neuronal stress pathways promote NMNAT2 turnover (34, 35), leading to the loss of the labile NMNAT2 antitoxin and the subsequent activation of the toxin SARM1, resulting in rapid and efficient axonal self-destruction. This may be a physiological function of SARM1, enabling the phagocytosis and clearance of damaged axons before their contents leak and potentially harm adjacent axons or cells or induce inflammation. Indeed, this scenario was recently demonstrated in a mouse model of ulcerative colitis, in which SARM1 promotes axon degeneration in the enteric nervous system and thereby limits inflammation in the colon (36). It will be interesting to determine whether NMNAT2 is also lost in this colitis model, or in response to rabies infection, when SARM1 is activated and axons degenerate as an antiviral defense (14).

SARM1 Is the Founding Member of the TIR Domain Family of Innate Immune NADases

The role of SARM1 and its connection to ancient surveillance mechanisms extends beyond its TA relationship with NMNAT2 to its mechanism of degeneration, NAD^+ cleavage. In many bacterial TA relationships, the toxins are, like SARM1, NAD^+ glycohydrolases (37). Examples include the toxins SPN (*S. pyogenes*), TNT and MbcT (*M. tuberculosis*), Tne2 (*P. protegens*), and RES (*P. luminescens*), all of which deplete NAD^+ to induce cellular dysfunction or death and are neutralized by their respective antitoxins (38–42). The newly discovered bacterial TIR domain NADases also participate in phage defense as they are essential components of the Thoeris phage defense mechanism (43, 44). Moreover, bacteria not only use TIR NADases to defend against phage invasion, but also to disrupt mammalian innate immune mechanisms *via* metabolic disruption of the host cell (45). In addition, plant TIR domain innate immune receptors are active NADases and this enzymatic function is essential for the cell death that confers disease resistance (20, 46–48). The conservation of TIR NADase function (**Table 1**) in bacterial, plant, and animal response to infection suggests that TIR-mediated NAD^+ cleavage is a primordial innate immune function.

Finally, SARM1 and evolutionarily diverse TIR domain proteins not only share NADase function, but can also possess regulatory domains controlled *via* allosteric binding to cellular metabolites. The SARM1 TIR domain NADase is fused to a metabolic sensing ARM domain that acts to inhibit the NADase activity until specific environmental signals are present. This is likely a general regulatory mechanism for TIR NADase activation, as organisms from all kingdoms of life encode proteins with TIR domains fused to a variety of other motifs, such as leucine-rich repeat (LRR), tetratricopeptide repeat (TPR), WD repeat, and coiled coil (CC) domains (51–53), that may function as environmental sensors to tune the NADase activity of the fused effector TIR domains. Indeed, this precise regulatory relationship occurs in ancient STING (stimulator of

TABLE 1 | TIR domain-containing proteins with demonstrated intrinsic NADase activity.

TIR domain protein	Organism	References
SARM1	*H. sapiens*	Essuman et al., *Neuron*, (16); Zhao et al., *iScience*, (27); Horsefield et al., *Science*, (47)
SARM1	*M. musculus*	Essuman et al., *Neuron*, (16)
SARM1	*D. rerio*	Essuman et al., *Neuron*, (16)
dSarm	*D. melanogaster*	Essuman et al., *Neuron*, (16)
TirS	*S. aureus*	Essuman et al., *Curr Biol*, (49)
AbTir	*A. baumannii*	Essuman et al., *Curr Biol*, (49)
TcpC	*E. coli*	Essuman et al., *Curr Biol*, (49)
BtpA	*Brucella*	Essuman et al., *Curr Biol*, (49); Coronas-Serna et al., *PLoS Pathog*, (45)
BtpB	*Brucella*	Coronas-Serna et al., *PLoS Pathog*, (45)
PdTir	*P. denitrificans*	Essuman et al., *Curr Biol*, (49)
TcpF	*E. faecalis*	Essuman et al., *Curr Biol*, (49)
ApTir	*Actinoplanes sp.*	Essuman et al., *Curr Biol*, (49)
TcpA	*T. archaeon*	Essuman et al., *Curr Biol*, (49)
TcpO	*M. olleaye*	Essuman et al., *Curr Biol*, (49); Wan et al., *Science* (46)
RBA1	*A. thaliana*	Wan et al., *Science* (46)
RPS4	*A. thaliana*	Wan et al., *Science* (46); Horsefield et al., *Science*, (47)
RPP1	*A. thaliana*	Wan et al., *Science* (46); Horsefield et al., *Science*, (47); Ma et al., *Science*, (20)
BdTIR	*B. distachyon*	Wan et al., *Science* (46); Ofir et al., *bioRxiv*, (44)
L6	*L. usitatissimum*	Horsefield et al., *Science*, (47)
RUN1	*M. rotundifolia*	Horsefield et al., *Science*, (47)
tir-1	*C. elegans*	Horsefield et al., *Science*, (47)
SNC1	*A. thaliana*	Horsefield et al., *Science*, (47)
ROQ1	*N. benthamiana*	Horsefield et al., *Science*, (47)
RPV1	*M. rotundifolia*	Horsefield et al., *Science*, (47)
*Sf*STING	*S. faecium*	Morehouse et al., *Nature*, (50)
ThsB	*B. cereus*	Ofir et al., *bioRxiv*, (44)
ThsB TIR1/TIR2	*B. dafuensis*	Ofir et al., *bioRxiv*, (44)

interferon genes) receptor proteins. In both prokaryotes and lower eukaryotes, STING domains are fused with TIR NADase domains. This effectively couples STING domain sensing of cyclic dinucleotides produced by the cyclic AMP-GMP synthase (cGAS) cellular surveillance system to TIR domain NADase activity (50). We suggest that additional multi-domain proteins encoding TIR NADases are likely regulated *via* metabolite binding, sensing changes in cellular metabolism and responding *via* NAD^+ hydrolysis.

DISCUSSION

This brief survey of SARM1 and the family of TIR domain NADases demonstrates that mechanistic insights into SARM1 regulation and function have enabled major breakthroughs in our understanding of TIR domain proteins across the domains of life. Key insights from these studies are the identification of SARM1/NMNAT2 as the first candidate mammalian toxin/antitoxin pair, the recognition of multidomain TIR containing proteins as coordinated metabolic sensors and effectors, and the realization that there is a striking commonality between mechanisms of neurodegeneration and the primordial battle between bacteria and bacteriophages.

AUTHOR CONTRIBUTIONS

Writing – original draft, all authors. Writing – review and editing, all authors. Funding acquisition, AD and JM. Supervision, AD and JM. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by National Institutes of Health grants R37NS065053 to AD and RO1NS087632 to AD and JM.

REFERENCES

1. Figley MD, DiAntonio A. The SARM1 Axon Degeneration Pathway: Control of the NAD+ Metabolome Regulates Axon Survival in Health and Disease. *Curr Opin Neurobiol* (2020) 63:59–66. doi: 10.1016/j.conb.2020.02.012
2. Geisler S, Doan RA, Strickland A, Huang X, Milbrandt J, DiAntonio A. Prevention of Vincristine-Induced Peripheral Neuropathy by Genetic Deletion of SARM1 in Mice. *Brain* (2016) 139:3092–108. doi: 10.1093/brain/aww251
3. Geisler S, Doan RA, Cheng GC, Cetinkaya-Fisgin A, Huang SX, Höke A, et al. Vincristine and Bortezomib Use Distinct Upstream Mechanisms to Activate a

Common SARM1-Dependent Axon Degeneration Program. *JCI Insight* (2019) 4(17):e129920. doi: 10.1172/jci.insight.129920
4. Cetinkaya-Fisgin A, Luan X, Reed N, Jeong YE, Oh BC, Hoke A. Cisplatin Induced Neurotoxicity is Mediated by Sarm1 and Calpain Activation. *Sci Rep* (2020) 10:21889. doi: 10.1038/s41598-020-78896-w
5. Gould SA, White M, Wilbrey AL, Pór E, Coleman MP, Adalbert R. Protection Against Oxaliplatin-Induced Mechanical and Thermal Hypersensitivity in Sarm1-/- Mice. *Exp Neurol* (2021) 338:113607. doi: 10.1016/j.expneurol.2021.113607
6. Maynard ME, Redell JB, Zhao J, Hood KN, Vita SM, Kobori N, et al. Sarm1 Loss Reduces Axonal Damage and Improves Cognitive Outcome After Repetitive Mild Closed Head Injury. *Exp Neurol* (2020) 327:113207. doi: 10.1016/j.expneurol.2020.113207
7. Marion CM, McDaniel DP, Armstrong RC. Sarm1 Deletion Reduces Axon Damage, Demyelination, and White Matter Atrophy After Experimental Traumatic Brain Injury. *Exp Neurol* (2019) 321:113040. doi: 10.1016/j.expneurol.2019.113040
8. Bradshaw DVJr, Knutsen AK, Korotcov A, Sullivan GM, Radomski KL, Dardzinski BJ, et al. Genetic Inactivation of SARM1 Axon Degeneration Pathway Improves Outcome Trajectory After Experimental Traumatic Brain Injury Based on Pathological, Radiological, and Functional Measures. *Acta Neuropathol Commun* (2021) 9:89. doi: 10.1186/s40478-021-01193-8
9. Ziogas NK, Koliatsos VE. Primary Traumatic Axonopathy in Mice Subjected to Impact Acceleration: A Reappraisal of Pathology and Mechanisms With High-Resolution Anatomical Methods. *J Neurosci* (2018) 38:4031–47. doi: 10.1523/JNEUROSCI.2343-17.2018
10. Henninger N, Bouley J, Sikoglu EM, An J, Moore CM, King JA, et al. Attenuated Traumatic Axonal Injury and Improved Functional Outcome After Traumatic Brain Injury in Mice Lacking Sarm1. *Brain* (2016) 139:1094–105. doi: 10.1093/brain/aww001
11. Ko KW, Milbrandt J, DiAntonio A. SARM1 Acts Downstream of Neuroinflammatory and Necroptotic Signaling to Induce Axon Degeneration. *J Cell Biol* (2020) 219(8):e201912047. doi: 10.1083/jcb.201912047
12. Ozaki E, Gibbons L, Neto NG, Kenna P, Carty M, Humphries M, et al. SARM1 Deficiency Promotes Rod and Cone Photoreceptor Cell Survival in a Model of Retinal Degeneration. *Life Sci Alliance* (2020) 3(5):e201900618. doi: 10.26508/lsa.201900618
13. Sasaki Y, Kakita H, Kubota S, Sene A, Lee TJ, Ban N, et al. SARM1 Depletion Rescues NMNAT1-Dependent Photoreceptor Cell Death and Retinal Degeneration. *Elife* (2020) 9:e62027. doi: 10.7554/eLife.62027
14. Sundaramoorthy V, Green D, Locke K, O'Brien CM, Dearnley M, Bingham J. Novel Role of SARM1 Mediated Axonal Degeneration in the Pathogenesis of Rabies. *PloS Pathog* (2020) 16:e1008343. doi: 10.1371/journal.ppat.1008343
15. Mukherjee P, Woods TA, Moore RA, Peterson KE. Activation of the Innate Signaling Molecule MAVS by Bunyavirus Infection Upregulates the Adaptor Protein SARM1, Leading to Neuronal Death. *Immunity* (2013) 38:705–16. doi: 10.1016/j.immuni.2013.02.013
16. Essuman K, Summers DW, Sasaki Y, Mao X, DiAntonio A, Milbrandt J. The SARM1 Toll/Interleukin-1 Receptor Domain Possesses Intrinsic NAD+ Cleavage Activity That Promotes Pathological Axonal Degeneration. *Neuron* (2017) 93:1334–1343.e5. doi: 10.1016/j.neuron.2017.02.022
17. Gerdts J, Summers DW, Sasaki Y, DiAntonio A, Milbrandt J. Sarm1-Mediated Axon Degeneration Requires Both SAM and TIR Interactions. *J Neurosci* (2013) 33:13569–80. doi: 10.1523/JNEUROSCI.1197-13.2013
18. Shen C, Vohra M, Zhang P, Mao X, Figley MD, Zhu J, et al. Multiple Domain Interfaces Mediate SARM1 Autoinhibition. *Proc Natl Acad Sci USA* (2021) 118(4):e2023151118. doi: 10.1073/pnas.2023151118
19. Summers DW, Gibson DA, DiAntonio A, Milbrandt J. SARM1-Specific Motifs in the TIR Domain Enable NAD+ Loss and Regulate Injury-Induced SARM1 Activation. *Proc Natl Acad Sci USA* (2016) 113:E6271–80. doi: 10.1073/pnas.1601506113
20. Ma S, Lapin D, Liu L, Sun Y, Song W, Zhang X, et al. Direct Pathogen-Induced Assembly of an NLR Immune Receptor Complex to Form a Holoenzyme. *Science* (2020) 370(6521):eabe3069. doi: 10.1126/science.abe3069
21. Jiang Y, Liu T, Lee C-H, Chang Q, Yang J, Zhang Z. The NAD+-Mediated Self-Inhibition Mechanism of Pro-Neurodegenerative SARM1. *Nature* (2020) 588:658–63. doi: 10.1038/s41586-020-2862-z
22. Sporny M, Guez-Haddad J, Khazma T, Yaron A, Dessau M, Shkolnisky Y, et al. Structural Basis for SARM1 Inhibition and Activation Under Energetic Stress. *Elife* (2020) 9:e62021. doi: 10.7554/eLife.62021
23. Figley MD, Gu W, Nanson JD, Shi Y, Sasaki Y, Cunnea K, et al. SARM1 is a Metabolic Sensor Activated by an Increased NMN/NAD+ Ratio to Trigger Axon Degeneration. *Neuron* (2021) 109:1118–36.e11. doi: 10.1016/j.neuron.2021.02.009
24. Di Stefano M, Nascimento-Ferreira I, Orsomando G, Mori V, Gilley J, Brown R, et al. A Rise in NAD Precursor Nicotinamide Mononucleotide (NMN) After Injury Promotes Axon Degeneration. *Cell Death Differ* (2015) 22:731–42. doi: 10.1038/cdd.2014.164
25. Sasaki Y, Nakagawa T, Mao X, DiAntonio A, Milbrandt J. NMNAT1 Inhibits Axon Degeneration *via* Blockade of SARM1-Mediated NAD+ Depletion. *Elife* (2016) 5:1010. doi: 10.7554/eLife.19749
26. Di Stefano M, Loreto A, Orsomando G, Mori V, Zamporlini F, Hulse RP, et al. NMN Deamidase Delays Wallerian Degeneration and Rescues Axonal Defects Caused by NMNAT2 Deficiency *In Vivo*. *Curr Biol* (2017) 27:784–94. doi: 10.1016/j.cub.2017.01.070
27. Zhao ZY, Xie XJ, Li WH, Liu J, Chen Z, Zhang B, et al. A Cell-Permeant Mimetic of NMN Activates SARM1 to Produce Cyclic ADP-Ribose and Induce Non-Apoptotic Cell Death. *iScience* (2019) 15:452–66. doi: 10.1016/j.isci.2019.05.001
28. Gilley J, Coleman MP. Endogenous Nmnat2 is an Essential Survival Factor for Maintenance of Healthy Axons. *PloS Biol* (2010) 8:e1000300. doi: 10.1371/journal.pbio.1000300
29. Harms A, Brodersen DE, Mitarai N, Gerdes K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell* (2018) 70:768–84. doi: 10.1016/j.molcel.2018.01.003
30. Osterloh JM, Yang J, Rooney TM, Fox AN, Adalbert R, Powell EH, et al. Dsarm/Sarm1 is Required for Activation of an Injury-Induced Axon Death Pathway. *Science* (2012) 337:481–4. doi: 10.1126/science.1223899
31. Hicks AN, Lorenzetti D, Gilley J, Lu B, Andersson K-E, Miligan C, et al. Nicotinamide Mononucleotide Adenylyltransferase 2 (Nmnat2) Regulates Axon Integrity in the Mouse Embryo. *PloS One* (2012) 7:e47869–10. doi: 10.1371/journal.pone.0047869
32. Gilley J, Adalbert R, Yu G, Coleman MP. Rescue of Peripheral and CNS Axon Defects in Mice Lacking NMNAT2. *J Neurosci* (2013) 33:13410–24. doi: 10.1523/JNEUROSCI.1534-13.2013
33. Gilley J, Orsomando G, Nascimento-Ferreira I, Coleman MP. Absence of SARM1 Rescues Development and Survival of NMNAT2-Deficient Axons. *Cell Rep* (2015) 10:1974–81. doi: 10.1016/j.celrep.2015.02.060
34. Summers DW, Frey E, Walker LJ, Milbrandt J, DiAntonio A. DLK Activation Synergizes With Mitochondrial Dysfunction to Downregulate Axon Survival Factors and Promote SARM1-Dependent Axon Degeneration. *Mol Neurobiol* (2020) 57:1146–58. doi: 10.1007/s12035-019-01796-2
35. Walker LJ, Summers DW, Sasaki Y, Brace EJ, Milbrandt J, DiAntonio A. MAPK Signaling Promotes Axonal Degeneration by Speeding the Turnover of the Axonal Maintenance Factor NMNAT2. *Elife* (2017) 6:1–20. doi: 10.7554/eLife.22540
36. Sun Y, Wang Q, Wang Y, Ren W, Cao Y, Li J, et al. Sarm1-Mediated Neurodegeneration Within the Enteric Nervous System Protects Against Local Inflammation of the Colon. *Protein Cell* (2021) 12:621–38. doi: 10.1007/s13238-021-00835-w
37. Roussin M, Salcedo SP. NAD+-Targeting by Bacteria: An Emerging Weapon in Pathogenesis. *FEMS Microbiol Rev* (2021) fuab037. doi: 10.1093/femsre/fuab037
38. Freire DM, Gutierrez C, Garza-Garcia A, Grabowska AD, Sala AJ, Ariyachaokun K, et al. An NAD+ Phosphorylase Toxin Triggers Mycobacterium Tuberculosis Cell Death. *Mol Cell* (2019) 73:1282–91.e8. doi: 10.1016/j.molcel.2019.01.028
39. Skjerning RB, Senissar M, Winther KS, Gerdes K, Brodersen DE. The RES Domain Toxins of RES-Xre Toxin-Antitoxin Modules Induce Cell Stasis by Degrading NAD+. *Mol Microbiol* (2019) 111:221–36. doi: 10.1111/mmi.14150
40. Smith CL, Ghosh J, Elam JS, Pinkner JS, Hultgren SJ, Caparon MG, et al. Structural Basis of Streptococcus Pyogenes Immunity to Its NAD+ Glycohydrolase Toxin. *Structure* (2011) 19:192–202. doi: 10.1016/j.str.2010.12.013
41. Sun J, Siroy A, Lokareddy RK, Speer A, Doornbos KS, Cingolani G, et al. The Tuberculosis Necrotizing Toxin Kills Macrophages by Hydrolyzing NAD. *Nat Struct Mol Biol* (2015) 22:672–8. doi: 10.1038/nsmb.3064
42. Tang JY, Bullen NP, Ahmad S, Whitney JC. Diverse NADase Effector Families Mediate Interbacterial Antagonism *via* the Type VI Secretion System. *J Biol Chem* (2018) 293:1504–14. doi: 10.1074/jbc.RA117.000178

43. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic Discovery of Antiphage Defense Systems in the Microbial Pangenome. *Science* (2018) 3:eaar4120–18. doi: 10.1126/science.aar4120
44. Ofir G, Herbst E, Baroz M, Cohen D, Millman A, Doron S, et al. Antiviral Activity of Bacterial TIR Domains *via* Signaling Molecules That Trigger Cell Death. *bioRxiv* (2021). doi: 10.1101/2021.01.06.425286
45. Coronas-Serna JM, Louche A, Rodríguez-Escudero M, Roussin M, Imbert PRC, Rodríguez-Escudero I, et al. The TIR-Domain Containing Effectors BtpA and BtpB From Brucella Abortus Impact NAD Metabolism. *PloS Pathog* (2020) 16:e1007979. doi: 10.1371/journal.ppat.1007979
46. Wan L, Essuman K, Anderson RG, Sasaki Y, Monteiro F, Chung E-H, et al. TIR Domains of Plant Immune Receptors Are NAD+-Cleaving Enzymes That Promote Cell Death. *Science* (2019) 365:799–803. doi: 10.1126/science.aax1771
47. Horsefield S, Burdett H, Zhang X, Manik MK, Shi Y, Chen J, et al. NAD+ Cleavage Activity by Animal and Plant TIR Domains in Cell Death Pathways. *Science* (2019) 365:793–9. doi: 10.1126/science.aax1911
48. Swiderski MR, Birker D, Jones JDG. The TIR Domain of TIR-NB-LRR Resistance Proteins Is a Signaling Domain Involved in Cell Death Induction. *Mol Plant Microbe Interact* (2009) 22:157–65. doi: 10.1094/MPMI-22-2-0157
49. Essuman K, Summers DW, Sasaki Y, Mao X, Yim AKY, DiAntonio A, et al. TIR Domain Proteins Are an Ancient Family of NAD+-Consuming Enzymes. *Curr Biol* (2018) 28:421–30.e4. doi: 10.1016/j.cub.2017.12.024
50. Morehouse BR, Govande AA, Millman A, Keszei AFA, Lowey B, Ofir G, et al. STING Cyclic Dinucleotide Sensing Originated in Bacteria. *Nature* (2020) 586:429–33. doi: 10.1038/s41586-020-2719-5
51. Zhang Q, Zmasek CM, Cai X, Godzik A. TIR Domain-Containing Adaptor SARM Is a Late Addition to the Ongoing Microbe-Host Dialog. *Dev Comp Immunol* (2011) 35:461–8. doi: 10.1016/j.dci.2010.11.013
52. Nanson JD, Kobe B, Ve T. Death, TIR, and RHIM: Self-Assembling Domains Involved in Innate Immunity and Cell-Death Signaling. *J Leukoc Biol* (2019) 105:363–75. doi: 10.1002/JLB.MR0318-123R
53. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L. Comparative Genomic Analyses Reveal a Vast, Novel Network of Nucleotide-Centric Systems in Biological Conflicts, Immunity and Signaling. *Nucleic Acids Res* (2015) 43:10633–54. doi: 10.1093/nar/gkv1267

Conflict of Interest: AD and JM are co-founders, scientific advisory board members, and shareholders of Disarm Therapeutics, Inc. (a wholly-owned subsidiary of Eli Lilly and Company). The editor declared a past co-authorship with all the authors.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

REVIEW
published: 01 November 2021
doi: 10.3389/fimmu.2021.769543

Running With Scissors: Evolutionary Conflicts Between Viral Proteases and the Host Immune System

*Brian V. Tsu†, Elizabeth J. Fay†, Katelyn T. Nguyen, Miles R. Corley, Bindhu Hosuru, Viviana A. Dominguez and Matthew D. Daugherty**

Division of Biological Sciences, University of California, San Diego, CA, United States

Many pathogens encode proteases that serve to antagonize the host immune system. In particular, viruses with a positive-sense single-stranded RNA genome [(+)ssRNA], including picornaviruses, flaviviruses, and coronaviruses, encode proteases that are not only required for processing viral polyproteins into functional units but also manipulate crucial host cellular processes through their proteolytic activity. Because these proteases must cleave numerous polyprotein sites as well as diverse host targets, evolution of these viral proteases is expected to be highly constrained. However, despite this strong evolutionary constraint, mounting evidence suggests that viral proteases such as picornavirus 3C, flavivirus NS3, and coronavirus 3CL, are engaged in molecular 'arms races' with their targeted host factors, resulting in host- and virus-specific determinants of protease cleavage. In cases where protease-mediated cleavage results in host immune inactivation, recurrent host gene evolution can result in avoidance of cleavage by viral proteases. In other cases, such as recently described examples in NLRP1 and CARD8, hosts have evolved 'tripwire' sequences that mimic protease cleavage sites and activate an immune response upon cleavage. In both cases, host evolution may be responsible for driving viral protease evolution, helping explain why viral proteases and polyprotein sites are divergent among related viruses despite such strong evolutionary constraint. Importantly, these evolutionary conflicts result in diverse protease-host interactions even within closely related host and viral species, thereby contributing to host range, zoonotic potential, and pathogenicity of viral infection. Such examples highlight the importance of examining viral protease-host interactions through an evolutionary lens.

Keywords: viral proteases, host-virus evolution, innate antiviral immunity, molecular arms races, effector-triggered immunity, inflammasome

OPEN ACCESS

Edited by:
Bostjan Kobe,
The University of Queensland, Australia

Reviewed by:
Ross Thomas Barnard,
The University of Queensland, Australia
Junfa Yuan,
Huazhong Agricultural University, China

***Correspondence:**
Matthew D. Daugherty
mddaugherty@ucsd.edu

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *02 September 2021*
Accepted: *08 October 2021*
Published: *01 November 2021*

Citation:
Tsu BV, Fay EJ, Nguyen KT, Corley MR, Hosuru B, Dominguez VA and Daugherty MD (2021) Running With Scissors: Evolutionary Conflicts Between Viral Proteases and the Host Immune System. Front. Immunol. 12:769543. doi: 10.3389/fimmu.2021.769543

INTRODUCTION

Positive-sense single-stranded RNA [(+)ssRNA, see **Table 1** for glossary of abbreviations] viruses represent the largest group of RNA viruses, spanning 30 divergent viral families that include important human pathogens in *Flaviviridae*, *Picornaviridae*, and *Coronaviridae* such as dengue virus, poliovirus, and SARS-CoV-2 (1). Despite their diversity, many viruses in this group share a common replication strategy: their (+)ssRNA viral genomes are delivered to host cells as a

TABLE 1 | List of abbreviations and alternative names used throughout this review.

Acronym/Abbreviation	Alternative Names	Definition
General terms		
ssRNA		Single-stranded RNA
dsRNA		Double-stranded RNA
PRR		Pathogen recognition receptor
ETI		Effector-triggered immunity
LF		Lethal Factor
Host factors		
NLRP1	NALP1	NACHT, LRR, and PYD domains-containing protein 1
CARD8	CARDINAL	Caspase Recruitment Domain Family Member 8
eIF4F		Eukaryotic translation initiation factor 4F, composed of subunits eIF4A, EIF4E, and eIF4G
PABP	PABPC1	PolyA binding protein
eIF4A		Eukaryotic translation initiation factor 4A
eIF4G		Eukaryotic translation initiation factor 4G
G3BP1		Ras GTPase-activating protein-binding protein 1
RIG-I	DDX58	Retinoic acid-inducible gene-I-like receptor; DEXD/H-box helicase 58
MDA5	IFIH1	Melanoma differentiation-associated protein 5; interferon-induced with helicase C domain 1
cGAS	MB21D1; C6orf150	Cyclic GMP–AMP synthase; Mab-21 domain containing 1;
IFN		Interferon
ISG		Interferon-stimulated gene
STING	TMEM173	Stimulator of interferon genes; transmembrane protein 173
MAVS	IPS-1, CARDIF, VISA	Mitochondrial antiviral-signaling protein; IFN-β promoter stimulator 1
NF-κB		Nuclear transcription factor kB, often composed of p65 (RelA) and p50 (NFKB1) subunits
NEMO	IKBKG, IKK-gamma	Nuclear transcription factor κB essential modulator
STAT2		Signal transducer and activator of transcription 2
IL-1		Interleukin-1
IκBα	NFKBIA	NF-κB inhibitor alpha
IKK		IκB kinase complex, includes NEMO
Picornaviridae		
PV		Poliovirus
CVB3		Coxsackievirus B3
FMDV		Foot-and-mouth disease virus
HepA		Hepatitis A virus
EMCV		Encephalomyocarditis virus
Coronaviridae		
3CL	NSP5; Mpro	3C-like; nonstructural protein 5; Main protease
CoV		Coronavirus
PLP		Papain-like protease
SARS-CoV-2		Severe acute respiratory syndrome-associated coronavirus-2, causative agent on COVID-19
MHV		Murine hepatitis virus
hCoV 229E		Human coronavirus 229E
bCoV HKU4		Bat coronavirus HKU4
hCoV-OC43		Human coronavirus OC43
hCoV-HKU1		Human coronavirus HKU1
PDCoV		Porcine deltacoronavirus
Flaviviridae		
NS3		Nonstructural protein 3
HCV		Hepatitis C Virus
DENV		Dengue virus
YFV		Yellow fever virus
WNV		West Nile virus
JEV		Japanese encephalitis virus
ZIKV		Zika virus

translation-ready mRNA that encodes a multidomain viral polyprotein. Following translation of the viral polyprotein by host ribosomes, one or more embedded viral proteases cleave the polyprotein into individual, functional proteins at numerous sequence-specific positions (**Figure 1A**). Polyprotein cleavage at these specific sites is necessary for sustained virus replication and propagation, making viral proteases an attractive target for development of antiviral therapeutics (2, 3).

In addition to their essential role in the viral life cycle, (+) ssRNA viral proteases also cleave host proteins to manipulate host processes, including the host innate antiviral immune response (4). Importantly, host targets are cleaved with the same sequence specificity as sites within the viral polyprotein (**Figure 1B**). These dual roles place viral proteases at the intersection of two opposing selective pressures. On one side, the virus and its polyprotein site targets are under strong

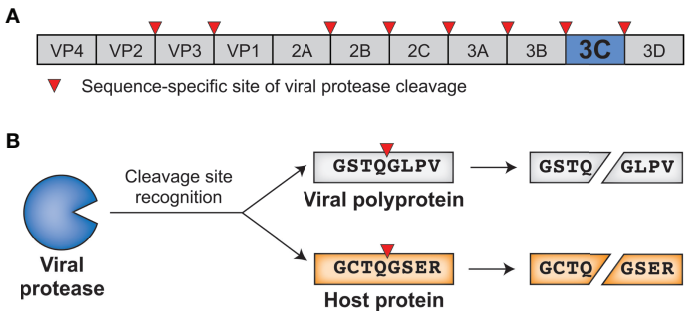


FIGURE 1 | Viral proteases cleave specific sites within the viral polyprotein and host proteins. **(A)** Schematic of an enterovirus (family: *Picornaviridae*) polyprotein, with the position of the 3C protease and sites of 3C-mediated cleavage shown. **(B)** 3C protease recognizes and cleaves viral polyprotein sites and host proteins with the same sequence specificity.

pressure to be conserved, as any changes to the protease sequence specificity or protease sites without concomitant changes to the other would be deleterious for viral fitness. On the other side, viral fitness may be expected to benefit from a protease's ability to adapt to and cleave new host targets, newly evolved sequences in the same host, or divergent sequences in a different host to facilitate cross-species transmission. This type of direct engagement between viral proteases and host factors thus generates an evolutionary conflict where both sides may be driven to adapt in a type of escalating molecular 'arms race' (5–8).

Molecular arms races exist as a result of the competing evolutionary interests of viruses and their hosts. Such competing interests establish an evolutionary equilibrium that is characterized by cyclical adaptations that exemplify a so-called 'Red Queen' genetic conflict (9). In these cases, viral adaptations that allow for successful infection of a host will provide a temporary advantage to the virus. However, host adaptations may restore the advantage to the host, applying selection pressure back to the virus (**Figure 2A**). Thus, molecular interactions between viruses and their hosts, particularly those interactions that contribute to potentiation or inhibition of virus

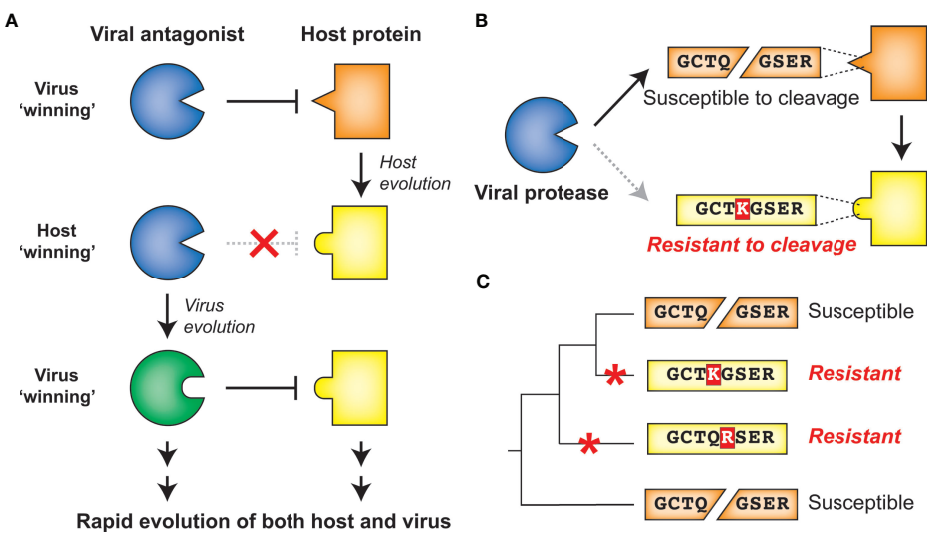


FIGURE 2 | Host-virus evolutionary arms races can be driven by protease-target interactions. **(A)** Host-virus arms races occur when there is direct interaction between host and viral factors, which places evolutionary pressure to select for variants. In this scenario, a viral antagonist recognizes and inactivates a host protein, driving host evolution away from this interaction. The necessity of host target cleavage for virus replication in turn drives evolution of the viral antagonist to reestablish host target recognition. **(B)** Single amino acid changes in the sequence-specific cleavage motif can eliminate cleavage by a viral protease. **(C)** Across a phylogenetic tree, changes can occur recurrently resulting in differential susceptibility between even closely related species. Red asterisks mark the branch in which an amino acid change occurred that alters cleavage susceptibility.

replication, are shaped by immense evolutionary pressure on both parties: hosts are driven to both maintain interactions that activate or carry out antiviral defenses and evade virus interactions that prevent these responses, and viruses are driven to do the opposite. The result is recurrent adaptation of both virus and host to promote either virus replication or host antiviral mechanisms, respectively (**Figure 2A**) (5–8). Due to the fact that the direct molecular host-virus interfaces are those that are being remodeled during such molecular arms races, single amino acid changes can change the outcome of these conflicts (5, 7). Indeed, traces of these host-virus conflicts can be detected in host genomes by identifying gene codons that show evolutionary signatures of recurrent diversifying (positive) selection (5, 7, 10). Similarly, viruses are known to adapt during or following cross-species transmission to a novel host, and such adaptations can also be characterized by signatures of positive selection in viral genomes (11–14). Importantly, whether the host has evolved to the virus or the virus has evolved to the host, the resulting genetic and molecular changes determine the host range and pathogenesis of viruses, including influencing the ability of viruses to zoonotically transmit into the human population (7, 8, 13).

Due to the importance of sequence specificity to protease-host interactions, evolutionary arms races at the interfaces of proteases and their targets would be expected to exist. For instance, a single amino acid change in a targeted host protein at a position that is important for sequence-specific protease cleavage could completely reverse cleavage susceptibility (**Figure 2B**). As a result, single lineage-specific changes at any number of positions in the cleavage motif would be expected to alter cleavage susceptibility even among closely related hosts, establishing species-specific host-virus interactions that could drive viral host range (**Figure 2C**). Indeed, while a great deal of research on viral proteases has focused on conserved elements of protease function, emerging evidence suggests that both hosts and viruses are evolving in ways that can impact the host- and virus-specificity of cleavage. Here, we review the host-viral molecular conflicts engaged by the main proteases of flaviviruses, picornaviruses, and coronaviruses to emphasize how proteases of (+)ssRNA viruses act as evolutionary drivers of host innate immunity, and how viral proteases are being shaped by these same molecular conflicts. This evolutionary perspective highlights the importance of viral proteases and their host targets as being an important determinant of viral host range, tissue tropism and pathogenesis, and zoonotic potential of (+)ssRNA viruses.

DESPITE EVOLUTIONARY CONSTRAINTS, MAIN PROTEASES OF (+)SSRNA VIRUSES CONTINUE TO EVOLVE

Virus-encoded proteases are essential to the life cycle of numerous (+)ssRNA viruses. Newly synthesized viral polyproteins mature into individual, functional proteins *via* a series of cleavage events carried out by virus-encoded and host proteases. For *Picornaviridae* and *Coronaviridae*, the viral cysteine proteases 3C and 3C-Like (3CL) respectively, are responsible for the majority of polyprotein processing events (**Figures 3A, B**) (4, 16–19). Most picornaviruses have six or more 3C cleavage sites throughout the polyprotein (**Figure 1A**), and there is a preference to cleave between a glutamine (Q) in the P1 position and a small residue [e.g. glycine (G) or serine (S)] in the P1' position (**Figure 3A**) (15, 16, 20). Likewise, coronaviruses (CoVs) have ten or more cleavage sites for the 3CL protease (also known as MPro or nsp5 in several CoVs including SARS-CoV-2) (**Figure 3B**) (17, 18, 21). Numerous other viral families, including members of *Caliciviridae* (e.g. norovirus) (22) and *Potyviridae* (*e.g.* tobacco etch virus) (23) encode a cysteine protease with a similar specificity for cleavage between a Q and a small residue, whereas members of *Togaviridae* (*e.g.* Chikungunya virus) use a cysteine protease with different cleavage specificity (24). Other viral families use a serine protease, including *Flaviviridae*, where the serine protease NS3 processes at least four polyprotein cleavage sites (**Figure 3C**) (25, 26). Here and in subsequent sections, we will predominantly discuss activities of the 3C, 3CL, and NS3 proteases of *Picornaviridae*, *Coronaviridae*, and *Flaviviridae*, respectively, due to their known roles in cleaving mammalian host factors. It is important to point out that because the polyprotein is sequentially processed, and not always to completion, protease activity may also be carried out when 3C, 3CL, or NS3 remains fused or associated with additional viral proteins (27, 28). This is especially true in the *Flaviviridae*, where the NS3 protease usually functions in association with NS2B (in the case of flaviviruses such as dengue and Zika viruses) or NS4A [in the case of hepatitis C virus (HCV)] (26, 29). However, for the sake of clarity, we will subsequently only refer to the protease domains of 3C, 3CL, or NS3. Moreover, many (+)ssRNA viruses encode additional proteases involved in both polyprotein processing and host antagonism, including the 2A protease in some picornaviruses and the papain-like protease (PLP) in coronaviruses (18, 30). Finally, viruses other than (+) ssRNA viruses can encode proteases that are important for polyprotein processing, most notably, the retrovirally-encoded aspartyl protease (31). While all of these additional proteases from (+)ssRNA viruses and retroviruses play important host antagonism roles, and likely shape host and viral evolution, they will not be extensively explored here.

The functions of the (+)ssRNA viral proteases described above are, by definition of being required for completion of the viral life cycle, well conserved. In addition to homology between the proteases themselves, the positions and sequences of the polyprotein cleavage motifs are often similar between members of the same viral family. Indeed, this conservation of polyprotein cleavage motifs has made it possible to compile sequences surrounding the cleavage site from genome sequences alone to generate a consensus motif for the viral protease that can be used to predict host and viral targets (15, 20, 32, 33) (**Figure 3A**). These consensus motifs are often generated using many diverse viruses, relying on the assumption that protease sequence

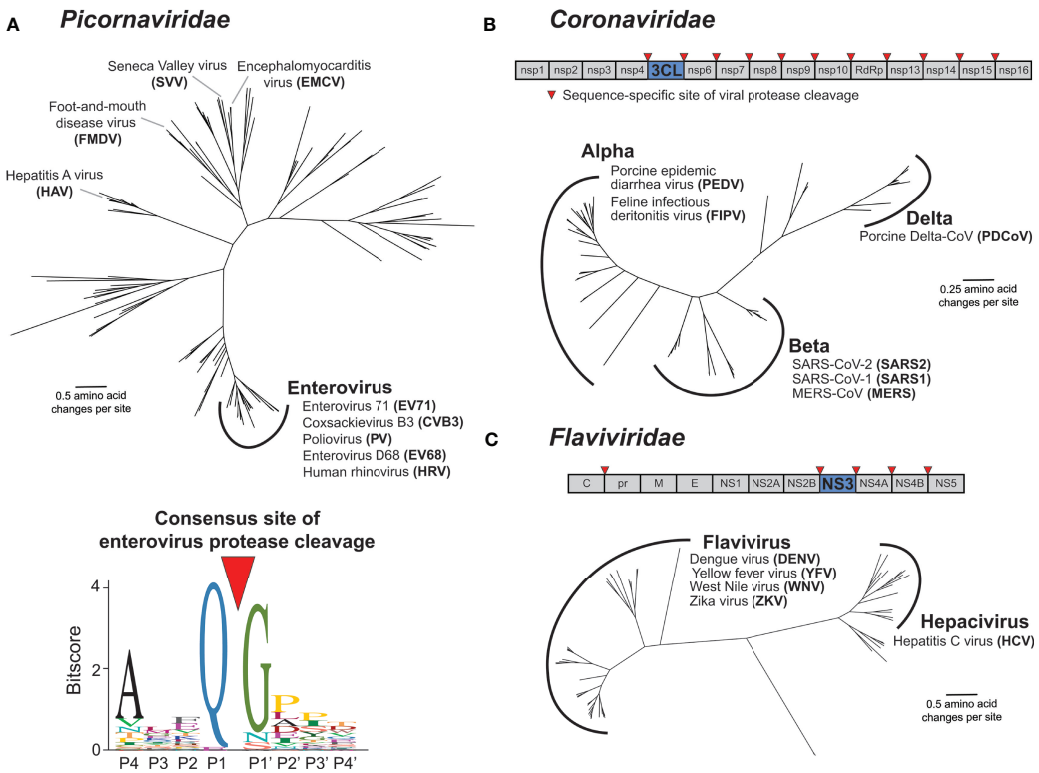


FIGURE 3 | Main proteases in *Picornaviridae, Coronaviridae*, and *Flaviviridae*. **(A)** Phylogenetic tree of available RefSeq *Picornaviridae* 3C protease protein sequences (151 total, top). Names of viruses with human relevance or referenced throughout the text are listed next to their respective genus or singular node. The consensus enterovirus 3C cleavage motif (bottom) as was generated previously (15). The cleavage site is shown flanked by four amino acids upstream (labeled P4 through P1) and four amino acids downstream (labeled P1' through P4'). **(B)** Schematic of the SARS-CoV-2 (family: *Coronaviridae*) nonstructural (ORF1ab) polyprotein, with the position of the 3CL protease and sites of 3CL-mediated cleavage shown. Phylogenetic tree of available RefSeq *Coronaviridae* 3CL protease protein sequences (64 total). Names of viruses with human relevance or referenced throughout the text are listed next to their respective genus. **(C)** Schematic of the dengue virus (DENV) (family: *Flaviviridae*) polyprotein, with the position of the NS3 protease and sites of NS3-mediated cleavage shown. Phylogenetic tree of available RefSeq *Flaviviridae* NS3 protease protein sequences (68 total). Names of viruses with human relevance or referenced throughout the text are listed next to their respective genus.

specificity is well conserved among virus species. Interestingly, despite the evolutionary constraint to maintain cleavage across multiple sites in the polyprotein, virus-encoded proteases are substantially divergent across viruses (**Figure 3**). For instance, picornavirus 3C proteases can share less than 20% amino acid sequence identity, despite sharing an overall similar fold and many homologous cleavage sites (15). Similar evolutionary distances are observed with other families of proteases, including *Coronaviridae* 3CL and *Flaviviridae* NS3 (**Figure 3**).

Even with the divergence of protease sequences, protease sequence specificity is expected to be well conserved within closely related viruses given the essentiality of cleaving multiple site-specific polyprotein sites. Surprisingly, there is mounting evidence that this is not the case. For instance, among closely related serotypes of dengue virus (DENV), biochemical substrate profiling has revealed a subtle but clear shift in the NS3 protease cleavage sequence specificity profile (34). This type of in-depth comparative biochemical analysis of other (+) ssRNA proteases has not been conducted, but assays on model substrates have revealed differences in cleavage specificity even among 3C proteases within the Enterovirus genus of *Picornaviridae* (35). Some of the best evidence that protease sequence specificity is changing between related viruses has come from studies using chimeric viruses in which the protease of one virus species is inserted into the backbone of another virus. If such protease swaps result in insufficient or improper cleavage of the polyprotein and reduced viral replication, it would suggest divergence in protease sequence specificity between the parental viruses. For example, among enteroviruses (**Figure 3A**), replacing the poliovirus (PV) 3C protease with 3C proteases from human rhinovirus 14 or coxsackievirus B3 (CVB3) resulted in reduced, changed, or loss of cleavage products (36). Likewise, within the Flavivirus genus of *Flaviviridae* (**Figure 3C**), swapping the protease domain of DENV NS3 for the protease domain of yellow fever virus (YFV) ablates processing of polyproteins containing DENV cleavage sites (37). Additionally, West Nile virus (WNV) NS3 can cleave a polyprotein site in only one of two closely related DENV2 strains, where the only difference is in the residue in the P1' position (38). While some of these differences may be attributed

to the requirement for NS3 proteases to bind to lineage-specific activating cofactors to further augment cleavage specificity (39–47), it is also likely that these changes in cleavage specificity are dependent on non-conserved residues in the binding pocket of the NS3 protease (48). Similarly, within *Coronaviridae*, replication competent chimeric murine hepatitis virus (MHV) could not be recovered when the 3CL protease was replaced with one of many related alpha- or beta-coronavirus proteases including SARS-CoV, hCoV-229E and bat CoV-HKU4 (**Figure 3B**). Only when the MHV 3CL was replaced with the two most closely related beta-coronaviruses, hCoV-OC43 and hCoV-HKU1, could virus be recovered, but with a substantial fitness cost (49). Altogether, these biochemical and chimeric virus studies illustrate that (+)ssRNA viruses have undergone lineage-specific evolution in both their protease sequence specificity as well as their many polyprotein cleavage sites.

(+)SSRNA VIRAL PROTEASES HAVE EVOLVED IN CONFLICT WITH THEIR HOSTS

The above-described changes in protease sequence specificity do not require invocation of adaptation. Indeed, evolutionary drift could result in changes to the viral protease and its cleavage sites, including those that result in loss of fitness for chimeric viruses. However, there is another selective pressure that likely shapes viral protease evolution: the advantage that viruses gain by cleaving host targets. Proteins in multiple cellular processes have been identified as targets of viral proteases, many of which are involved in the host antiviral immune response (4). Many of these host targets are divergent between species, potentially establishing molecular barriers to cross-species transmission. Although the ability to cleave the viral polyprotein is an invariant function of viral proteases, we posit that cleavage of specific host proteins may be selected for during viral evolution, especially during or following cross-species transmission. Indeed, pathogenicity of a mouse-adapted SARS coronavirus required two mutations in 3CL to facilitate rapid, robust virus replication (50). Although it has not been established whether these 3CL changes result in changes in host target cleavage, these data indicate that protease evolution may be required for successful adaptation to a novel host species.

Several excellent reviews have been written describing the diverse host targets that are cleaved by viral proteases (4, 16, 17, 19, 51). In many cases, the described host-virus interaction has focused on a single or a small number of related viral proteases and only a single host species, often humans. Thus, the importance of host and virus diversity in these interactions is often poorly understood. However, evidence is accumulating that viral proteases and their host targets are engaged in species-specific interactions. Below, we highlight such cases in which host and viral diversity alter the outcome of the interaction between host pathways and proteases of picornaviruses, flaviviruses and coronaviruses, illustrating this ongoing molecular arms race.

Viral Proteases Target Essential Host Processes in a Virus-Specific Manner

Some of the best studied targets of viral proteases, especially from picornaviruses, are involved in well conserved processes such as translation initiation or translation control (52, 53). In many cases, the functional outcome is similar: viral proteases antagonize a host molecular function in a way that benefits the virus. However, the specific host protein or specific site within that host protein can be divergent between different viruses, highlighting differences in protease cleavage specificity between related viruses, as well as the convergence of viral protease cleavage on the same host pathways. Thus, even for host functions that are 'well conserved' targets of protease cleavage, there is surprising mechanistic diversity. Below we highlight two such examples in well described targets of picornavirus proteases, but likely many other similar examples exist.

Translation of picornavirus mRNAs occurs *via* an internal ribosome entry site (IRES) (54). This bypasses the need to engage with host cap-dependent translation machinery and offers the opportunity to induce a 'host-shutoff' of translation of host antiviral proteins while maintaining production of viral proteins. Many picornaviruses inhibit host translation in a protease-dependent manner *via* cleavage of subunits of the eIF4F cap-binding complex, which binds to host mRNA cap structures to establish the initiation complex, or poly-A binding protein (PABP), which binds the 3' polyA tail of mRNAs and eIF4G to circularize mRNAs for optimal translation initiation (55) (**Figure 4A**). For instance, Foot-and-mouth disease virus (FMDV) 3C cleaves the eIF4G and eIF4A subunits of eIF4F (56). Interestingly, neither hepatitis A (HepA) virus nor encephalomyocarditis virus (EMCV) 3C target eIF4G for cleavage, but both target PABP (57, 58). Convergently, PV also targets PABP, but at a site that is ~100 residues away from the cleavage site of EMCV (59) and additionally uses its 2A protease to cleave eIF4G (60). Despite cleaving different host targets and/or host sites, these interactions all result in host translation shut-off. These data highlight functional conservation, rather than molecular conservation, of picornavirus 3C-mediated inhibition of host translation and suggests that even among related viruses, there are important differences in the viral specificity of host target cleavage.

A similar phenomenon is observed in another well-established target of picornavirus 3C proteases, the stress granule protein G3BP1 (**Figure 4B**). Numerous viruses manipulate stress granule formation for their benefit, as this is a major intersection point between translation control and cellular stress responses (53, 61). Among the mapped cleavage sites in G3BP1, PV 3C cleaves at Q326 (62) while FMDV 3C cleaves at E284 (63), but both of these cleavage events benefit the virus by manipulating stress granule formation. These findings further demonstrate the convergence of 3C cleavage onto the same host target, while highlighting how subtle differences in cleavage specificity can impact viral targeting of host factors. Of note, the P1 and P3' positions of the PV cleavage site are altered in a way that would prevent cleavage of mouse G3BP1, which is otherwise >90% identical to the human protein (**Figure 4B**).

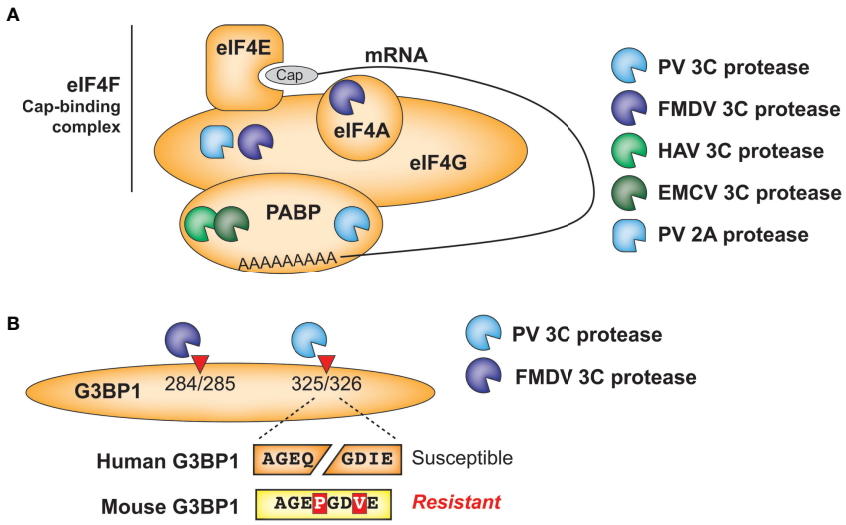


FIGURE 4 | Antagonism of host cellular processes by viral proteases. **(A)** Diverse viral proteases inhibit translation of host mRNA through cleavage of initiation factors and/or poly(A)-binding protein. **(B)** Host and virus species-specific cleavage of the stress granule protein G3BP1 by picornavirus proteases.

Whether host G3BP1 is cleaved by enteroviruses that infect rodents, and at what site, has yet to be determined.

Proteins in the Innate Antiviral Immune Response Are Common Targets of Viral Proteases

n addition to essential cellular processes, proteins in the innate antiviral immune response are common targets of diverse viral proteases. The host antiviral response is initiated when cells detect viral products (**Figure 5A**). Following entry into a host cell, viral nucleic acids can be detected by host pattern recognition receptors (PRRs) such as RIG-I, MDA5, and cGAS. While RIG-I and MDA5 directly detect viral ssRNA or dsRNA as a product of (+)ssRNA virus replication (67–71), the cytosolic DNA sensor cGAS can be indirectly activated *via* virus-induced mitochondrial damage and subsequent release of mitochondrial DNA that can occur during (+)ssRNA viral infection (72). After ligand binding, PRRs recruit a series of adaptor proteins, ultimately resulting in the production and secretion of type I and III interferons (IFN-I and IFN-III) (73). IFN-I and IFN-III are antiviral cytokines that signal in an autocrine or paracrine manner to induce expression of interferon-stimulated genes (ISGs), which act to directly and indirectly inhibit virus replication and establish an antiviral state in the host (74) (**Figure 5A**).

Induction of IFN and subsequent upregulation of ISGs is critical to the host antiviral defense. Therefore, proteins involved in these pathways are common targets of viral antagonism (73), including several that are cleaved by (+)ssRNA viral proteases (**Table 2** and **Figure 5A**). For instance, NS3 from DENV and other flaviviruses can cleave and inactivate STING to prevent sensing of cytoplasmic mitochondrial DNA (65, 75), whereas PV and possibly other 3C proteases cleave RIG-I during infection (76, 96) (**Figure 5A**). Tellingly, many proteases convergently cleave the same host targets. For instance, CVB3 3C and HCV NS3 are both able to cleave MAVS (77, 78), a critical innate immune adaptor for both MDA5 and RIG-I (**Figure 5A**). 3CL proteases from Porcine Epidemic Diarrhea Virus (PEDV), porcine deltacoronavirus (PDCoV), and feline infectious peritonitis virus (FIPV), as well as 3C proteases from FMDV and HepA can also inhibit RIG-I/MDA5 pathways by cleaving nuclear transcription factor κB (NF-κB) essential modulator (NEMO), a bridging adaptor protein involved in activating both NF-κB and interferon-regulatory factor signaling pathways (89–93, 97) (**Figure 5A**). Finally, STAT2, one of the critical transcription factors that transmits the signaling of IFN to ISG production (**Figure 5A**), is cleaved by the 3CL from PDCoV (85), although whether other 3CLs cleave this protein is unknown. Altogether these data show that viral protease-mediated cleavage of innate immune signaling proteins is a common strategy across (+)ssRNA viruses to prevent the antiviral response and promote virus replication.

Many proteins in the innate antiviral immune response are rapidly evolving within and between host populations (98–100). One potential consequence of these host changes is that a cleavage site for a viral protease may be present in one host but not another. If there is strong selection for the virus to restore antagonism of that host function, there would be selection for viral proteases that would change the sequence specificity of host target cleavage to either restore cleavage of the original site, cleave another site on the host protein, or cleave another protein in the host pathway (**Figure 5B**). Such an evolutionary model can be used to understand the genetic bases for host- and viral-specificity of protease cleavage. For many of the known interactions between host immunity proteins and viral proteases, there is little information on how host and viral evolution shapes the outcome. However, analyses on two host

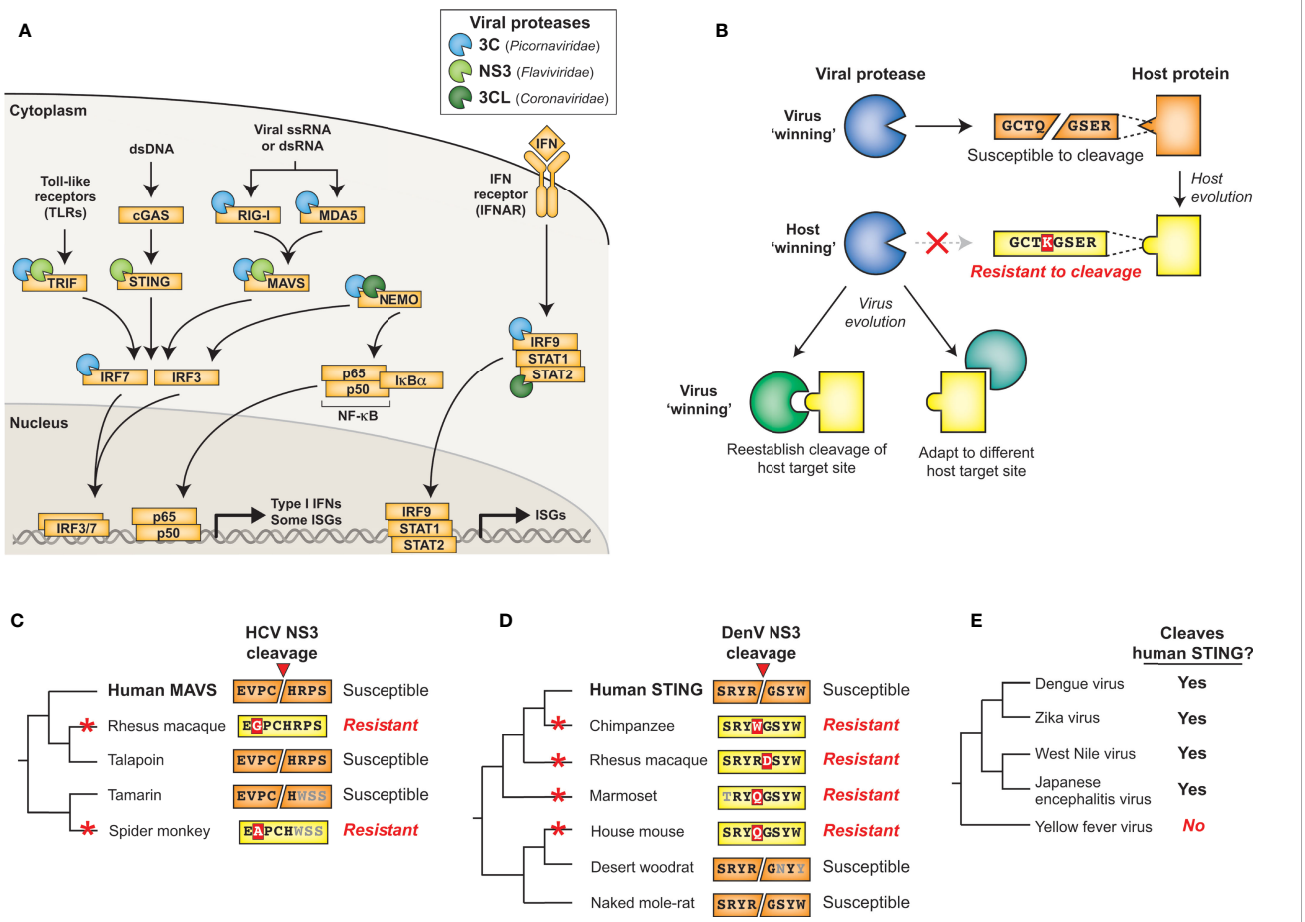


FIGURE 5 | Protease antagonism of IFN induction and signaling pathways. **(A)** Examples of viral proteases that antagonize the innate antiviral immune response, including antagonism of IFN-induction (left) or signaling downstream of IFN (right). **(B)** Model for how protease sequence specificity may be driven to evolve by conflicts with host factors. Following host evolution, or cross-species transmission, viral proteases may no longer be able to antagonize a given host factor. To re-establish host antagonism, the protease can evolve to cut a different sequence at same host site (left) or may evolve to cut a new site elsewhere in the host protein (right). **(C, D)** Evolution of MAVS (64) **(C)** and STING (65) **(D)** across primates and other mammals confers resistance or susceptibility to flaviviral protease cleavage. Red asterisks mark the inferred branch in which an amino acid change occurred that alters cleavage susceptibility. **(E)** Human STING cleavage by flavivirus NS3 proteases is virus species-specific. Data adapted from (66).

TABLE 2 | Select list of IFN pathway-related targets of (+)ssRNA virus proteases.

Host target	Viral protease	References
STING	NS3 (NS2B3) (ZIKV, JEV, WNV, YFV, DENV)	(65, 66, 75)
RIG-I	3C (PV)	(76)
MAVS	NS3 (NS3-4A) (HCV, GBV-B)	(64, 77–82)
	3C (CVB3, SVV)	
Riplet	NS3 (NS3-4A) (HCV)	(83)
MDA5	3C (FMDV)	(84)
STAT2	3CL (PDCoV)	(85)
TRIF	NS3 (NS3-4A) (HCV)	(78, 81, 86–88)
	3C (CVB3, SVV, EV68)	
	3CD (HAV)	
NEMO	3CL (PEDV, FIPV, PDCoV)	(89–93)
	3C (FMDV, HAV)	
IRF7	3C (EV68)	(94)
IRF9	3C (EV71)	(95)

Viral abbreviations are as follows: ZIKV, Zika virus; JEV, Japanese encephalitis virus; WNV, West Nile virus; YFV, Yellow fever virus; DENV, Dengue virus; PV, Poliovirus; HCV, Hepatitis C virus; GBV-B, GB virus B or Pegivirus B; CVB3, Coxsackievirus B3; SVV, Seneca Valley virus; FMDV, Foot and mouth disease virus; PDCoV, Porcine deltacoronavirus; EV68, Enterovirus D68; HAV, Hepatitis A virus; PEDV, Porcine epidemic diarrhea virus; FIPV, Feline infectious peritonitis virus.

targets, described in more detail below, provide evidence for an arms races between host immunity proteins and viral proteases.

MAVS and STING Have Evolved in Conflict With Viral Proteases

One well-characterized instance of viral proteases shaping host gene evolution is in HCV NS3 protease antagonism of the host protein MAVS. MAVS serves as a critical signaling node to integrate signals from the nucleic acid sensors RIG-I and MDA5 to downstream IFN production (**Figure 5A**). Early observations indicated that MAVS cleavage by HCV NS3 was site specific and important for viral evasion of the immune system (101). Subsequent evolutionary analyses revealed that one residue within the HCV cleavage site in MAVS has evolved under recurrent positive selection, suggestive that MAVS evolution has been shaped by NS3 antagonism (64). Variation at this site across primates affects susceptibility to cleavage by HCV NS3. Importantly, primate MAVS proteins that have evolved resistance to cleavage retain a functional IFN response during HCV infection, providing a potential explanation for the restricted host range of HCV (**Figure 5C**) (64). This work also identified a site evolving under positive selection that is known to be antagonized by the CVB3 3C protease, and variation at this site across primates could also alter protease-mediated antagonism and antiviral signaling through MAVS (64, 78).

Another adaptor protein that connects nucleic acid sensing to the IFN response is STING, which operates downstream of the cytoplasmic DNA sensor cGAS (**Figure 5A**). Originally described as a species-specific target of DENV NS3 cleavage (65, 75), STING has evolved under positive selection in primates and the NS3 cleavage site within STING contains several amino acid differences across primates that alter the outcome of cleavage (**Figure 5D**) (65). Expanding this analysis to a broader panel of mammals, the NS3 site of cleavage in human STING has evolved to be cleavage resistant in mice, pigs, and ground squirrels, whereas naked mole rat and desert woodrat are susceptible to cleavage (65) (**Figure 5D**). Interestingly, differences in protein sequences that affect cleavage do not just occur between host species; polymorphisms within a host can also alter the ability of a viral protease to cleave a given target. Evidence of this process can be observed in human STING polymorphisms, where the three most common human STING haplotypes are differentially cleaved by DENV NS3 (102). Not only is host diversity important, but viral diversity is as well. For instance, ZIKV, DENV, JEV, and WNV NS2B3 can cleave human but not mouse STING, whereas YFV NS3 cannot cleave STING from either species (66) (**Figure 5E**). Additional work to identify more divergent flaviviral protease interactions will further define evolution of STING antagonism.

IMMUNE SENSORS OF VIRAL PROTEASE ACTIVITY: WHO IS CHASING WHOM?

Cleavage of host proteins by viral proteases often inactivates the host protein and results in a fitness advantage for the virus. In these cases, host evolutionary signatures reveal adaptations that are presumed to evade cleavage. However, another possibility exists, in which the host protein can sense the presence of the viral protease in the cytoplasm through an evolved sequence that mimics the viral polyprotein cleavage site. Sensing of pathogen-encoded activities such as toxins and effector enzymes, known as effector-triggered immunity (ETI), is well-described in plants but is also emerging as an important immune mechanism in animals (103–106). Three such signaling pathways, described below, are known to detect the main protease activity of human viruses.

NLRP1 Mimics Diverse Picornaviral 3C Cleavage to Trigger Inflammation

One of the best described cases of mammalian ETI involves NLRP1 (NACHT, LRR, and PYD domains-containing protein 1; **Figure 6A**), a critical sensor for the innate immune complex known as the inflammasome. Mouse NLRP1B was identified in a genetic screen as a determinant of differential susceptibility between mouse strains to Lethal Toxin, a virulence factor responsible for the major pathologies seen during infection by the bacterial pathogen *Bacillus anthracis* (107). Further research identified that NLRP1B was a target of cleavage by the secreted bacterial protease component of Lethal Toxin, termed Lethal Factor (LF). Interestingly, mice with a cleavage-susceptible variant of NLRP1B were protected from *B. anthracis* challenge, indicating that cleavage of NLRP1B was immunologically protective (108, 109). The mechanism by which this occurs, termed 'functional degradation' (110, 111), depends on the FIIND domain encoded within NLRP1B, which undergoes a constitutive self-cleavage event (known as 'auto-processing') such that the N-terminal domains and C-terminal CARD-containing fragment of NRLP1B exist as two distinct, noncovalently associated polypeptides (112, 113) (**Figure 6B**). Once LF cleaves upstream of the FIIND domain in NLRP1B, the released product has a new N-terminus that is recognized by the N-end rule cellular machinery and targets it for proteasome-mediated degradation. However, as a result of the break in the polyprotein backbone within the FIIND domain, proteasome-mediated degradation of NLRP1B ceases after degrading the N-terminal domains, leaving the bioactive C-terminal fragment intact and able to assemble into an active inflammasome (**Figure 6B**) (110, 111). The unusual domain architecture of NLRP1B thus facilitates the mounting of the inflammasome response upon proteolytic cleavage of the N-terminus.

LF cleaves within the rapidly evolving 'tripwire' region of mouse NLRP1B but fails to cleave or activate human NLRP1. Interestingly, human NLRP1 has an analogous rapidly-evolving 'tripwire' region (**Figure 6A**), and cleavage of human NLRP1 *via* an engineered tobacco etch virus (TEV) protease cleavage site can activate the inflammasome (114). These data suggested that human NLRP1 may also detect pathogen-encoded proteases and activate the inflammasome *via* a functional degradation mechanism. Indeed, we and others recently identified that human NLRP1 recognizes picornavirus 3C protease activity and serves as a tripwire for inflammatory cell death and downstream

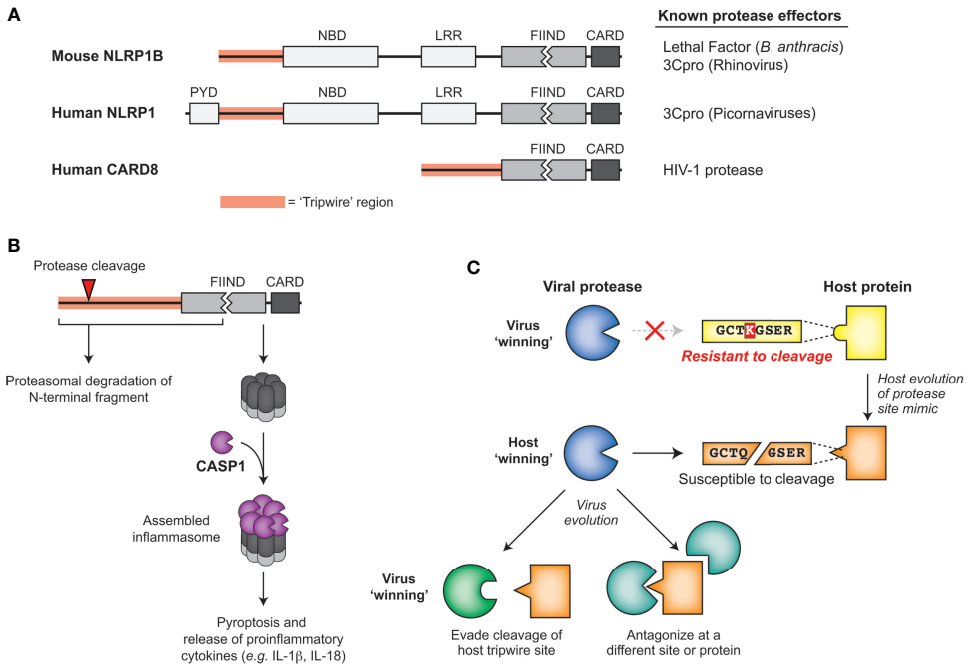


FIGURE 6 | Sensing of pathogen-encoded protease activities by host 'tripwires'. **(A)** NLRP1 and CARD8 serve as effector-triggered immunity (ETI) sensors to detect cleavage by viral proteases. Schematic of mouse NLRP1B, human NLRP1, and human CARD8, highlighting the tripwire region (left) and the known protease effectors (right). **(B)** Model for how protease cleavage initiates functional degradation of the N-terminal region of inflammasome activators. Activation recruits and activates caspase-1, which cleaves multiple host proteins, including processing proinflammatory cytokines such as IL-1β, into their mature, bioactive form. **(C)** Model for how evolution of host protease site mimics may drive viral protease evolution to either evade cleavage of the host tripwire or antagonize the host in other parts of the protein or pathway.

inflammatory signaling (15, 115). During enterovirus infection, 3C cleavage of NLRP1 results in assembly of the active inflammasome and subsequent pro-inflammatory cytokine release (15, 115), including in human primary airway epithelial cells (115). Interestingly, based on phylogenetic analyses, the 3C-protease site mimic in this specific region of NLRP1 only evolved in the primate lineage, and is only cleavable in some primates (15). Differences across simian primates and a SNP within the human population prevent cleavage and inflammasome activation (15). Although mice lack this human-aligned cleavage site, we discovered a similar phenomenon where picornavirus 3C proteases cleave NLRP1B at different sites to activate the inflammasome in a virus- and mouse-strain-specific manner (15).

In addition to host diversity, viral diversity also determines NLRP1 cleavage. While all enteroviruses cleave the same site within NLRP1 and activate the inflammasome, other picornaviruses cleave NLRP1 at different sites within the N-terminal domain or do not cleave NLRP1 (15). For instance, the 3C protease of EMCV does not cleave NLRP1, and resultingly no activation of the NLRP1 inflammasome was observed upon EMCV infection (15). As numerous sites in the protease-sensing N-terminal region of NLRP1 are evolving under positive selection (114), other independently evolved tripwire sites within NLRP1 may sense divergent 3C or other viral proteases.

Intracellular HIV-1 Protease Activity Triggers Inflammation *via* CARD8

Another inflammasome mediator, CARD8, is known to share the unusual C-terminal domain structure critical for the sensing mechanism of NLRP1 – the FIIND domain followed by a CARD domain (**Figure 6A**) (116, 117). In addition to these domain similarities, CARD8 inflammasome assembly can also be activated by the same small molecules as NLRP1 (117). Such similarities initially suggested that CARD8 could also be activated using a functional degradation model to act as a tripwire sensor of pathogen-encoded activities (116). Indeed, the protease of human immunodeficiency virus 1 (HIV-1) can cleave and activate the CARD8 inflammasome in an activation mechanism that resembles NLRP1 (118). While HIV-1 protease is normally important for cleaving viral polyproteins in the maturing capsid, treatment with specific non-nucleoside reverse transcriptase inhibitors (NNRTIs) can result in protease activity in the cytoplasm (119). Under these NNRTI treatment conditions, HIV-1 proteases from four prevalent HIV-1 subtypes cleave CARD8 and activate the inflammasome, resulting in pro-inflammatory cytokine release and influencing clearance of latent HIV-1 in primary CD4+ T cells (118). While the extent to which host evolution or evolution of other viruses influences the activation of the CARD8 system remains unknown, these findings reveal a broader role for host encoded

tripwires for viral proteases that can activate a robust immune response using mimicry of viral protease cleavage sites.

3C-Mediated Cleavage of a Regulator of NF-kB Triggers Apoptosis

'Tripwire' mechanisms such as NLRP1 and CARD8 rely on a specific elegant, but rare, domain architecture that allows for coupling of a cleavage event to generation of a bioactive signaling molecule. An additional mechanism for sensing of viral proteases arises from the intricate ways that the innate immune response is negatively regulated. For instance, downstream of NLRP1 and CARD8, mature inflammatory cytokines are detected by the IL-1 receptor to activate the transcription factor NF-κB, which can amplify the inflammatory response (120). NF-κB is an essential transcription factor involved in many innate immune pathways and can mediate a variety of downstream responses depending on the input stimuli (121), including pro- or anti-apoptotic responses (122). Within the cytoplasm, the NF-κB heterodimer, composed of the Rel family proteins p65 and p50, remains bound and inactive by members of the inhibitor of κB (IκB) family, including IκBα (**Figure 5A**). In response to cytokines such as IL-1β, IκB kinase (IKK) family proteins phosphorylate IκB proteins, releasing the active transcription factor to translocate into the nucleus (120). A previous study demonstrated that IκBα senses CVB3 3C protease activity (123). The 3C protease was shown to cleave IκBα, producing a fragment that stably complexes with p65 and translocates to the nucleus. This stable complex blocks NF-κB transcriptional activation, resulting in increased cell apoptosis and decreased viral replication (123). Thus, cleavage of IκBα may have evolved as another way to sense viral protease activity and induce cell death to prevent further virus propagation. Many viral proteases are known to cleave proteins in the NF-κB pathway (**Table 2**). Additional characterization of these virus-host interactions may reveal additional antiviral mechanisms associated with this critical immune pathway.

Evolutionary Advantages of ETI

In the continual evolutionary conflict between viruses and their hosts, cleavage mimicry encoded in NLRP1, CARD8 and NF-kB serve as examples of a successful strategy emerging in host organisms to exploit highly constrained pathogenic processes. Viruses are known to use molecular mimicry to antagonize or subvert the host immune response (124). In the cases of ETI described above, the host is turning the tables and using mimicry of viral protease cleavage sites to support the antiviral response. Rather than mimicry of entire proteins or protein domains, mimicry of these cleavage sites as 'short linear motifs' (SLIMs) require only a small number of amino acids to hijack the highly conserved protease activity (125, 126). In order to avoid these 'tripwires' and negative regulators of the immune response, these viruses must either evolve their respective main proteases along with all affiliate cleavage sites or antagonize the process some other way (**Figure 6C**). Supporting this idea, 3C proteases from some picornaviruses cleave NLRP1 but do not activate the NLRP1 inflammasome, suggesting that 3C proteases have evolved to evade detection by NLRP1 by antagonizing NLRP1 function elsewhere (15). We expect that this work may lead to the discovery that protease-driven ETI strategies may have evolved more broadly at other sites of host-pathogen conflicts.

DISCUSSION

The proteases of (+)ssRNA viruses have multiple roles in establishing and maintaining virus infection within a host. First and foremost, virally-encoded proteases cleave numerous sequence-specific sites within the viral polyprotein, which is essential for completion of the viral replication cycle. As a consequence of this essential activity, the ability of proteases to evolve novel sequence specificity is highly constrained. However, viral proteases also serve to manipulate numerous host processes in the infected cell through site-specific cleavage of host targets. In this context, changes in protease sequence specificity would allow the virus to cleave new host targets that might benefit the virus, or avoid cleaving host targets that are detrimental to the virus. It is at this intersection that viral proteases are engaged in evolutionary 'arms races' with the host, resulting in varied interactions across viral and host species and across evolutionary time. Several examples, including virus-specific cleavage of essential mRNA translation machinery and host-specific evasion of cleavage of innate antiviral immune components, highlight the consequences of these evolutionary conflicts. More recently, the discovery of host-encoded effector-triggered immunity (ETI) sensors such as NLRP1 and CARD8 suggest that host mimicry of viral protease cleavage sites is an efficient strategy to detect the cellular activity of viral proteases.

The extent to which viral protease evolution, and host target diversity, shape viral host range and pathogenesis remains unknown and is an exciting area of future research. The majority of characterized protease-host interactions have been described for a single virus against a single host, leaving open the opportunity for more detailed exploration of the evolutionary dynamics of these interactions. Indeed, examples such as cleavage of host proteins such as MAVS, STING, and NLRP1 highlight the insights that can be gained from additional analyses of host and viral diversity in these interactions. Likewise, future studies aiming to discover additional host targets of viral proteases, especially those that may be cleaved in a virus-specific manner, will advance our knowledge of the ways that protease-host interactions shape viral phenotypes. Finally, ETI sensors such as NLRP1 and CARD8 may represent just the start of host proteins that mimic viral protease cleavage sites to induce an immune response. Further studies aimed to identify ETI mechanisms against both viral and other pathogen-encoded proteases will likely continue to reveal novel mechanisms and evolutionary principles of the host innate immune response.

AUTHOR CONTRIBUTIONS

All authors discussed relevant literature. BT, EF, and MD wrote the first draft. All authors contributed to figure generation and

editing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Institutes of Health (R35 GM133633), Pew Biomedical Scholars Program, Hellman Fellows Program, and Burroughs Wellcome Investigators in the Pathogenesis of Infectious Disease Program to MD and an NIH T32 grant (GM007240) to BT.

ACKNOWLEDGMENTS

We thank all members of the Daugherty laboratory for helpful discussions. We apologize to the colleagues whose work was not included in this review owing to space limitations.

REFERENCES

1. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Dempsey DM, et al. Changes to Virus Taxonomy and the Statutes Ratified by the International Committee on Taxonomy of Viruses (2020). *Arch Virol* (2020) 165:2737–48. doi: 10.1007/s00705-020-04752-x
2. Anirudhan V, Lee H, Cheng H, Cooper L, Rong L. Targeting SARS-CoV-2 Viral Proteases as a Therapeutic Strategy to Treat COVID-19. *J Med Virol* (2021) 93:2722–34. doi: 10.1002/jmv.26814
3. Steuber H, Hilgenfeld R. Recent Advances in Targeting Viral Proteases for the Discovery of Novel Antivirals. *Curr Top Med Chem* (2010) 10:323–45. doi: 10.2174/156802610790725470
4. Lei J, Hilgenfeld R. RNA-Virus Proteases Counteracting Host Innate Immunity. *FEBS Lett* (2017) 591:3190–210. doi: 10.1002/1873-3468.12827
5. Daugherty MD, Malik HS. Rules of Engagement: Molecular Insights From Host-Virus Arms Races. *Annu Rev Genet* (2012) 46:677–700. doi: 10.1146/annurev-genet-110711-155522
6. Duggal NK, Emerman M. Evolutionary Conflicts Between Viruses and Restriction Factors Shape Immunity. *Nat Rev Immunol* (2012) 12:687–95. doi: 10.1038/nri3295
7. Meyerson NR, Sawyer SL. Two-Stepping Through Time: Mammals and Viruses. *Trends Microbiol* (2011) 19:286–94. doi: 10.1016/j.tim.2011.03.006
8. Rothenburg S, Brennan G. Species-Specific Host-Virus Interactions: Implications for Viral Host Range and Virulence. *Trends Microbiol* (2020) 28:46–56. doi: 10.1016/j.tim.2019.08.007
9. Van Valen L. A New Evolutionary Law. *Evol Theory* (1973) 1:1–30. doi: 10.7208/9780226115504-
10. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary Insights Into Host-Pathogen Interactions From Mammalian Sequence Data. *Nat Rev Genet* (2015) 16:224–36. doi: 10.1038/nrg3905
11. Diehl WE, Lin AE, Grubaugh ND, Carvalho LM, Kim K, Kyawe PP, et al. Ebola Virus Glycoprotein With Increased Infectivity Dominated the 2013-2016 Epidemic. *Cell* (2016) 167:1088–98.e6. doi: 10.1016/j.cell.2016.10.014
12. Martin DP, Weaver S, Tegally H, San JE, Shank SD, Wilkinson E, et al. The Emergence and Ongoing Convergent Evolution of the SARS-CoV-2 N501Y Lineages. *Cell* (2021) 184:5189–200.e7. doi: 10.1016/j.cell.2021.09.003
13. Sawyer SL, Elde NC. A Cross-Species View on Viruses. *Curr Opin Virol* (2012) 2:561–8. doi: 10.1016/j.coviro.2012.07.003
14. Wain LV, Bailes E, Bibollet-Ruche F, Decker JM, Keele BF, Van Heuverswyn F, et al. Adaptation of HIV-1 to Its Human Host. *Mol Biol Evol* (2007) 24:1853–60. doi: 10.1093/molbev/msm110
15. Tsu BV, Beierschmitt C, Ryan AP, Agarwal R, Mitchell PS, Daugherty MD. Diverse Viral Proteases Activate the NLRP1 Inflammasome. *Elife* (2021) 10:e60609. doi: 10.7554/eLife.60609
16. Laitinen OH, Svedin E, Kapell S, Nurminen A, Hytonen VP, Flodstrom-Tullberg M. Enteroviral Proteases: Structure, Host Interactions and Pathogenicity. *Rev Med Virol* (2016) 26:251–67. doi: 10.1002/rmv.1883
17. Ng CS, Stobart CC, Luo H. Innate Immune Evasion Mediated by Picornaviral 3C Protease: Possible Lessons for Coronaviral 3C-Like Protease? *Rev Med Virol* (2021) e2206. doi: 10.1002/rmv.2206
18. V'Kovski P, Kratzel A, Steiner S, Stalder H, Thiel V. Coronavirus Biology and Replication: Implications for SARS-CoV-2. *Nat Rev Microbiol* (2021) 19:155–70. doi: 10.1038/s41579-020-00468-6
19. Yi J, Peng J, Yang W, Zhu G, Ren J, Li D, et al. Picornavirus 3C - a Protease Ensuring Virus Replication and Subverting Host Responses. *J Cell Sci* (2021) 134:jcs253237. doi: 10.1242/jcs.253237
20. Blom N, Hansen J, Blaas D, Brunak S. Cleavage Site Analysis in Picornaviral Polyproteins: Discovering Cellular Targets by Neural Networks. *Protein Sci* (1996) 5:2203–16. doi: 10.1002/pro.5560051107
21. Roe MK, Junod NA, Young AR, Beachboard DC, Stobart CC. Targeting Novel Structural and Functional Features of Coronavirus Protease Nsp5 (3CL(Pro), M(pro)) in the Age of COVID-19. *J Gen Virol* (2021) 102:1558. doi: 1099/jgv.0.001558
22. Clarke IN, Lambden PR. Organization and Expression of Calicivirus Genes. *J Infect Dis* (2000) 181 Suppl 2:S309–16. doi: 10.1086/315575
23. Valli AA, Gallo A, Rodamilans B, Lopez-Moya JJ, Garcia JA. The HCPro From the Potyviridae Family: An Enviable Multitasking Helper Component That Every Virus Would Like to Have. *Mol Plant Pathol* (2018) 19:744–63. doi: 10.1111/mpp.12553
24. Ten Dam E, Flint M, Ryan MD. Virus-Encoded Proteinases of the Togaviridae. *J Gen Virol* (1999) 80(Pt 8):1879–88. doi: 10.1099/0022-1317-80-8-1879
25. Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus Genome Organization, Expression, and Replication. *Annu Rev Microbiol* (1990) 44:649–88. doi: 10.1146/annurev.mi.44.100190.003245
26. Barrows NJ, Campos RK, Liao KC, Prasanth KR, Soto-Acosta R, Yeh SC, et al. Biochemistry and Molecular Biology of Flaviviruses. *Chem Rev* (2018) 118:4448–82. doi: 10.1021/acs.chemrev.7b00719
27. Lindenbach BD, Rice CM. Evasive Maneuvers by Hepatitis C Virus. *Hepatology* (2003) 38:769–71. doi: 10.1002/hep.510380327
28. Winston DS, Boehr DD. The Picornavirus Precursor 3cd Has Different Conformational Dynamics Compared to 3C(Pro) and 3D(Pol) in Functionally Relevant Regions. *Viruses* (2021) 13:442. doi: 10.3390/v13030442
29. Moradpour D, Penin F. Hepatitis C Virus Proteins: From Structure to Function. *Curr Top Microbiol Immunol* (2013) 369:113–42. doi: 10.1007/978-3-642-27340-7_5
30. Baggen J, Thibaut HJ, Strating J, van Kuppeveld FJM. The Life Cycle of Non-Polio Enteroviruses and How to Target it. *Nat Rev Microbiol* (2018) 16:368–81. doi: 10.1038/s41579-018-0005-4
31. Navia MA, McKeever BM. A Role for the Aspartyl Protease From the Human Immunodeficiency Virus Type 1 (HIV-1) in the Orchestration of Virus Assembly. *Ann N Y Acad Sci* (1990) 616:73–85. doi: 10.1111/j.1749-6632.1990.tb17829.x
32. Kiemer L, Lund O, Brunak S, Blom N. Coronavirus 3clpro Proteinase Cleavage Sites: Possible Relevance to SARS Virus Pathology. *BMC Bioinf* (2004) 5:72. doi: 10.1186/1471-2105-5-72
33. Stanley JT, Gilchrist AR, Stabell AC, Allen MA, Sawyer SL, Dowell RD. Two-Stage ML Classifier for Identifying Host Protein Targets of the Dengue Protease. *Pac Symp Biocomput* (2020) 25:487–98. doi: 10.1142/9789811215636_0043
34. Li J, Lim SP, Beer D, Patel V, Wen D, Tumanut C, et al. Functional Profiling of Recombinant NS3 Proteases From All Four Serotypes of Dengue Virus Using Tetrapeptide and Octapeptide Substrate Libraries. *J Biol Chem* (2005) 280:28766–74. doi: 10.1074/jbc.M500588200
35. O'Donoghue AJ, Eroy-Reveles AA, Knudsen GM, Ingram J, Zhou M, Statnekov JB, et al. Global Identification of Peptidase Specificity by Multiplex Substrate Profiling. *Nat Methods* (2012) 9:1095–100. doi: 10.1038/nmeth.2182

36. Dewalt PG, Lawson MA, Colonno RJ, Semler BL. Chimeric Picornavirus Polyproteins Demonstrate a Common 3C Proteinase Substrate Specificity. *J Virol* (1989) 63:3444–52. doi: 10.1128/jvi.63.8.3444-3452.1989
37. Preugschat F, Lenches EM, Strauss JH. Flavivirus Enzyme-Substrate Interactions Studied With Chimeric Proteinases: Identification of an Intragenic Locus Important for Substrate Recognition. *J Virol* (1991) 65:4749–58. doi: 10.1128/jvi.65.9.4749-4758.1991
38. VanBlargan LA, Davis KA, Dowd KA, Akey DL, Smith JL, Pierson TC. Context-Dependent Cleavage of the Capsid Protein by the West Nile Virus Protease Modulates the Efficiency of Virus Assembly. *J Virol* (2015) 89:8632–42. doi: 10.1128/JVI.01253-15
39. Jan LR, Yang CS, Trent DW, Falgout B, Lai CJ. Processing of Japanese Encephalitis Virus Non-Structural Proteins: NS2B-NS3 Complex and Heterologous Proteases. *J Gen Virol* (1995) 76(Pt 3):573–80. doi: 10.1099/0022-1317-76-3-573
40. Falgout B, Pethel M, Zhang YM, Lai CJ. Both Nonstructural Proteins NS2B and NS3 Are Required for the Proteolytic Processing of Dengue Virus Nonstructural Proteins. *J Virol* (1991) 65:2467–75. doi: 10.1128/jvi.65.5.2467-2475.1991
41. Falgout B, Miller RH, Lai CJ. Deletion Analysis of Dengue Virus Type 4 Nonstructural Protein NS2B: Identification of a Domain Required for NS2B-NS3 Protease Activity. *J Virol* (1993) 67:2034–42. doi: 10.1128/jvi.67.4.2034-2042.1993
42. Cahour A, Falgout B, Lai CJ. Cleavage of the Dengue Virus Polyprotein at the NS3/NS4A and NS4B/NS5 Junctions Is Mediated by Viral Protease NS2B-NS3, Whereas NS4A/NS4B may be Processed by a Cellular Protease. *J Virol* (1992) 66:1535–42. doi: 10.1128/jvi.66.3.1535-1542.1992
43. Chambers TJ, Grakoui A, Rice CM. Processing of the Yellow Fever Virus Nonstructural Polyprotein: A Catalytically Active NS3 Proteinase Domain and NS2B Are Required for Cleavages at Dibasic Sites. *J Virol* (1991) 65:6042–50. doi: 10.1128/jvi.65.11.6042-6050.1991
44. Sbardellati A, Scarselli E, Amati V, Falcinelli S, Kekule AS, Traboni C. Processing of GB Virus B Non-Structural Proteins in Cultured Cells Requires Both NS3 Protease and NS4A Cofactor. *J Gen Virol* (2000) 81:2183–8. doi: 10.1099/0022-1317-81-9-2183
45. Butkiewicz N, Yao N, Zhong W, Wright-Minogue J, Ingravallo P, Zhang R, et al. Virus-Specific Cofactor Requirement and Chimeric Hepatitis C Virus/GB Virus B Nonstructural Protein 3. *J Virol* (2000) 74:4291–301. doi: 10.1128/JVI.74.9.4291-4301.2000
46. Xu J, Mendez E, Caron PR, Lin C, Murcko MA, Collett MS, et al. Bovine Viral Diarrhea Virus NS3 Serine Proteinase: Polyprotein Cleavage Sites, Cofactor Requirements, and Molecular Model of an Enzyme Essential for Pestivirus Replication. *J Virol* (1997) 71:5312–22. doi: 10.1128/jvi.71.7.5312-5322.1997
47. Wiskerchen M, Collett MS. Pestivirus Gene Expression: Protein P80 of Bovine Viral Diarrhea Virus Is a Proteinase Involved in Polyprotein Processing. *Virology* (1991) 184:341–50. doi: 10.1016/0042-6822(91)90850-B
48. Junaid M, Chalayut C, Sehgelmeble Torrejon A, Angsuthanasombat C, Shutava I, Lapins M, et al. Enzymatic Analysis of Recombinant Japanese Encephalitis Virus NS2B(H)-NS3pro Protease With Fluorogenic Model Peptide Substrates. *PloS One* (2012) 7:e36872. doi: 10.1371/journal.pone.0036872
49. Stobart CC, Sexton NR, Munjal H, Lu X, Molland KL, Tomar S, et al. Chimeric Exchange of Coronavirus Nsp5 Proteases (3clpro) Identifies Common and Divergent Regulatory Determinants of Protease Activity. *J Virol* (2013) 87:12611–8. doi: 10.1128/JVI.02050-13
50. Roberts A, Deming D, Paddock CD, Cheng A, Yount B, Vogel L, et al. A Mouse-Adapted SARS-Coronavirus Causes Disease and Mortality in BALB/c Mice. *PloS Pathog* (2007) 3:e5. doi: 10.1371/journal.ppat.0030005
51. Chen S, Wu Z, Wang M, Cheng A. Innate Immune Evasion Mediated by Flaviviridae Non-Structural Proteins. *Viruses* (2017) 9:291. doi: 10.3390/v9100291
52. Lloyd RE. Enterovirus Control of Translation and RNA Granule Stress Responses. *Viruses* (2016) 8:93. doi: 10.3390/v8040093
53. McCormick C, Khaperskyy DA. Translation Inhibition and Stress Granules in the Antiviral Immune Response. *Nat Rev Immunol* (2017) 17:647–60. doi: 10.1038/nri.2017.63
54. Lee KM, Chen CJ, Shih SR. Regulation Mechanisms of Viral IRES-Driven Translation. *Trends Microbiol* (2017) 25:546–61. doi: 10.1016/j.tim.2017.01.010
55. Jackson RJ, Hellen CU, Pestova TV. The Mechanism of Eukaryotic Translation Initiation and Principles of its Regulation. *Nat Rev Mol Cell Biol* (2010) 11:113–27. doi: 10.1038/nrm2838
56. Belsham GJ, McInerney GM, Ross-Smith N. Foot-And-Mouth Disease Virus 3C Protease Induces Cleavage of Translation Initiation Factors Eif4a and Eif4g Within Infected Cells. *J Virol* (2000) 74:272–80. doi: 10.1128/JVI.74.1.272-280.2000
57. Kobayashi M, Arias C, Garabedian A, Palmenberg AC, Mohr I. Site-Specific Cleavage of the Host Poly(A) Binding Protein by the Encephalomyocarditis Virus 3C Proteinase Stimulates Viral Replication. *J Virol* (2012) 86:10686–94. doi: 10.1128/JVI.00896-12
58. Zhang B, Morace G, Gauss-Muller V, Kusov Y. Poly(A) Binding Protein, C-Terminally Truncated by the Hepatitis A Virus Proteinase 3C, Inhibits Viral Translation. *Nucleic Acids Res* (2007) 35:5975–84. doi: 10.1093/nar/gkm645
59. Bonderoff JM, Larey JL, Lloyd RE. Cleavage of Poly(A)-Binding Protein by Poliovirus 3C Proteinase Inhibits Viral Internal Ribosome Entry Site-Mediated Translation. *J Virol* (2008) 82:9389–99. doi: 10.1128/JVI.00006-08
60. Gradi A, Svitkin YV, Imataka H, Sonenberg N. Proteolysis of Human Eukaryotic Translation Initiation Factor Eif4gii, But Not Eif4gi, Coincides With the Shutoff of Host Protein Synthesis After Poliovirus Infection. *Proc Natl Acad Sci USA* (1998) 95:11089–94. doi: 10.1073/pnas.95.19.11089
61. White JP, Lloyd RE. Regulation of Stress Granules in Virus Systems. *Trends Microbiol* (2012) 20:175–83. doi: 10.1016/j.tim.2012.02.001
62. White JP, Cardenas AM, Marissen WE, Lloyd RE. Inhibition of Cytoplasmic mRNA Stress Granule Formation by a Viral Proteinase. *Cell Host Microbe* (2007) 2:295–305. doi: 10.1016/j.chom.2007.08.006
63. Ye X, Pan T, Wang D, Fang L, Ma J, Zhu X, et al. Foot-And-Mouth Disease Virus Counteracts on Internal Ribosome Entry Site Suppression by G3BP1 and Inhibits G3BP1-Mediated Stress Granule Assembly *via* Post-Translational Mechanisms. *Front Immunol* (2018) 9:1142. doi: 10.3389/fimmu.2018.01142
64. Patel MR, Loo YM, Horner SM, Gale MJr., Malik HS. Convergent Evolution of Escape From Hepaciviral Antagonism in Primates. *PloS Biol* (2012) 10:e1001282. doi: 10.1371/journal.pbio.1001282
65. Stabell AC, Meyerson NR, Gullberg RC, Gilchrist AR, Webb KJ, Old WM, et al. Dengue Viruses Cleave STING in Humans But Not in Nonhuman Primates, Their Presumed Natural Reservoir. *Elife* (2018) 7:e31919. doi: 10.7554/eLife.31919
66. Ding Q, Gaska JM, Douam F, Wei L, Kim D, Balev M, et al. Species-Specific Disruption of STING-Dependent Antiviral Cellular Defenses by the Zika Virus NS2B3 Protease. *Proc Natl Acad Sci USA* (2018) 115:E6310–8. doi: 10.1073/pnas.1803406115
67. van Kasteren PB, Beugeling C, Ninaber DK, Frias-Staheli N, van Boheemen S, Garcia-Sastre A, et al. Arterivirus and Nairovirus Ovarian Tumor Domain-Containing Deubiquitinases Target Activated RIG-I to Control Innate Immune Signaling. *J Virol* (2012) 86:773–85. doi: 10.1128/JVI.06277-11
68. Luo R, Xiao S, Jiang Y, Jin H, Wang D, Liu M, et al. Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) Suppresses Interferon-Beta Production by Interfering With the RIG-I Signaling Pathway. *Mol Immunol* (2008) 45:2839–46. doi: 10.1016/j.molimm.2008.01.028
69. Deddouche S, Goubau D, Rehwinkel J, Chakravarty P, Begum S, Maillard PV, et al. Identification of an LGP2-Associated MDA5 Agonist in Picornavirus-Infected Cells. *Elife* (2014) 3:e01535. doi: 10.7554/eLife.01535
70. Loo YM, Fornek J, Crochet N, Bajwa G, Perwitasari O, Martinez-Sobrido L, et al. Distinct RIG-I and MDA5 Signaling by RNA Viruses in Innate Immunity. *J Virol* (2008) 82:335–45. doi: 10.1128/JVI.01080-07
71. Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, et al. Differential Roles of MDA5 and RIG-I Helicases in the Recognition of RNA Viruses. *Nature* (2006) 441:101–5. doi: 10.1038/nature04734
72. West AP, Khoury-Hanold W, Staron M, Tal MC, Pineda CM, Lang SM, et al. Mitochondrial DNA Stress Primes the Antiviral Innate Immune Response. *Nature* (2015) 520:553–7. doi: 10.1038/nature14156
73. Hoffmann HH, Schneider WM, Rice CM. Interferons and Viruses: An Evolutionary Arms Race of Molecular Interactions. *Trends Immunol* (2015) 36:124–38. doi: 10.1016/j.it.2015.01.004

74. Schoggins JW. Interferon-Stimulated Genes: Roles in Viral Pathogenesis. *Curr Opin Virol* (2014) 6:40–6. doi: 10.1016/j.coviro.2014.03.006
75. Aguirre S, Maestre AM, Pagni S, Patel JR, Savage T, Gutman D, et al. DENV Inhibits Type I IFN Production in Infected Cells by Cleaving Human STING. *PloS Pathog* (2012) 8:e1002934. doi: 10.1371/journal.ppat.1002934
76. Barral PM, Sarkar D, Fisher PB, Racaniello VR. RIG-I Is Cleaved During Picornavirus Infection. *Virology* (2009) 391:171–6. doi: 10.1016/j.virol.2009.06.045
77. Meylan E, Curran J, Hofmann K, Moradpour D, Binder M, Bartenschlager R, et al. Cardif Is an Adaptor Protein in the RIG-I Antiviral Pathway and Is Targeted by Hepatitis C Virus. *Nature* (2005) 437:1167–72. doi: 10.1038/nature04193
78. Mukherjee A, Morosky SA, Delorme-Axford E, Dybdahl-Sissoko N, Oberste MS, Wang T, et al. The Coxsackievirus B 3C Protease Cleaves MAVS and TRIF to Attenuate Host Type I Interferon and Apoptotic Signaling. *PloS Pathog* (2011) 7:e1001311. doi: 10.1371/journal.ppat.1001311
79. Chen Z, Benureau Y, Rijnbrand R, Yi J, Wang T, Warter L, et al. GB Virus B Disrupts RIG-I Signaling by NS3/4A-Mediated Cleavage of the Adaptor Protein MAVS. *J Virol* (2007) 81:964–76. doi: 10.1128/JVI.02076-06
80. Ferreira AR, Magalhaes AC, Camoes F, Gouveia A, Vieira M, Kagan JC, et al. Hepatitis C Virus NS3-4A Inhibits the Peroxisomal MAVS-Dependent Antiviral Signalling Response. *J Cell Mol Med* (2016) 20:750–7. doi: 10.1111/jcmm.12801
81. Qian S, Fan W, Liu T, Wu M, Zhang H, Cui X, et al. Seneca Valley Virus Suppresses Host Type I Interferon Production by Targeting Adaptor Proteins MAVS, TRIF, and TANK for Cleavage. *J Virol* (2017) 91:e00823-17. doi: 10.1128/JVI.00823-17
82. Anggakusuma RJP, Brown DH, Banda D, Todt G, Vieyres E, Steinmann, et al. Hepacivirus NS3/4A Proteases Interfere With MAVS Signaling in Both Their Cognate Animal Hosts and Humans: Implications for Zoonotic Transmission. *J Virol* (2016) 90:10670–81. doi: 10.1128/JVI.01634-16
83. Vazquez C, Tan CY, Horner SM. Hepatitis C Virus Infection Is Inhibited by a Noncanonical Antiviral Signaling Pathway Targeted by NS3-NS4A. *J Virol* (2019) 93:e00725-19. doi: 10.1128/JVI.00725-19
84. Kim H, Kim AY, Choi J, Park SY, Park SH, Kim JS, et al. Foot-And-Mouth Disease Virus Evades Innate Immune Response by 3C-Targeting of MDA5. *Cells* (2021) 10:271. doi: 10.3390/cells10020271
85. Zhu X, Wang D, Zhou J, Pan T, Chen J, Yang Y, et al. Porcine Deltacoronavirus Nsp5 Antagonizes Type I Interferon Signaling by Cleaving Stat2. *J Virol* (2017) 91:e00003-17. doi: 10.1128/JVI.00003-17
86. Li K, Foy E, Ferreon JC, Nakamura M, Ferreon AC, Ikeda M, et al. Immune Evasion by Hepatitis C Virus NS3/4A Protease-Mediated Cleavage of the Toll-Like Receptor 3 Adaptor Protein TRIF. *Proc Natl Acad Sci USA* (2005) 102:2992–7. doi: 10.1073/pnas.0408824102
87. Qu L, Feng Z, Yamane D, Liang Y, Lanford RE, Li K, et al. Disruption of TLR3 Signaling Due to Cleavage of TRIF by the Hepatitis A Virus Protease-Polymerase Processing Intermediate, 3CD. *PloS Pathog* (2011) 7:e1002169. doi: 10.1371/journal.ppat.1002169
88. Xiang Z, Li L, Lei X, Zhou H, Zhou Z, He B, et al. Enterovirus 68 3C Protease Cleaves TRIF to Attenuate Antiviral Responses Mediated by Toll-Like Receptor 3. *J Virol* (2014) 88:6650–9. doi: 10.1128/JVI.03138-13
89. Chen S, Tian J, Li Z, Kang H, Zhang J, Huang J, et al. Feline Infectious Peritonitis Virus Nsp5 Inhibits Type I Interferon Production by Cleaving NEMO at Multiple Sites. *Viruses* (2019) 12:43. doi: 10.3390/v12010043
90. Wang D, Fang L, Li K, Zhong H, Fan J, Ouyang C, et al. Foot-And-Mouth Disease Virus 3C Protease Cleaves NEMO to Impair Innate Immune Signaling. *J Virol* (2012) 86:9311–22. doi: 10.1128/JVI.00722-12
91. Wang D, Fang L, Shi Y, Zhang H, Gao L, Peng G, et al. Porcine Epidemic Diarrhea Virus 3c-Like Protease Regulates Its Interferon Antagonism by Cleaving NEMO. *J Virol* (2016) 90:2090–101. doi: 10.1128/JVI.02514-15
92. Wang D, Fang L, Wei D, Zhang H, Luo R, Chen H, et al. Hepatitis A Virus 3C Protease Cleaves NEMO to Impair Induction of Beta Interferon. *J Virol* (2014) 88:10252–8. doi: 10.1128/JVI.00869-14
93. Zhu X, Fang L, Wang D, Yang Y, Chen J, Ye X, et al. Porcine Deltacoronavirus Nsp5 Inhibits Interferon-Beta Production Through the Cleavage of NEMO. *Virology* (2017) 502:33–8. doi: 10.1016/j.virol.2016.12.005
94. Xiang Z, Liu L, Lei X, Zhou Z, He B, Wang J. 3c Protease of Enterovirus D68 Inhibits Cellular Defense Mediated by Interferon Regulatory Factor 7. *J Virol* (2016) 90:1613–21. doi: 10.1128/JVI.02395-15
95. Hung HC, Wang HC, Shih SR, Teng IF, Tseng CP, Hsu JT. Synergistic Inhibition of Enterovirus 71 Replication by Interferon and Rupintrivir. *J Infect Dis* (2011) 203:1784–90. doi: 10.1093/infdis/jir174
96. Papon L, Oteiza A, Imaizumi T, Kato H, Brocchi E, Lawson TG, et al. The Viral RNA Recognition Sensor RIG-I Is Degraded During Encephalomyocarditis Virus (EMCV) Infection. *Virology* (2009) 393:311–8. doi: 10.1016/j.virol.2009.08.009
97. Chen J, Wang D, Sun Z, Gao L, Zhu X, Guo J, et al. Arterivirus Nsp4 Antagonizes Interferon Beta Production by Proteolytically Cleaving NEMO at Multiple Sites. *J Virol* (2019) 93:e00385-19. doi: 10.1128/JVI.00385-19
98. Enard D, Cai L, Gwennap C, Petrov DA. Viruses Are a Dominant Driver of Protein Adaptation in Mammals. *Elife* (2016) 5:e12469. doi: 10.7554/eLife.12469
99. Judd EN, Gilchrist AR, Meyerson NR, Sawyer SL. Positive Natural Selection in Primate Genes of the Type I Interferon Response. *BMC Ecol Evol* (2021) 21:65. doi: 10.1186/s12862-021-01783-z
100. Shultz AJ, Sackton TB. Immune Genes Are Hotspots of Shared Positive Selection Across Birds and Mammals. *Elife* (2019) 8:e41815. doi: 10.7554/eLife.41815
101. Li XD, Sun L, Seth RB, Pineda G, Chen ZJ. Hepatitis C Virus Protease NS3/4A Cleaves Mitochondrial Antiviral Signaling Protein Off the Mitochondria to Evade Innate Immunity. *Proc Natl Acad Sci USA* (2005) 102:17717–22. doi: 10.1073/pnas.0508531102
102. Su CI, Kao YT, Chang CC, Chang Y, Ho TS, Sun HS, et al. DNA-Induced 2'3'-cGAMP Enhances Haplotype-Specific Human STING Cleavage by Dengue Protease. *Proc Natl Acad Sci USA* (2020) 117:15947–54.
103. Cui H, Tsuda K, Parker JE. Effector-Triggered Immunity: From Pathogen Perception to Robust Defense. *Annu Rev Plant Biol* (2015) 66:487–511. doi: 10.1146/annurev-arplant-050213-040012
104. Fischer NL, Naseer N, Shin S, Brodsky IE. Effector-Triggered Immunity and Pathogen Sensing in Metazoans. *Nat Microbiol* (2020) 5:528. doi: 10.1038/s41564-020-0682-4
105. Jones JD, Vance RE, Dangl JL. Intracellular Innate Immune Surveillance Devices in Plants and Animals. *Science* (2016) 354:aaf6395. doi: 10.1126/science.aaf6395
106. Vance RE, Isberg RR, Portnoy DA. Patterns of Pathogenesis: Discrimination of Pathogenic and Nonpathogenic Microbes by the Innate Immune System. *Cell Host Microbe* (2009) 6:10–21. doi: 10.1016/j.chom.2009.06.007
107. Boyden ED, Dietrich WF. Nalp1b Controls Mouse Macrophage Susceptibility to Anthrax Lethal Toxin. *Nat Genet* (2006) 38:240–4. doi: 10.1038/ng1724
108. Moayeri M, Crown D, Newman ZL, Okugawa S, Eckhaus M, Cataisson C, et al. Inflammasome Sensor Nlrp1b-Dependent Resistance to Anthrax Is Mediated by Caspase-1, IL-1 Signaling and Neutrophil Recruitment. *PloS Pathog* (2010) 6:e1001222. doi: 10.1371/journal.ppat.1001222
109. Terra JK, Cote CK, France B, Jenkins AL, Bozue JA, Welkos SL, et al. Cutting Edge: Resistance to Bacillus Anthracis Infection Mediated by a Lethal Toxin Sensitive Allele of Nalp1b/Nlrp1b. *J Immunol* (2010) 184:17–20. doi: 10.4049/jimmunol.0903114
110. Chui AJ, Okondo MC, Rao SD, Gai K, Griswold AR, Johnson DC, et al. N-Terminal Degradation Activates the NLRP1B Inflammasome. *Science* (2019) 364:82–5. doi: 10.1126/science.aau1208
111. Sandstrom A, Mitchell PS, Goers L, Mu EW, Lesser CF, Vance RE. Functional Degradation: A Mechanism of NLRP1 Inflammasome Activation by Diverse Pathogen Enzymes. *Science* (2019) 364:eaau1330. doi: 10.1126/science.aau1330
112. D'Osualdo A, Weichenberger CX, Wagner RN, Godzik A, Wooley J, Reed JC. CARD8 and NLRP1 Undergo Autoproteolytic Processing Through a ZU5-Like Domain. *PloS One* (2011) 6:e27396. doi: 10.1371/journal.pone.0027396
113. Finger JN, Lich JD, Dare LC, Cook MN, Brown KK, Duraiswami C, et al. Autolytic Proteolysis Within the Function to Find Domain (FIIND) Is Required for NLRP1 Inflammasome Activity. *J Biol Chem* (2012) 287:25030–7. doi: 10.1074/jbc.M112.378323
114. Chavarria-Smith J, Mitchell PS, Ho AM, Daugherty MD, Vance RE. Functional and Evolutionary Analyses Identify Proteolysis as a General Mechanism for NLRP1 Inflammasome Activation. *PloS Pathog* (2016) 12:e1006052. doi: 10.1371/journal.ppat.1006052

115. Robinson KS, Teo DET, Tan KS, Toh GA, Ong HH, Lim CK, et al. Enteroviral 3C Protease Activates the Human NLRP1 Inflammasome in Airway Epithelia. *Science* (2020) 370:eaay2002. doi: 10.1126/science.aay2002
116. Mitchell PS, Sandstrom A, Vance RE. The NLRP1 Inflammasome: New Mechanistic Insights and Unresolved Mysteries. *Curr Opin Immunol* (2019) 60:37–45. doi: 10.1016/j.coi.2019.04.015
117. Taabazuing CY, Griswold AR, Bachovchin DA. The NLRP1 and CARD8 Inflammasomes. *Immunol Rev* (2020) 297:13–25. doi: 10.1111/imr.12884
118. Wang Q, Gao H, Clark KM, Mugisha CS, Davis K, Tang JP, et al. CARD8 Is an Inflammasome Sensor for HIV-1 Protease Activity. *Science* (2021) 371: eabe1707. doi: 10.1126/science.abe1707
119. Figueiredo A, Moore KL, Mak J, Sluis-Cremer N, de Bethune MP, Tachedjian G. Potent Nonnucleoside Reverse Transcriptase Inhibitors Target HIV-1 Gag-Pol. *PloS Pathog* (2006) 2:e119. doi: 10.1371/journal.ppat.0020119
120. Li X, Massa PE, Hanidu A, Peet GW, Aro P, Savitt A, et al. IKKalpha, IKKbeta, and NEMO/IKKgamma Are Each Required for the NF-Kappa B-Mediated Inflammatory Response Program. *J Biol Chem* (2002) 277:45129–40. doi: 10.1074/jbc.M205165200
121. Oeckinghaus A, Hayden MS, Ghosh S. Crosstalk in NF-kappaB Signaling Pathways. *Nat Immunol* (2011) 12:695–708. doi: 10.1038/ni.2065
122. Radhakrishnan SK, Kamalakaran S. Pro-Apoptotic Role of NF-Kappab: Implications for Cancer Therapy. *Biochim Biophys Acta* (2006) 1766:53–62. doi: 10.1016/j.bbcan.2006.02.001
123. Zaragoza C, Saura M, Padalko EY, Lopez-Rivera E, Lizarbe TR, Lamas S, et al. Viral Protease Cleavage of Inhibitor of Kappabalpha Triggers Host Cell Apoptosis. *Proc Natl Acad Sci USA* (2006) 103:19051–6. doi: 10.1073/pnas.0606019103
124. Elde NC, Malik HS. The Evolutionary Conundrum of Pathogen Mimicry. *Nat Rev Microbiol* (2009) 7:787–97. doi: 10.1038/nrmicro2222
125. Chemes LB, de Prat-Gay G, Sanchez IE. Convergent Evolution and Mimicry of Protein Linear Motifs in Host-Pathogen Interactions. *Curr Opin Struct Biol* (2015) 32:91–101. doi: 10.1016/j.sbi.2015.03.004
126. Hagai T, Azia A, Babu MM, Andino R. Use of Host-Like Peptide Motifs in Viral Proteins Is a Prevalent Strategy in Host-Virus Interactions. *Cell Rep* (2014) 7:1729–39. doi: 10.1016/j.celrep.2014.04.052

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 15 November 2021
doi: 10.3389/fimmu.2021.744783

Sequence Diversity, Locus Structure, and Evolutionary History of the *SpTransformer* Genes in the Sea Urchin Genome

Megan A. Barela Hudgell and L. Courtney Smith *

Department of Biological Sciences, George Washington University, Washington, DC, United States

The generation of large immune gene families is often driven by evolutionary pressure exerted on host genomes by their pathogens, which has been described as the immunological arms race. The *SpTransformer* (*SpTrf*) gene family from the California purple sea urchin, *Strongylocentrotus purpuratus*, is upregulated upon immune challenge and encodes the SpTrf proteins that interact with pathogens during an immune response. Native SpTrf proteins bind both bacteria and yeast, and augment phagocytosis of a marine *Vibrio*, while a recombinant SpTrf protein (rSpTrf-E1) binds a subset of pathogens and a range of pathogen associated molecular patterns. In the sequenced sea urchin genome, there are four *SpTrf* gene clusters for a total of 17 genes. Here, we report an in-depth analysis of these genes to understand the sequence complexities of this family, its genomic structure, and to derive a putative evolutionary history for the formation of the gene clusters. We report a detailed characterization of gene structure including the intron type and UTRs with conserved transcriptional start sites, the start codon and multiple stop codons, and locations of polyadenylation signals. Phylogenetic and percent mismatch analyses of the genes and the intergenic regions allowed us to predict the last common ancestral *SpTrf* gene and a theoretical evolutionary history of the gene family. The appearance of the gene clusters from the theoretical ancestral gene may have been driven by multiple duplication and deletion events of regions containing *SpTrf* genes. Duplications and ectopic insertion events, indels, and point mutations in the exons likely resulted in the extant genes and family structure. This theoretical evolutionary history is consistent with the involvement of these genes in the arms race in responses to pathogens and suggests that the diversification of these genes and their encoded proteins have been selected for based on the survival benefits of pathogen binding and host protection.

Keywords: sea urchin, invertebrate immunity, *Strongylocentrotus purpuratus*, gene family evolution, large gene families

OPEN ACCESS

Edited by:
Bostjan Kobe,
The University of Queensland, Australia

Reviewed by:
Yuko Ota,
University of Maryland, Baltimore, United States
Anthony K. Redmond,
Trinity College Dublin, Ireland

***Correspondence:**
L. Courtney Smith
csmith@gwu.edu

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *20 July 2021*
Accepted: *12 October 2021*
Published: *15 November 2021*

Citation:
Barela Hudgell MA and Smith LC (2021) Sequence Diversity, Locus Structure, and Evolutionary History of the SpTransformer Genes in the Sea Urchin Genome. Front. Immunol. 12:744783. doi: 10.3389/fimmu.2021.744783

INTRODUCTION

Large, expanded immune gene families in echinoids were first identified in the genome sequence of the purple sea urchin, *Strongylocentrotus purpuratus* (1, 2). They include the Toll-like receptor (*TLR*) gene family that is composed of 253 members (3), the nucleotide oligomerization domain (*NOD*) and the NACHT leucine-rich repeat and PYD containing (*NALP*) gene families (1, 2), the cysteine rich scavenger receptor gene family (1, 4, 5), the *IL-17* cytokine genes (6), and the *SpTransformer* (*SpTrf*) genes of which 15 have been reported previously but whose copy number is likely to vary among individual sea urchins (7, 8). Most of the expanded gene families in the *S. purpuratus* genome encode, or are predicted to encode, proteins with immune function based on i) homologous genes in other species (9), ii) upregulation upon immune challenge (6, 10–13), or iii) patterns of expression and expected markers of gene sequence evolution (see below). The *SpTrf* gene family is upregulated swiftly upon immune challenge in sea urchin immune cells, called coelomocytes (10–13), although expression is restricted to the phagocyte subclass of coelomocytes in adults (13, 14) and the blastocoelar cells in larvae (15). As genes that encode proteins with immune function, native SpTrf proteins opsonize bacteria and augment phagocytosis (16), and one recombinant protein, rSpTrf-E1, binds to Gram negative bacteria, yeast, and several PAMPs (17, 18). The *SpTrf* genes consist of two exons with exon 2 composed 25-27 blocks of sequence called elements, which are present in a mosaic pattern and whose mosaicism makes up the 51 known element patterns (12, 19) that result in a wide range of sequences in exon 2 (12). The *SpTrf* genes also display allelic polymorphism (7) that increases the diversity of the family in individual animals and in the population (20). Allelic polymorphisms impart important diversity in small immune gene families such as those associated with allorecognition including the major histocompatibility complex (MHC) locus in higher vertebrates (21) and the fusion/histocompatibility (Fu/HC) locus in tunicates [reviewed in (22)]. Allelic polymorphism is also observed in large gene families such as the disease resistance (*R*) genes in plants (23). Differences between alleles at specific loci contribute to variation in the immune genes that improves fitness of the host to block and/or survive pathogen infection.

Large gene families can be generated through several processes of genome diversification [reviewed in (24)] that include duplications of large genomic regions, single or tandem duplications that can include coding regions, duplications that result in ectopic insertions as have been reported for *R* genes in plants [reviewed in (25–27)], inversions, meiotic mispairing of clustered genes with similar sequence, unequal crossing over of both intergenic and intragenic regions, and gene conversion in which a sequence from one gene is copied into a nonallelic gene of similar sequence (28). These processes that produce large gene families are the outcome of, and are promoted by, genomic instability (9). These traits are observable in genes under pathogen pressure based on the hypothesis that they are beneficial for maintaining the diversity in immune gene families to optimize fitness in response to pressure from pathogen interactions. In keeping with genomic instability, each *SpTrf* gene is flanked by GA short tandem repeats (STRs) that often includes GAT STRs plus long streatches of GA STR islands that flank two of the clusters (7). Furthermore, there are six different types of imperfect repeats in exon 2, which make up the mosaic pattern within the coding region of these genes and was the basis of the repeat-based alignement (see below) (29). While repeats are common in the sea urchin genome, the placement of the STRs around the genes in this gene family is unique and have been proposed to promote *SpTrf* gene duplication or deletion (7, 30). STRs are known to be highly unstable based on mutation rates that are up to 10 times greater than non-STR genomic DNA, which leads to genomic instability (31–34) and is likely a factor of strand-slipage, unequal crossing over, and/or conversion (35, 36).

The process of maintaining duplicated and altered immune genes is thought to be a response to pathogen pressure followed by selection for improved host fitness. However, the pathogen also responds with counter measures selected to avoid or defeat these new or modified host immune genes and that provide the benefits of successful infection and survival [reviewed in (37, 38)]. Both the host and the pathogen exert fitness pressure in a co-evolutionary arms race, which is known as the Red Queen hypothesis (39). Like the race between Alice and the Red Queen in Luis Carroll's Through the Looking Glass (40) where the two run a long and hard race only to stay in the same place, infers that a host can survive pathogen pressure only by rapidly changing genes that influence susceptibility or resistance to pathogen infection [(41), reviewed in (42)]. The pressure imposed on the host and the pathogen often leads to genomic regions with large expansive gene families (9, 20, 23, 24). Characteristics of genes involved in an arms race typically show signatures of positive selection, gene multiplicity, elevated recombination rates, and sequence variation that appear as elevated polymorphism and increased species level diversity (37, 38, 43–45). These processes can lead to the generation of complex and highly variable gene families that have the potential benefit of a greater range of pathogen recognition [e.g., (20)]. Some of the more common examples are the human killer cell immunoglobulin-like receptor (*KIR*) genes (46), fibrinogen-related protein genes (*FREP*s) in snails (47), variable region-containing chitin-binding protein (*VCBP*) genes in marine protochordates [(48), reviewed in (49)], *R* genes in higher plants [reviewed in (23)], and NOD-like receptor genes (*NLR*s) in animals (50). This phenomenon of multigene families is also common in other types of receptors, most notably the G-protein coupled receptors (GPCRs), which are mounted on the surface of cells and detect diverse types of external stimuli. These include chemoreceptors (51) such as olfactory receptors that are the largest multigene family in vertebrates (52, 53), some of the taste receptors (54, 55), and other GPCRs that identify large numbers of environmental molecules and trigger signaling pathways (51–55).

Here we present an in-depth bioinformatic and phylogenetic analysis of the sequence diversity of the *SpTrf* gene family that is encoded in the *S. purpuratus* genome sequence. We report an additional cluster of the *SpTrf* genes and describe details of both

the coding and flanking regions of the genes. The results enable a proposed theoretical evolutionary history for these genes originating from a last common ancestral (LCA) *SpTrf* gene, which subsequently underwent a number of tandem duplications, ectopic insertions, inversions, and intergenic indels and point mutations to generate the extant clustered genes in the genome sequence. While the genes identified in the sea urchin genome sequence are limited to a single animal, the analysis of these genes can be used as a basis for further work to understand genomic instability in the *SpTrf* gene loci in other *S. purpuratus* individuals that have different genotypes and different numbers of the *SpTrf* genes. These initial results suggest that genomic instability may be a key mechanism to promote diversification of immune gene families in echinoids that are locked in arms races with their pathogens.

MATERIALS AND METHODS

Bacterial Artificial Chromosome Clones

The sea urchin BAC library that was used to generate the genome assembly was the source of the BACs used in this analysis (56). They included BAC 10B1 (GenBank accession number KU668451; 157472 nt), BAC 10K9 (GenBank accession number KU668453; 144627 nt), BAC 10M18 (GenBank accession number KU668450; 74402 nt), and BAC 3104P4 (GenBank accession number KU668454; 118584 nt) (7). The identification of *SpTrf* genes in BAC insert sequences, plus the characterization of element patterns, untranslated regions, introns, and open reading frames were carried out according to Oren et al. (7). GenePalette[1], a universal software tool for genome sequence visualization and analysis (57), was used to identify the locations of individual *SpTrf* genes within each BAC insert sequence based on the locations of the *SpTrf* primer sequences (R1, F2, F5, F6, R9; see **Supplementary Table S1**). The 5′ and 3′ ends of the genes were identified using conserved primer sequences [5′UTR and 3′UTR; **Supplementary Table S1**; and see (7)]. *SpTrf* genes identified in the BAC insert sequences were added to a pre-aligned set of 121 unique *SpTrf* genes and 689 cDNAs with known and identified element patterns as previously reported (11, 19). The deduced amino acid sequences were aligned by hand in BioEdit (ver 7.2.5) (58) to identify the exons and to produce a repeat-based alignment and a cDNA-based alignment as previously reported (12, 19). The exons and the elements were identified and labeled for each *SpTrf* gene and verified based on previously reported genes. Introns were identified for each gene using the repeat-based alignment in BioEdit in which the 3′ end of exon 1 was used to identify the conserved GT splice signal that was approximately 54 nucleotides (nt) from the start codon and the conserved AG splice signal that was located approximately 550 nt from the start codon [see (59)]. Introns were removed from genes to determine whether all genes had open reading frames using NBCI Open Reading Frame Finder[2].

The cDNA sequence of *Sp0273* [GenBank accession number CK828488.1 (10)] was used to identify the 5′UTR and the TATA box, and the *Sp0065* cDNA sequence [GenBank accession number CK828780 (10)] was used to identify the poly adenylation sites. GenePalette was used to identify additional polyadenylation sites in the 3′UTR region of the genes. The 5′UTR and 3′UTR sequences were identified in GenePalette and verified from partial cDNA sequence data (10).

PRANK Analysis

Computational alignments of the deduced SpTrf proteins were done using GUIDANCE2[3] (60–62). Codons were aligned using the multiple sequence alignment (MSA) algorithm in PRANK[4] (63), an alignment-based software program that processes and identifies the placements of indels. The program was set to trust insertions (F+). Bootstrap guide-trees of 100 iterations were generated, which were further used to calculate 400 alternative alignments using PRANK with F+ before the GUIDANCE2 score was calculated to display whether the alignment was optimal. GUIDANCE scores were analyzed, however because the majority of sequences, columns, and amino acids with low GUIDANCE scores (here defined as >0.8) were associated with the outgroup sequences, the alignments were left unmodified prior to further analysis (data not shown). The deduced translated sequences for the 5′ and 3′ flanking regions (FRs), introns, and intergenic regions (IGRs) were also aligned with GUIDANCE2 using PRANK with the same parameters. The edges of the FRs included the 5′ and 3′ UTRs and extended to the location of the flanking GA STRs.

Sequence Similarity and Percent Mismatch Analysis of *SpTrf* Genes

Sequence similarity among genes with the same or relatively similar element patterns was evaluated with three approaches. i) Percent coverage and percent identity values were established using the basic local alignment search tool (BLAST[5]). ii) Sequence identity matrices were calculated in BioEdit (ver 7.0.5.3) based on the alignment of the deduced proteins. The number of identical residues were calculated while treating gaps as a fifth state to evaluate the similarities among the deduced proteins. iii) A pairwise distance matrix was calculated with Molecular Evolutionary Genetics Analysis [MEGA7[6], ver 7.0 for larger datasets (64)] using the codon alignment generated in PRANK with preset parameters. All three analyses were run on six regions of the genes that included the 5′FR, exon 1, the intron, exon 2, and the 3′FR, in addition to the IGRs. Percent mismatches were calculated according to the equation [pairwise distance/Ln^2], in which the results were generated from the average pairwise distance matrix data for each gene compared to every other gene, divided by Ln^2, in which the superscript 2 indicates the number of sequences that were compared. A graphical representation of these values was generated using Excel (Microsoft).

[1] http://www.genepalette.org
[2] http://www.ncbi.nlm.nih.gov/orffinder
[3] http://guidance.tau.ac.il/ver2/
[4] http://wasabiapp.org/software/prank/
[5] https://blast.ncbi.nlm.nih.gov/Blast.cgi
[6] http://www.megasoftware.net

Phylogenetic Trees

MEGA7 was used to generate phylogenetic trees from the PRANK alignments of the 5′FR, exon 2, the intron, and the 3′FR (~400 nt for the FRs) to determine the evolutionary relatedness among the sequences. Representative *Trf* sequences were selected from the sea urchin, *Heliocidaris erythrogramma* [*HeTrf;* GenBank accession numbers JQ780171-JQ780321; 29 genes; 39 introns (65)], which were used as the outgroup for phylogenetic analyses of both exon 2 and the intron of the *SpTrf* genes. Additional *SpTrf* genes (121 genes, 22 introns) were acquired from Buckley and Smith (19) and used to generate intron and expanded exon 2 trees (**Supplementary Figures S1, S2**). A single *Trf* gene identified from the *Lytechinus variegatus* genome sequence[7] [*LvTrf; Lv=185/333B3d;* NCBI Accession GCA_000239495.1; Scaffold 232, 80220 to 85000 (66)] was acquired and included 2500 nt on each side of the gene. The 5′FR and 3′FR (~400 nt) of the *LvTrf* gene were used as the outgroup in the FR alignment of the *SpTrf* genes. Maximum likelihood, neighbor joining, and maximum parsimony with pre-set parameters were used to generate phylogenetic trees. Bootstrap iterations were set to 500 and nodes with a bootstrap value of <50 were collapsed. All trees resulted in similar structure (**Supplementary Figures S3, S4**).

Dot Plots

Dot plots were generated using YASS[8] genomic similarity search tool to identify repeats and regions of similarity among genes within and among the four gene clusters. The e-value threshold ranged from 10,000 to e^{-30} as was optimal for different analyses (67). Dot plots from YASS were used to evaluate the sequence variations between allelic BAC 10B1 and BAC 10K9.

Analysis of Intergenic Regions Among Non-Duplicated Genes

Dot plots were generated using the e-value threshold set to e^{-20}. The IGRs between different genes were compared, which consisted of 3 kb flanking the 5′ and 3′ ends of the allelic *A2* and *A2*a (*A2*/a) genes, the entire 6.9 kb IGRs between the *E2*/a and the *E2*b/*01* genes, 3 kb to the 3′ side of the *E2*b/*01* genes, IGRs between the *D1*b/e and *E2*/a genes, and the IGRs between the *D1*h/f genes and the GAT STRs.

Verification of Allelic BACs

Sequence variations between allelic BAC 10M18 and BAC 3104P4 were analyzed using GenePalette in which GA and GAT STRs were mapped using the sequences GAGAGA and GATGATGAT, respectively, while allowing for a single mismatch. Primers GA1F-GA3F and GA1R-GA3R (**Supplementary Table S1**) were designed to amplify large regions of STRs to evaluate variations in STR lengths using PrimeSTAR GLX high fidelity DNA Polymerase (Takara) to ensure as little polymerase slippage as possible. The PrimeSTAR GLX protocol was 1X PrimeSTAR GXL buffer, 200 µM of each dNTP, 10-15 pmol of each primer, 10 ng BAC DNA, 0.5 U of PrimeSTAR polymerase in a final volume of 20 µL. The PCR program was 95°C for 5 min, followed by 25 cycles of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 4 min with a final extension of 72°C for 7 min and a hold at 4°C. Amplicons were separated on a 0.75% agarose gel with Tris-acetate-EDTA buffer (TAE: 40 mM Tris-acetate pH 8.0, 1.0 mM EDTA).

Synonymous *vs*. Nonsynonymous Nucleotide Changes

Exon 2 of *SpTrf* genes with the same element pattern were compared to identify synonymous *vs*. nonsynonymous single nucleotide polymorphisms (SNPs). SNPs were catalogued by eye from the alignments and verified with Synonymous Non-synonymous Analysis Program (SNAP[9]; ver 2.1.1) (68) and Single-Likelihood Ancestor Counting (SLAC) (69) in datamonkey[10] (70–72). The dN/dS value for each gene was calculated based on the Jukes-Cantor corrections (73–75). SNAP was used as an alternative method to evaluate the dN/dS and the number of synonymous *vs*. nonsynonymous substitutions because SNAP was capable of calculating dN/dS values between two genes rather than a group of genes required by SLAC. SLAC was used to confirm results for the seven *D1* genes. Purifying selection was defined as dN/dS of < 1, whereas diversifying selection was defined as > 1. Because the *D1*f and *D1*h gene sequences were identical, they were combined and noted as *D1*f/h for comparisons to the other *D1* genes.

RESULTS

Pairs of BAC Inserts Are Likely Allelic Rather Than Clones That Cover Identical Genomic Regions

Clusters 3 and 4 Are Allelic

Previous work defined three *SpTrf* gene clusters from BAC insert sequence analysis, of which Clusters 1 and 2 were defined as allelic based on the nearly identical sequences of the genomic regions that flank these two *SpTrf* gene clusters (7). The allelic region for Cluster 3 was not reported because the two tightly linked *SpTrf* genes in Cluster 3 were present in both BAC 10M18 and BAC 3104P4 and were reported as replicates of the same region of the genome (7). This assumption was feasible given the 25X coverage of the BAC library (56). To verify whether these two BAC inserts were identical or allelic, the sequences were re-evaluated by dot plot comparison followed by verification using PCR amplification of the gene clusters and three large flanking STRs. The two *SpTrf* genes on BACs 10M18 and 3104P4 encompassed about 10 kb, which was verified by PCR, and had identical sequences based on comparisons using GenePallete (**Figure 1A**, purple angle arrows; **Figure 1B**). Three large GA STR islands were associated with the gene clusters based on GenePallete (**Figure 1A** and **Table 1**). PCR amplicons of the STRs indicated different sizes for the STR2 amplicon for the two BACs (**Figure 1C**). This suggested that BAC 10M18 (Cluster 3) and BAC 3104P4 (Cluster 4) were likely allelic and were identified as Locus 2 for the *SpTrf* gene family in the sea urchin genome.

[7] http://whis.caltech.edu/Echinobase/LvAbout
[8] http://bioinfo.lifl.fr/yass/index.php
[9] https://www.hiv.lanl.gov/content/sequence/SNAP/SNAP.html
[10] http://classic.datamonkey.org/dataupload.php

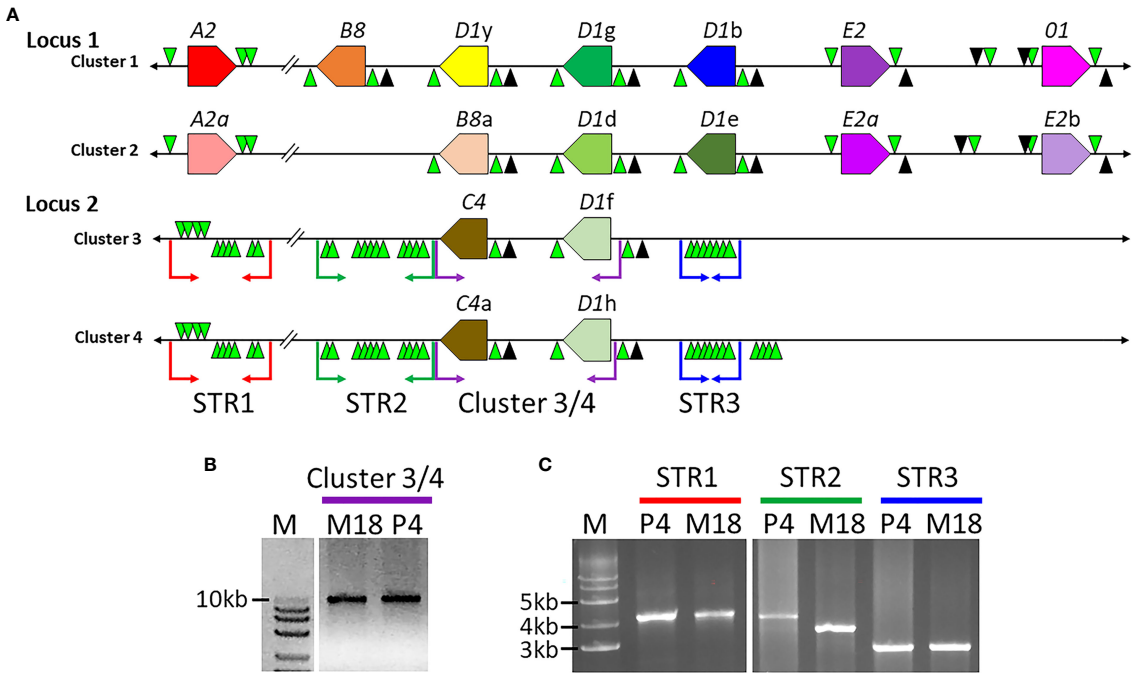


FIGURE 1 | The structure of the *SpTrf* gene loci. **(A)** A representative map of the *SpTrf* loci. Locus 1 has allelic regions with unequal numbers of genes. Although Clusters 3 and 4 in Locus 2 appear identical, the different sizes of the flanking STR islands indicate that these two clusters are allelic. The colored polygons indicate the *SpTrf* genes located in the clusters with the pointed end of the polygon indicating the transcription direction. GA STRs (green triangles) and GAT STRs (black triangles) flank each gene and large GA STR islands flank Clusters 3 and 4 in Locus 2. The black horizontal line indicates the DNA extending from the 5′ and 3′ ends of the clusters. The colored angle-arrows in Locus 2 indicate the regions amplified by PCR and correspond to the colored bars over lanes in the DNA gels shown in **(B, C)**. **(B)** Clusters 3 and 4 in Locus 2 are the same size. The BAC templates for PCR are indicated above the lanes as M18 (BAC 10M18; Cluster 3) or P4 (BAC 3104P4; Cluster 4). Amplicons of Clusters 3 (BAC 10M18) and 4 (BAC 3104P4) indicate identical size. **(C)** Clusters 3 and 4 in Locus 2 have varying sizes of large GA STR islands. PCR was carried out for the P4 and M18 BAC clones to amplify the GA STR islands. M indicates the all-purpose Hi-Lo DNA marker (BioNexus), and sizes of the relevant bands are indicated to the left in **(B, C)**.

Cluster 1 and Cluster 2 Have an Intergenic Region of Dissimilarity

Previous analysis of Clusters 1 and 2 of Locus 1 (**Figure 1A**) report different numbers of *SpTrf* genes, of which some genes are unique to a particular cluster based on different element patterns (7). Cluster 1 (BAC 10B1) has seven *SpTrf* genes while Cluster 2 (BAC 10K9) has six (7, 30). However, the flanking regions of these two allelic regions show approximately 99% sequence identity, which was the basis for reporting their allelic relationship rather than as two different loci (7). Dot plots of the BAC inserts for Locus 1 verified their allelic status, but also identified regions with significant sequence variations in the intergenic regions (IGRs) and in the flanking regions that surround the clusters (**Figure 2**). Although most of the region flanking the clusters generally aligned, Cluster 2 had a large deletion (**Figure 2A**, blue bar), in agreement with the previous report (7). The IGRs between the *A2* and *B8* genes in Clusters 1 and 2 were different in size and sequence that remained evident after increasing e-value threshold for the dot plot (67) (**Figures 2A, B**). The sequence identity of this region of dissimilarity in the two clusters was 42% to 48.1% based on BLAST and BioEdit analysis, respectively. When these sequences were used to search for other matches in the sea urchin genome in the NCBI database only poor matches were identified with percent mismatches of ~45.7% identity (based on results using MEGA7). A more detailed analysis of the IGRs between the *A2* an *B8* genes showed that there were two discrete areas of variation (**Figure 2C**). The first was 3.8 kb that was only present in Cluster 2, which was flanked on both sides by regions of high similarity with Cluster 1 (**Figures 2A, C**; red stripes). At the 5′ end of the IGR near the *A2* genes in each cluster was 1.1 kb of non-coding sequence that included the GA STRs. In the 3′ direction was 718 nt in Cluster 2 that matched with 96% identity to 730 nt in Cluster 1 (**Figure 2C**, yellow). The second region of dissimilarity (<40% identity) was 6.1 kb (Cluster 2) and

TABLE 1 | The second STR island in Locus 2 alleles are different lengths[1].

BAC	Cluster	Size (nt)[2]		
		STR1	STR2	STR3
10M18 (M18)[3]	3	4293	3869	2633
3104P4 (P4)	4	4314	3998	2635

[1]*These are results from sequence comparisons using GenePallet.*
[2]*The locations of the STR islands are shown in* ***Figure 1A****.*
[3]*Abbreviations for BAC numbers in parentheses correlate with* ***Figure 1C****.*

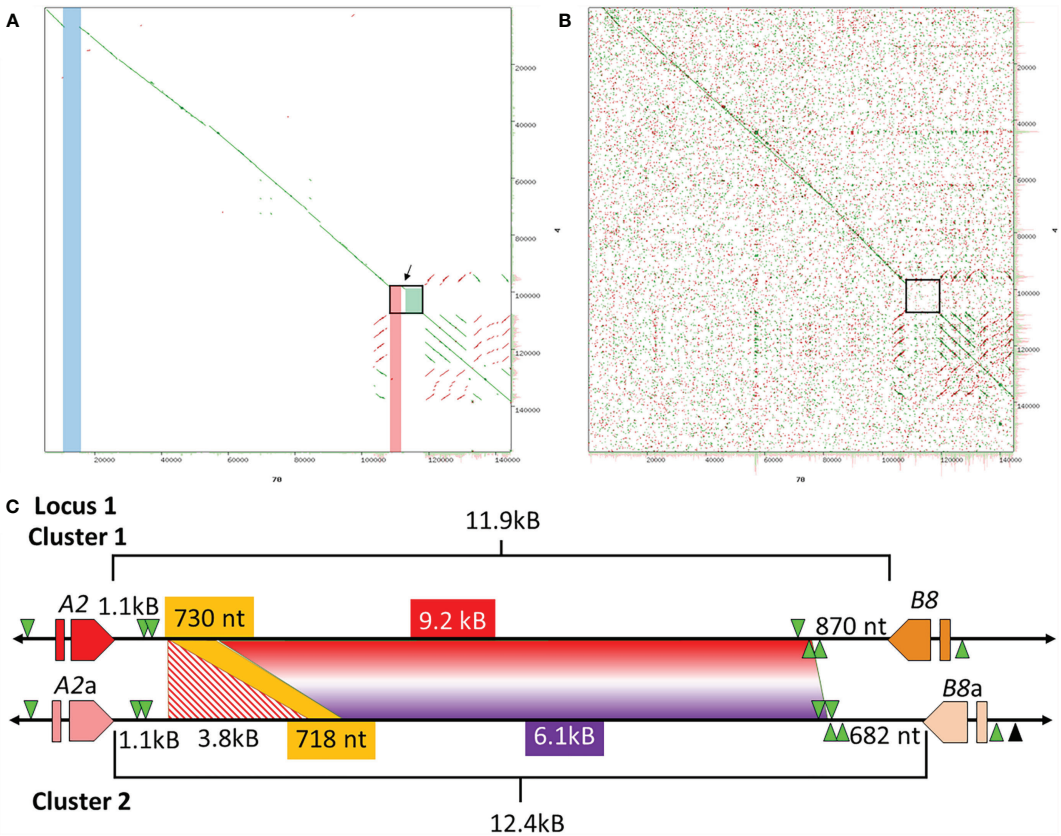


FIGURE 2 | The IGRs of Clusters 1 and 2 in Locus 1 include non-matching sequences. **(A, B)** Dot plots show the comparison between Cluster 1 and Cluster 2. The dot plot in **(A)** employed a preset e-value threshold of 10 whereas the dot plot in **(B)** employed an e-value threshold of 10,000. The central diagonal in the dot plots indicate the mostly identical sequence of the allelic regions, while lines offset from the central diagonal indicate repeats that are highly similar in either a tandem (green) or inverted (red) orientation. Highlighted, colored vertical bars in **(A)** indicate the locations of mismatched sequences between the two clusters. The blue and red bars show the locations of sequences in Cluster 2 that are absent from Cluster 1 and the green box indicates a region of complete dissimilarity. The arrow between the red bar and green box in **(A)** indicates a region of similarity that is located between the two regions of dissimilarity. The black boxes in **(A, B)** are expanded in **(C)** to show details. YASS[8] was used to generate dot plots with standard parameters (scoring matrix = +5, -4, -3 -4: composition bias correction: gap costs = -16, -4: X-drop threshold = 30). **(C)** The IGRs located between the *A2*/a and *B8*/a genes in Clusters 1 and 2 are a mixture of similar and dissimilar sequences. The red and orange polygons indicate the *SpTrf* genes, *A2*/a and *B8*/a genes, that flank these IGRs. GA STRs (green triangles) surround each gene. The horizontal black line indicates the DNA that extends from the 5′ and 3′ ends of each gene. The lengths of the IGRs between the *A2*/a and *B8*/a genes are indicated by upper and lower brackets. The sizes of the areas within the IGRs are indicated by colors that are coordinated when similar. The red and white striped region is a sequence that is only present in Cluster 2 and corresponds to the red bar in **(A)**. The yellow region is a short area of similarity, and the area of complete dissimilarity is shown as a polygon of a red and purple gradient. This figure in not drawn to scale.

9.2 kb (Cluster 1) and extended from the 718/730 nt region of similarity to the *B8* genes (**Figure 2C**). While Clusters 1 and 2 of Locus 1 show similarities within flanking regions, there were also large regions of dissimilarity outside and within the cluster, the largest variation being the IGRs between the *A2* and *B8* genes.

Stop Codons and Untranslated Regions in the *SpTrf* Genes

The locations of the TATA box and polyadenylation site were reported previously for six of the *SpTrf* genes in Cluster 1 (30) except for the *01* gene, which was identified as part of Cluster 1 in a subsequent report (7). Those initial reports plus a set of partial *SpTrf* cDNA sequences (10) were used to identify or verify the transcriptional start and stop sites and the sizes of the untranslated regions for all of the *SpTrf* genes in the BAC insert sequences. Results showed that the 5′UTR ranged in size from 146 nt to 149 nt for 16 of 17 genes with the TATA box positioned 101 nt to 111 nt from the start codon within the 5′UTR, in agreement with the TATA box positioning described in Miller et al. (30) (**Supplementary Figure S5**). However, the TATA box for the *D1*g, which was reported to have a point mutation of TATACA was not verified. Rather, the *D1*g TATA box had a TATAAA sequence that was similar to the other genes, with the exception of *D1*d with a sequence of TATATA. No other conserved TATA box sites were identified within the proximity of the 5′ end of the UTR (the next nearest was distant by 1.3 kb). A putative initiator (Inr) (76, 77) was identified in all genes and located 27 nt to the 3′ side of the TATA box with the degenerate

sequence of T(CA)A(+1)GTT in which the +1 A was conserved (**Figure 3A** and **Supplementary Figure S5**). This sequence is similar to the Inr sequence in *Drosophila* genes (76) and is considered a core promotor similar to the TATA box that can enhance binding affinity to a promotor element for either RNA polymerase or a transcription factor and, in some cases, can direct transcription without a functional TATA box (78).

3′UTRs are defined by the location of the stop codon and the polyadenylation sequence. Three stop codons have been reported for the *SpTrf* genes (19) and cDNAs (11, 12) and are present in the last element of exon 2 (**Figure 3A**, indicated as a, b, and c). Analysis of the genes in Clusters 3 and 4 identified a fourth stop codon in the *D1*f and *D1*h genes, in which a SNP at nucleotide 955 changed a tryptophan codon to a stop (**Figure 3A**, identified as d;

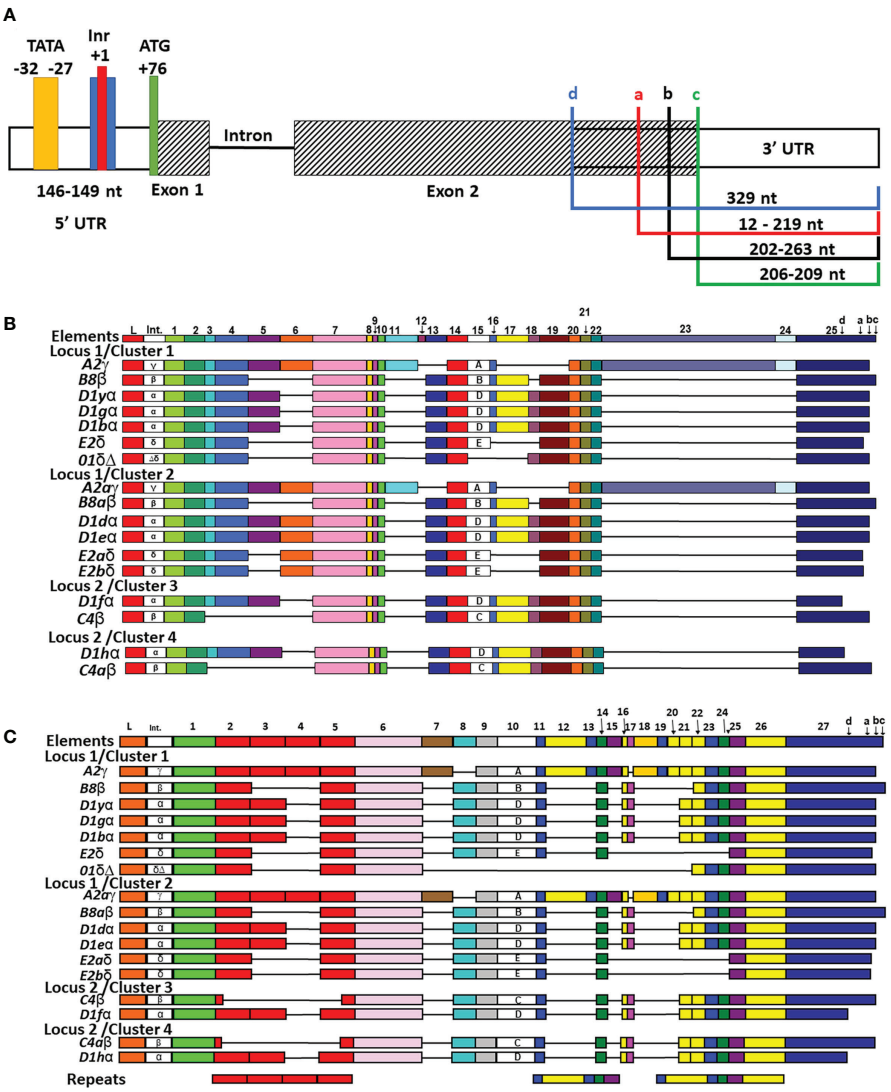


FIGURE 3 | All *SpTrf* genes are in frame, have identifiable TATA, Inr sites, one or more stop codons, and most can be aligned with the previously established repeat-based and cDNA alignments. **(A)** A representative map of the genes shows the 5′ UTR, exons, intron, and 3′ UTR. The 5′ and 3′ UTRs are indicated by white rectangles, the two exons are indicated as striped rectangles, and the intron is indicated by a solid black line. The range in lengths of the 5′ UTR among genes is indicated. The four colored boxes in 5′ UTR indicate putative 5′ regulatory elements and their locations + or – of the conserved +1 A of the start transcription site (red). The TATA box (yellow), the Inr (blue), and the ATG translation start (green) are indicated. The 3′ UTR is variable in length among genes and is indicated by colored brackets showing the four possible locations of stop codons, which are labeled in lowercase 'a'-'d'. **(B)** The cDNA alignment of genes from the four clusters. The manual alignment was done in BioEdit by adding the genes in the clusters to a pre-aligned set of cDNAs and genes according to previous publications (12, 19). All possible elements are numbered at the top and the four possible stop codons are indicated in element 26. The leader (L), the intron (Int), elements (colored rectangles), and gaps (horizontal lines) are indicated for each gene. Intron type and subtype of element 15 are labeled within each respective rectangle. **(C)** The repeat-based alignment of the genes from the three clusters. The manual alignment was done as in **(B)** according to Buckley and Smith (19). All possible elements are numbered at the top and the four stop codons are indicated in element 27. The leader, intron, intron type, elements, subtype of element 10, and gaps are indicated as in **(B)**. The six types of repeats in the gene sequence are indicated by rectangles of identical color at the bottom.

Supplementary Figure S6). This increased the size of the 3′ UTR by 116 nt and decreased the length of exon 2 shortening the protein by 38 amino acids (aa) relative to the other *D1* genes. Two types of polyadenylation sequences were identified, AATAAA and ATTAAA, of which most genes [13 of 17] had both (**Table 2**). Overall, the 3′UTR varied in length from 195 nt to 357 nt primarily based on the positions of the stop codons among the genes (**Figure 3A**). All of the *SpTrf* genes appeared to be functional with short UTRs, although the *D1*d gene in Cluster 2 had different sequences for transcription initiation and for the location of transcript trimming prior to polyadenylation. These results suggested that these genes have the minimal requirements for expression, although the regulatory regions for these genes have not been evaluated.

Exon 1 Is Conserved Whereas Exon 2 Is Highly Variable Among the Genes

Exon 1 in all *SpTrf* genes are either 51 or 54 nt in length and encode a conserved signal sequence of 18 or 19 aa (12, 19, 30). The difference is the presence or absence of the second codon for glutamic acid (**Supplementary Figure S7**), which has been reported previously (11, 12, 19). Eight additional variations in exon 1 were identified among the 17 genes in the four clusters, all of which were nonsynonymous polymorphisms that maintained the non-polar characteristic of the encoded leader. Although the function of the leader has not been tested formally, it is predicted to have characteristic hydrophobic and alpha helical structure (12, 18), which is consistent with secretion of the SpTrf proteins (16) and/or their localization to the plasma membrane (13). Overall, exon 1 of the genes in the four clusters was highly conserved both in sequence and hydrophobicity and did not show extensive sequence variation.

Manual alignments of exon 2 have been used previously because of the large gaps required to optimize the alignments, and these efforts have generated two possible alignment outcomes denoted as the cDNA alignment and the repeat-based alignment that are feasible because of the variety of repeats in exon 2 (**Figures 3B, C**) (11, 12, 19). To evaluate exon 2 for the 17 genes in the four clusters, the sequences were added to previously published alignments to understand how the genes in the clusters were related to the other *SpTrf* sequences including their element patterns (**Figures 3B, C** and **Supplementary Figure S6**) and intron types (**Figure 4A** and **Supplementary Figure S8**) (7, 12, 19, 30).

The *A2* Genes

The *A2* and *A2*a genes in Locus 1, as reported previously (7, 30), have 25 of 27 elements according to the repeat-based alignment and are only missing elements 8 and 17 that encode histidine rich regions of the proteins (**Figure 3C**). Sequence comparison of exon 2 for the *A2* genes showed that they were not identical (93% identical, 100% coverage; **Table 3**) because of an indel of 15 nt starting at nucleotide 950 of the A2 gene alignment (**Supplementary Figure S9**). Additional differences in the coding region for the *A2* genes were due to 12 SNPs, of which 10 changed the amino acid and seven changed either the charge or pI of the amino acid (**Supplementary Table S2**). The percent identity of the full-length *A2* genes, including the intron was 88%, in agreement with the minimum percent identities among all genes (19). Exon 2 had a 98% identity, and exon 1 had a 95% identity between the *A2* and *A2*a genes indicating that the majority of the sequence differences were in the introns (**Supplementary Tables S3, S4**). The *A2* and *A2*a genes had moderately dissimilar (88% identity) γ type introns that were positioned in different sister subclades for γ introns in the phylogenetic tree of introns (**Figure 4A**). Differences in the introns were due to one or two nt indels in addition to a region of significant variation from nt 354 to the 3′ end of the intron (**Supplementary Figure S9**). The sequence variation among the *A2* γ introns was greater than the introns in most other *SpTrf* genes with the same element pattern and same intron

TABLE 2 | The 3′and 5′UTRs for the *SpTrf* genes are short and all have conserved and identifiable transcription elements.

Gene	Full length transcript (nt)	5'UTR (nt)	Exon 1 (nt)	Intron (nt)	Exon 2 (nt)	3'UTR (nt)	Poly-A site[1]	Poly-A site variant[2]
A2	1893	148	54	413	1418	273	257	12
B8	1422	149	54	415	1023	196	206	N/A[3]
D1y	1503	148	51	411	1080	224	202	251
D1g	1511	148	51	412	1080	232	N/A	259
D1b	1511	148	51	411	1080	232	N/A	259
E2	1313	147	54	408	822	290	219	267
01	1197	148	51	348	771	227	208	167
A2a	1898	147	54	407	1431	266	248	13
B8a	1421	149	54	413	1023	195	209	N/A
D1d	1522	148	51	412	1080	243	263	214
D1e	1522	148	51	410	1080	243	215	264
E2a	1260	147	54	406	822	237	218	266
E2b	1237	148	51	407	810	228	208	167
C4	1725	149	54	383	909	230	205	N/A
D1f	1931	146	54	411	963	357	329	378
C4a	1725	149	54	383	909	230	205	N/A
D1h	1931	146	54	411	963	357	329	378

[1]*Nucleotide position of Poly-A site sequence (AATAAA) relative to the stop codon.*
[2]*Nucleotide position of Poly-A site sequence (ATTAAA) relative to the stop codon.*
[3]*N/A, not applicable, no Poly-A site or Poly-A site variant were found.*

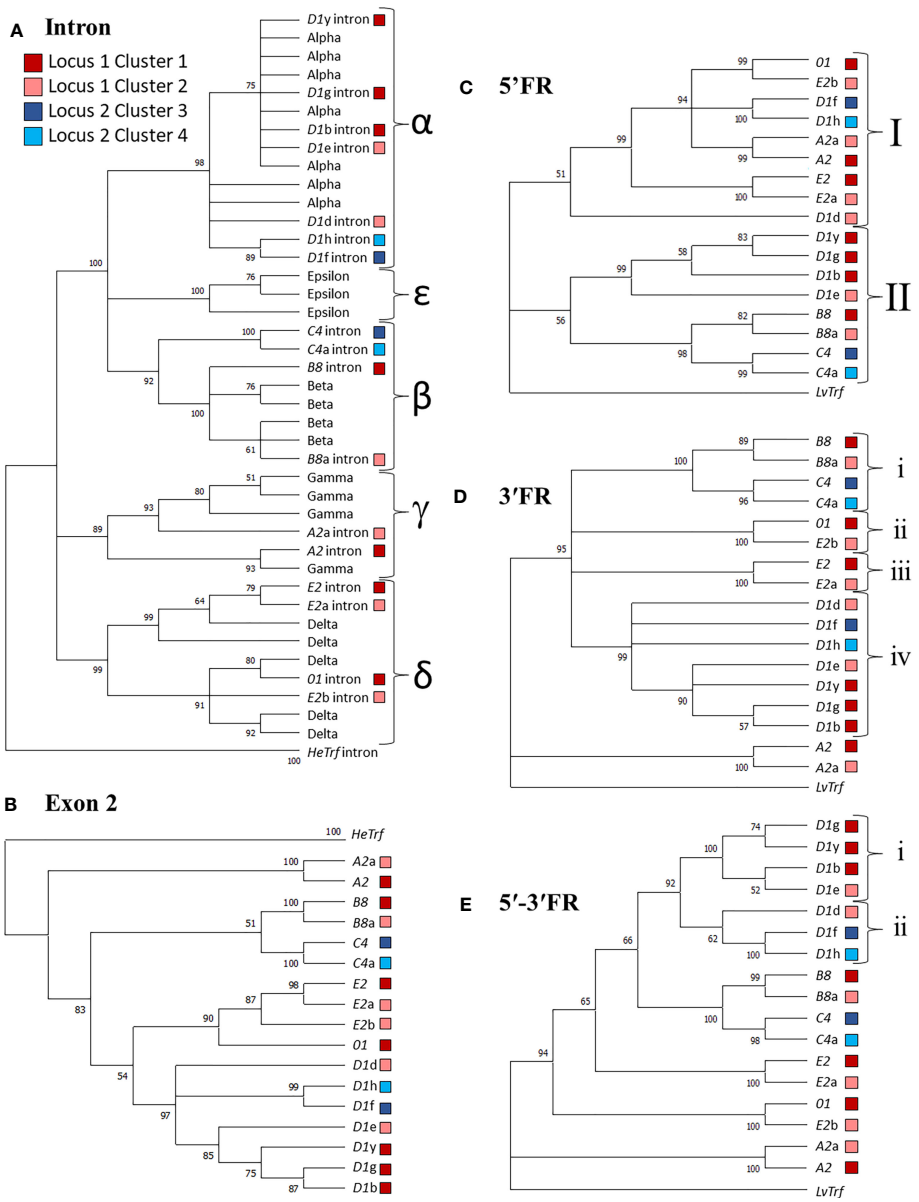


FIGURE 4 | Sequence similarities among the *SpTrf* genes and their putative evolutionary relationships are revealed by similar structures of maximum likelihood trees. Alignments were performed with PRANK, and Phylogenetic analysis was completed in MEGA7. Phylogenetic trees were generated using three approaches: neighbor joining, maximum parsimony (see **Supplementary Figures S1–S4**), and maximum likelihood (shown), all of which resulted in similar tree structure. Colored boxes shown in the legend indicate the cluster in which the gene is located. Bootstrap values from 500 iterations are indicated for each tree. **(A)** The intron tree. The intron types (indicated by α-ϵ labels for separate clades) for each gene was identified using a previously published alignment of introns (19) with the introns from *HeTrf* genes defined as the outgroup. **(B)** The exon 2 tree. Exon 2 from each gene was aligned and the exon 2 sequences from *HeTrf* genes were defined as the outgroup. **(C)** The 5′FR tree. The 5′FR for each gene was selected using GenePallete and corresponded to 400 nt upstream of the start codon. The 5′FR of the *LvTrf* gene was used as the outgroup. **(D)** The 3′FR tree. The 3′FR for each gene was selected using GenePallete and corresponded to 400 nt downstream of the stop codon. The 3′FR of the *LvTrf* gene was used as the outgroup. **(E)** The concatenated 5′-3′FR tree. The 5′FRs and 3′FRs used in **(C, D)** were aligned and then concatenated prior to phylogenetic analysis. The concatenated 5′-3′ FR of the *LvTrf* gene was used as the outgroup.

type. This is not consistent with introns from genes that shared the same intron type, which tended to have highly similar introns (93% to 100% identical; **Supplementary Table S3**). Overall, the *A2* genes were highly similar but not identical, with most of their sequence variations located in the intron.

The *B8* Genes

SpTrf gene sequence analysis from 10 sea urchins indicates that the genes with a *B* element pattern are likely common in the population (19) and that the gene copy number estimate for *B* genes in individual animals ranges from one to six (7). Two *B8*

TABLE 3 | Percent identity and coverage of the *SpTrf* genes of the same element pattern are highly similar*.

Genes compared		Coverage	Identity
A2	*A2*a	100	98
***B8*a**	*B8*	100	99
C4	*C4*a	100	100
***D1*f**	*D1*d	99	99
***D1*f**	*D1*e	99	98
***D1*f**	*D1*y	99	99
***D1*f**	*D1*g	99	98
***D1*f**	*D1*b	99	98
***D1*f**	*D1*h	100	100
***D1*h**	*D1*d	99	99
***D1*h**	*D1*e	99	98
***D1*h**	*D1*y	99	99
***D1*h**	*D1*g	99	98
***D1*h**	*D1*b	99	98
***D1*d**	*D1*e	100	98
***D1*d**	*D1*y	100	99
***D1*d**	*D1*g	100	98
***D1*d**	*D1*b	100	98
***D1*e**	*D1*y	100	99
***D1*e**	*D1*g	100	99
***D1*e**	*D1*b	100	99
***D1*y**	*D1*g	100	99
***D1*y**	*D1*b	100	99
***D1*g**	*D1*b	100	99
E2	*E2*a	100	99
E2	*E2*b	97	96
***E2*a**	*E2*b	97	96
01	*E2*	82	96
01	*E2*a	82	96
01	*E2*b	85	97

*These data were generated using NCBI BLAST.

genes are present in the BAC insert sequences; *B8* in Cluster 1 (30) and *B8*a in Cluster 2 (7) based on the elements defined by the repeat-based and cDNA-based alignments (**Figures 3B, C** and **Supplementary Figure S6**). The *B8* genes did not show new variation in their element pattern relative to previous reports of other *B8* cDNAs and genes (11, 12, 19). The full-length gene sequences for *B8* and *B8*a were 99% identical including the β introns (**Figure 4A**, **Table 3** and **Supplementary Tables S3, S4**), in agreement with previous results (30). Differences between the sequences showed 10 SNPs of which five were located in the exons and four altered the charge or pI of the encoded amino acid (**Supplementary Table S2**). A single nt indel was located in the intron at position 71 (**Supplementary Figure S10**), and both *B8* genes had a single stop codon in the 'c' position in element 27 (**Figure 3A**). Overall, the *B8* genes in Cluster 1 and 2 were highly similar with a few differences in SNPs and were consistent with *B8* sequences previously reported.

The *C4* Genes

Two *C* genes were identified in Locus 2, *C4* in Cluster 3 (7) and *C4*a in Cluster 4 (**Figure 1A**). The *C* genes are not common in genomic DNA from individual sea urchins (19), however an estimate of gene copy number suggests 1 to >5 in 9 of 10 sea urchins (7). The *C4* genes had identical sequences (**Table 3**, **Supplementary Tables S3, S4** and **Supplementary Figure S11**) and matched with 97% identity over a 92% coverage to *Sp0376* cDNA [GenBank accession number DQ183179.1 (12)], which is the only *C4* sequence in the *SpTrf* sequence database (**Supplementary Figure S12**). The *C4* genes contained a distinguishing deletion in the type 1 repeat region of exon 2, which made them distinct from genes with the *C2*, *C3*, and *C5* element patterns (12, 19). This deletion brought together the first 10 nt of element 2 and the last 35 nt of element 5 and maintained the reading frame (**Figure 3C**). The stop codons for both *C4* genes were in the 'b' position in element 27 (**Figure 3A**). Although all *C* genes previously sequenced from genomic DNA had α introns (19), the *C4* and *C4*a genes on Clusters 3 and 4 had β introns based on the phylogenetic intron analysis (**Figure 4A**). The intron alignment and the intron phylogenetic tree indicated that the *C4* intron sequence shared similarity with the *B8* intron from the 5′ end of the intron to nt 285 and from nt 373 to 450 at the 3′ end. However, the central region, from nt 286 to 372, shared similarity with the γ intron in the *A2*a gene, although it had an indel of 32 nt (**Supplementary Figure S13**). The fragments of shared intron sequence between the *C4* genes and a gene with a different element pattern was unique among the *SpTrf* genes in the genomic clusters. The *C4* genes in Locus 2 of the *SpTrf* gene family were distinct from the other *C* genes based on both exon sequence and intron type.

The *D1* Genes

There are three *D1* genes in Cluster 1 known as *D1*-yellow (*D1*y), *D1*-green (*D1*g), and *D1*-blue (*D1*b) (30), two *D1* genes in Cluster 2 known as *D1*d and *D1*e, and in Cluster 3 as *D1*f (7). Here, we report the *D1*h gene in Cluster 4 (**Figure 1A** and **Supplementary Figure S14**). All of the *D1* genes were highly similar (95% to 100% identical) in both the coding regions and the α introns with most of the differences identified as SNPs throughout the sequences (**Figure 4A**, **Table 3** and **Supplementary Figure S14** and **Supplementary Tables S3, S4**). The *D1*f and *D1*h genes had more SNPs compared to the *D1* genes in Locus 1, including a stop codon at nt 955 in position 'd' (**Figure 3A** and **Supplementary Figure S6**). The *D1* genes made up the largest group of genes in the *SpTrf* gene family (19) and were the most common element pattern in the sequenced BAC inserts as reported here and previously (7, 30).

The *E2* and *01* Genes

The *E* genes are as abundant as the *B* genes based on gene sequences identified from genomic DNA of individual sea urchins (19), and all sea urchins have at least one *E* gene copy with most predicted to have two to four and some as many as six copies (7). The *E* genes are the most highly expressed of the *SpTrf* gene family composing 546 of 689 cDNAs reported previously (11, 12). One *E* gene is present in Cluster 1, and two, *E2a* and *E2b*, are in Cluster 2 (7, 30). It is noteworthy that the allele position corresponding to *E2b* in Cluster 2 is the *01* gene in Cluster 1 rather than an *E2* gene. All *0* genes that have been identified from cDNA and gene sequences are named such because of a deletion of the key element used for naming (element 15 in the cDNA alignment or element 10 in the repeat-based alignment; **Figures 3B, C** and **Supplementary Figure S6**, blue box) (12). Hence, the allelic positioning of *E2*b

and *01* has been noted as unusual. The alignment of the *E2* and the *E2a* genes indicated 99% sequence identity with a 100% coverage. In comparison, *E2*b was 96% identical to the other *E2* genes over a 97% coverage (**Table 3**). The decreased percent identity for *E2*b was due to a gap of 12 nt in the first type 1 repeat (element 2 as defined by the repeat-based alignment), and another of 15 nt in element 27 at the 3′ end of exon 2 (**Figure 3C** and **Supplementary Figure S15**). Strikingly, the second gap in *E2*b matched to an identical gap in the *01* gene on Cluster 1. Because of this sequence similarity and because the *01* gene was positioned in the same allelic location as *E2*b (**Figure 1A**), analysis of the *01* gene was included in the comparison among the *E2* genes. The *01* gene had a 96% to 97% identity (85% and 82% coverage, an outcome of the deletion described above) with the *E2* genes (**Table 3**). The element pattern of the *01* gene was similar to the *E2* genes and shared elements 1 to 6, however, unlike the *E2* genes, *01* shared elements 22, 23, and 24 with all of the other genes in both loci based on the repeat-based alignment (**Figure 3C**). An alignment of the *E2* and *01* genes showed that the only differences among the four genes was a region of 32 nt that was preceded by a gap of 90 nt (**Supplementary Figure S15**, yellow highlights). The *E2* and *01* genes all had δ introns (**Figure 4A**), although the *01* intron had a deletion of 60 nt making it the shortest intron among the *SpTrf* genes (**Supplementary Figure S15**, yellow highlights). The *E2* genes all had stop codons in the 'a' position, while *01* had a stop codon in the 'b' position (**Figure 3A**). Overall, the *E2* genes showed sequence similarity not only to each other but also to the *01* gene. In turn, the *01* gene had the highest level of similarity with the *E2*b gene, with which it appears to be allelic.

The Majority of SNPs and Other Nucleotide Changes in Exon 2 Are Non-Synonymous

The *SpTrf* genes are expressed during sea urchin immune responses (10–12) and the encoded native proteins function as opsonins and augment phagocytosis (16). Genes that encode pathogen binding proteins are often under strong evolutionary pressure and selection from pathogen contact to optimize pathogen binding either to diversifying pathogens or to non-variable PAMPs. To determine whether the genes in the four clusters were diversifying at different rates relative to each other, the dN/dS scores were calculated among genes with the same element pattern (12). Comparisons among genes in these subsets of element patterns indicated both diversifying (dN/dS > 1) and purifying (dN/dS < 1) selection, although results did not typically vary by more than ±0.7 (**Table 4** and **Supplementary Table S5**). The two *A2* genes and the two *B8* genes had scores indicating purifying selection relative to each other suggesting that these alleles had not undergone much divergence. The average dN/dS value obtained for the *D1* genes (n = 7) varied depending on the analytic approach and was inconclusive (1.10565 from SLAC[10] and 0.9402 from SNAP[9]) (**Supplementary Table S5**). dN/dS values calculated in SNAP suggested that each of the *D1* genes was diversifying differently, and when pairs of *D1* genes were compared, results showed that some were undergoing diversifying selection (dN/dS > 1; *D1*f, *D1*h, and *D1*g) while others were undergoing purifying selection (dN/dS < 1; *D1*d, *D1*e, *D1*y, and *D1*b) (**Table 4** and **Supplementary Table S5**). When nonsynonymous and synonymous nucleotide changes were identified from an alignment of exon 2 from genes with the same element pattern they showed a variety of SNPs with the majority resulting in nonsynonymous changes in exon 2 that changed the encoded amino acid by either charge or pI (**Supplementary Table S2**). These results suggested that the genes were diversifying or evolving, but at different rates.

Phylogenetic Analysis Suggests Evolutionary Relationships Among the *SpTrf* Genes

Immune genes are often duplicated (reviewed in 37, 79) and the *SpTrf* gene family is no exception; duplicated genes are tightly clustered in discrete regions of the genome (11, 12, 19, 30). Given the nature of these genes and their function in sea urchin immune responses (16–18), attempts have been made to understand their theoretical evolutionary history (29). The previous analysis was limited to the exons and introns of the genes, and the six internal repeats in exon 2 because the sequences of the UTRs and IGRs were unavailable at the time. To address the question of *SpTrf* gene family evolution with the currently available sequence data, phylogenetic analyses were conducted for the *SpTrf* genes in the four clusters to evaluate the relationships among the 5′FR, the intron, exon 2, and the 3′FR. FRs were defined as sequences flanking both sides of the coding region that extended to the surrounding GA repeats and included the 5′ or 3′ UTRs. Sequences of the *Trf* genes from the sea urchin, *Heliocidaris erythrogramma* (*HeTrf*) (65), were used as the outgroup for analysis of exon 2 and the intron, while the 5′FR and 3′FR of the *Trf* sequences from the sea urchin, *Lytechinus variegatus* (*LvTrf*), were used as the outgroup for the 3′FR and 5′FR analysis. To date, *Trf* genes have been identified in six sea urchin species (65, 80–82), all of which are members of the Camarodonta order of euechinoids (83). Of these species, *Lytechinus* [LCA ~60 MYA (84)] and *Heliocidaris* [LCA ~75 MYA (84)] are not members of the Strongylocentrotid family (85) and therefore were appropriate choices as outgroups. The initial phylogenetic analysis of exon 2 from 138 *SpTrf* genes including those from the two genomic loci described here resulted in a polytomic tree structure that was an outcome of the large gaps required for optimal alignments (**Supplementary Figure S16**). Although this type of tree structure has been noted previously because of the mosaic element structure of exon 2, the structure was uninformative with regard to inferring evolutionary relationships among the *SpTrf* genes. Therefore, the dataset for exon 2 was decreased to only the genes in the clusters in an alternative approach to parse out putative relatedness among these genes. The resulting phylogenetic tree showed three major clades in which the earliest branch was composed of the *A2* genes, plus two sister clades that included a weakly supported cluster of the *B8* and *C4* genes, and a weakly supported cluster of the *E2*/*01* and *D1* genes (**Figure 4B**). Overall, the phylogenetic analysis of exon 2 suggested possible evolutionary relatedness among the genes.

TABLE 4 | dN/dS values for genes with the same element pattern show that some are undergoing positive selection while others are undergoing purifying selection.

Genes compared		S	*N*	*aa*[1]	*Sd*	*Sn*	*dN/dS*[2]
A2	*A2*a	6	10	7	6	10	0.445545
B8	*B8*a	3	5	4	3	5	0.463235
C4	*C4*a	0	0	0	0	0	N/A[3]
D1f	*D1*d	2	12	8	3	13	1.248227
D1f	*D1*e	3	16	13	3.5	16.5	1.365854
D1f	*D1*y	3	12	10	3	12	1.148936
D1f	*D1*g	4	15	11	4	15	1.074074
D1f	*D1*b	5	12	9	5	12	0.686441
D1f	*D1*h	0	0	0	0	0	N/A
D1h	*D1*d	2	12	8	3	13	1.248227
D1h	*D1*e	3	16	13	3.5	16.5	1.365854
D1h	*D1*y	3	12	10	3	12	1.148936
D1h	*D1*g	4	15	11	4	15	1.074074
D1h	*D1*b	5	12	9	5	12	0.686441
D1d	*D1*e	5	17	13	5.5	18.5	0.961373
D1d	*D1*y	5	12	8	5	13	0.735849
D1d	*D1*g	6	16	11	6	17	0.803922
D1d	*D1*b	7	12	9	7	13	0.52349
D1e	*D1*y	2	7	5	2.5	7.5	0.857143
D1e	*D1*g	3	11	6	3.5	11.5	0.932432
D1e	*D1*b	4	7	4	4.5	7.5	0.473684
D1y	*D1*g	1	6	3	1	6	1.714286
D1y	*D1*b	2	2	1	2	2	0.285714
D1g	*D1*b	1	4	2	1	4	1.142857
E2	*E2*a	3	3	3	3	3	0.283871
E2	*E2*b	3	7	6	5	9	0.522556
E2a	*E2*b	2	9	8	4	11	0.801887

[1]*The number of amino acid (aa) changes that encode either a change in polarity or pI.*

[2]*dN/dS values were generated for each gene both manually (first three columns) and by SNAP (last three columns). See* ***Supplementary Table S5*** *for more detailed calculations of these values.*

[3]*N/A, dN/dS values cannot be calculated.*

However, given the blocks of elements in exon 2, it was necessary to conduct additional detailed phylogenetic analyses to verify the pattern of the exon 2 tree. The phylogenetic tree for the introns initially employed to identify the intron types [see above (19)] was used to evaluate sequence similarities among introns from the four clusters with 39 introns from *HeTrf* genes (65) employed as an outgroup. The structure of the intron tree was composed of three strongly supported clades composed of δ introns, γ introns, plus a mixed clade of α, β, and ϵ introns (**Figure 4A** and **Supplementary Figures S2, S4**). The intron tree clearly identified the intron types for the *SpTrf* genes in the four clusters, which was in agreement with the previous report (19). The phylogenetic tree of the introns replicated the general structure of the tree for exon 2 (**Figures 4A, B**) strongly supporting the notion that genes with the same element pattern in exon 2 also share the same intron type.

The coding regions of immune genes are often poorly conserved either in sub-regions or throughout the coding regions because of host-pathogen arms race that drives selection for sequence diversification (9, 38, 45). Therefore, exon 2 may not be the optimal sequence to evaluate the relatedness among these genes. As an alternative approach to this problem, the FRs associated with the *SpTrf* coding regions were used in a phylogenetic analysis to avoid the variations in sequence and length for exon 2. This approach has been reported previously to understand the phylogeny of mini-genes encoding microRNAs (86). Phylogenetic trees of the 3′FRs and 5′FRs were generated in MEGA7 using PRANK alignments with the *LvTrf* 5′FR and 3′FR as the outgroup sequences. The 5′FR tree had two major clades in which Clade I consisted of the 5′FRs from the *01* gene, the *A2* genes, the *E2* genes, and three of the *D1* genes (**Figure 4C**). Clade II was composed of two sister sub-clades of which one contained the remaining *D1* 5′FRs and the second included the 5′FRs from the *B8* and *C4* genes. Although the bootstrap value was low for the node separating these two sister groups, this 5′FR tree structure was consistent with the structure of the intron tree (**Figure 4A**). The 3′FR tree showed good support for two major clades composed of the 3′FRs from the *A2* genes in one clade, and the 3′FRs from the other genes in the second clade. The 3′FRs from genes with similar element patterns clustered into four sub-clades composed of i) the *B8* and *C4* genes, ii) the *E2*b and *01* genes, iii) the *E2* and *E2*a genes, iv) and the *D1* genes (**Figure 4D**). Similarities among the structures of the phylogenetic trees for exon 2, and both of the FR trees indicated that FR sequences surrounding genes of the same element pattern were also similar and sufficiently different from those associated with genes of different element patterns to result in structural agreement among phylogenetic trees (**Figures 4A, C, D**).

Because of the variation in the structures of the 5′FR and 3′FR trees, a third assessment was carried out. The 5′FRs and 3′FRs (both ~ 400 nt in length) were aligned and then concatenated for each gene to generate the 5′-3′FR alignment and tree (**Figure 4E**). Alternatively, the 5′-3′FR sequences were concatenated and then aligned which gave tree structures that were essentially the same

(data not shown). This was done to understand the possible evolutionary relationships among the genes without the coding and intron regions that may have affected or driven the outcome of tree topographies due to both length and sequence complexity of those regions of the genes relative to the short sequences of 5′ FR and 3′FR. By analyzing the longer, concatenated 5′-3′FR sequences, each nucleotide and each difference was weighted less in the final tree calculations. The 5′-3′FR tree generated a more robust and definitive tree with regard to the sequence relationships among the genes (**Figure 4E**). Results showed that the 5′-3′FRs from the *D1* genes formed a single clade with two sub-clades (**Figures 4E, i, ii**) composed of i) the 5′-3′FRs from the *D1*g, *D1*y, *D1*b, and *D1*e genes and ii) the *D1*d, *D1*f, and *D1*h genes. Unexpectedly, the 5′-3′FRs from the *D1*e and *D1*d genes from Cluster 2 were separated into different sub-clades (**Figure 4E**, light red boxes). Furthermore, the 5′-3′FR from *D1*d clustered with the 5′-3′FRs from *D1* genes in Locus 2 (**Figure 4E**, light red *vs*. blue boxes). The 5′-3′FRs from the *B8* and *C4* genes clustered together consistently and were sister to the *D1* clade, and the 5′-3′FRs from the *E2* and *E2*a genes also clustered together. The overall structure of the 5′-3′FR tree showed two sister clades with a ladderlike structure for the rest of the tree. The similarities in the structures of the three FR trees (**Figures 4C–E**) indicated that the *SpTrf* genes could be separated into two major groups in which the *D1, B8*, and *C4* genes may have had a shared evolutionary history, while the *E2, 01*, and *A2* genes may have undergone a separate evolutionary history.

Percent Mismatches Highlight Sequence Similarities Among Genes of Different Element Patterns

A complementary approach to using phylogenetic trees to derive evolutionary relationships among the *SpTrf* genes is to calculate the percent mismatch between pairs of genes. These values give a general view of gene sequence similarities and whether those similarities may be due to random chance or to true similarity. A similar analysis was reported using a pairwise distance matrix for the full-length genes that included the introns and four flanking regions [see Figure 9 in (30)]. Here, we used the same approach to analyze the 5′FR, exon 1, the intron, exon 2, and the 3′FR to reveal the relatedness between each gene with every other gene based on the pairwise distance scores (**Supplementary Table S6**). The results are presented as percent mismatch scores for easier visualization (**Figure 5**). The *A2* genes showed low percent mismatch scores against each other for the 5′FR, the intron, exon 2, and the 3′FR, with the 5′FR showing the greatest mismatch (**Figure 5A**, red line). Although the 5′FRs of the two *A2* genes showed higher mismatches of 18% to 30%, exon 1 showed a percent mismatch that was within the range of scores against the other *SpTrf* genes, which was consistent with the sequence conservation of this exon. The percent mismatch scores for the *A2* introns *vs*. other *SpTrf* introns (14-22%) were similar to the percent mismatch scores for exon 2 (15% ± 1%), however the mismatch scores for the 3′FR were much higher (48-57%) (**Figure 5A**). These data verified that the *A2* genes were similar to one another and were equally dissimilar to all the other *SpTrf* genes.

The *D1* genes in each of the four clusters had nearly identical percent mismatch scores among them. Hence, the percent mismatch scores were averaged for those *D1* genes in the two clusters in Locus 1, which reduced the complexity of the data. The two *D1* genes in Locus 2 were identical and analyzed as a single sequence termed *D1*h/f. Pairwise comparisons among the *D1* gene sequences showed very low percent mismatches for the intron, exon 2, and the 3′FR, whereas the mismatches for the 5′FR and exon 1 had greater variation (**Figure 5B**, green lines; **Supplementary Table S7**). The two *D1* genes in Cluster 2, *D1*e and *D1*d, had different percent mismatches for the 5′FR compared to the set of *D1* genes in Cluster 1, indicating sequence differences between the *D1* genes in the two clusters of Locus 1. Furthermore, the mismatches for the 5′FR among *D1* genes from different loci and mismatches with genes of different element patterns showed a similar range of variation (**Figure 5B**). When the *D1* genes were compared to genes with different element patterns, the percent mismatch scores varied among regions and element patterns. The *E2* and *01* genes (**Figure 5B**, purple and pink lines) showed relatively high percent mismatches against the *D1* genes for the 5′FR and the intron but intermediate scores for exon 2 and the 3′FR. Comparisons between the *D1* genes and the *B8* genes (**Figure 5B**, orange line) and the *C4* genes (**Figure 5B**, brown line) showed intermediate percent mismatch scores for the intron with scores for exon 2 and the 3′FR that were similar to the scores for the *D1* genes *vs.* the *E2* and *01* genes (**Figure 5B**).

The comparison between *B8* and *B8*a showed nearly identical low mismatch scores for all regions (**Figure 5C**), similar to results for the *D1* genes. The percent mismatch scores for exon 1 between the *B8* genes and the other genes were within the same range (7% to 12%). There were two outcomes for the percent mismatches for the introns of the *B8* genes compared to introns from the other genes, with a relatively high mismatch scores against the *A2*, *01*, and *E2* genes, and low scores against the *D1* and *C4* introns (**Figure 5C**). Interestingly, the percent mismatches for both *B8* genes compared to the *C4* gene were low for the 5′FR (**Figure 5C**, green and brown lines) along with the 5′FR against the *D1*-y, g, b, e genes (**Supplementary Table S7**). The mismatch scores for the 5′FR of the *B8* genes against the *A2, E2, 01*, and *D1*f/h/d ranged from 20% to 33%, whereas the percent mismatch scores for the 3′FR were lower for all genes (16% to 20%) except between the allelic *B8* genes and the *C4* genes (**Supplementary Table S7**). The percent mismatch scores for the *C4* genes compared to the other *SpTrf* genes showed similar results as that for the *B8* genes (**Figures 5C, D**). The lowest percent mismatch scores for the *C4* genes across all regions was against the allelic *C4* followed by the *B8* genes and the *D1* genes (**Figure 5D**). These results were in agreement with the phylogenetic tree results, which suggested that the *D1, B8*, and *C4* genes shared greater sequence similarity with each other than with the *E2, A2* and *01* genes.

The comparison between the *E2* and *E2*a genes showed low mismatch scores throughout the sequences of these alleles (**Figure 5E**, lavender line), and although the scores against the

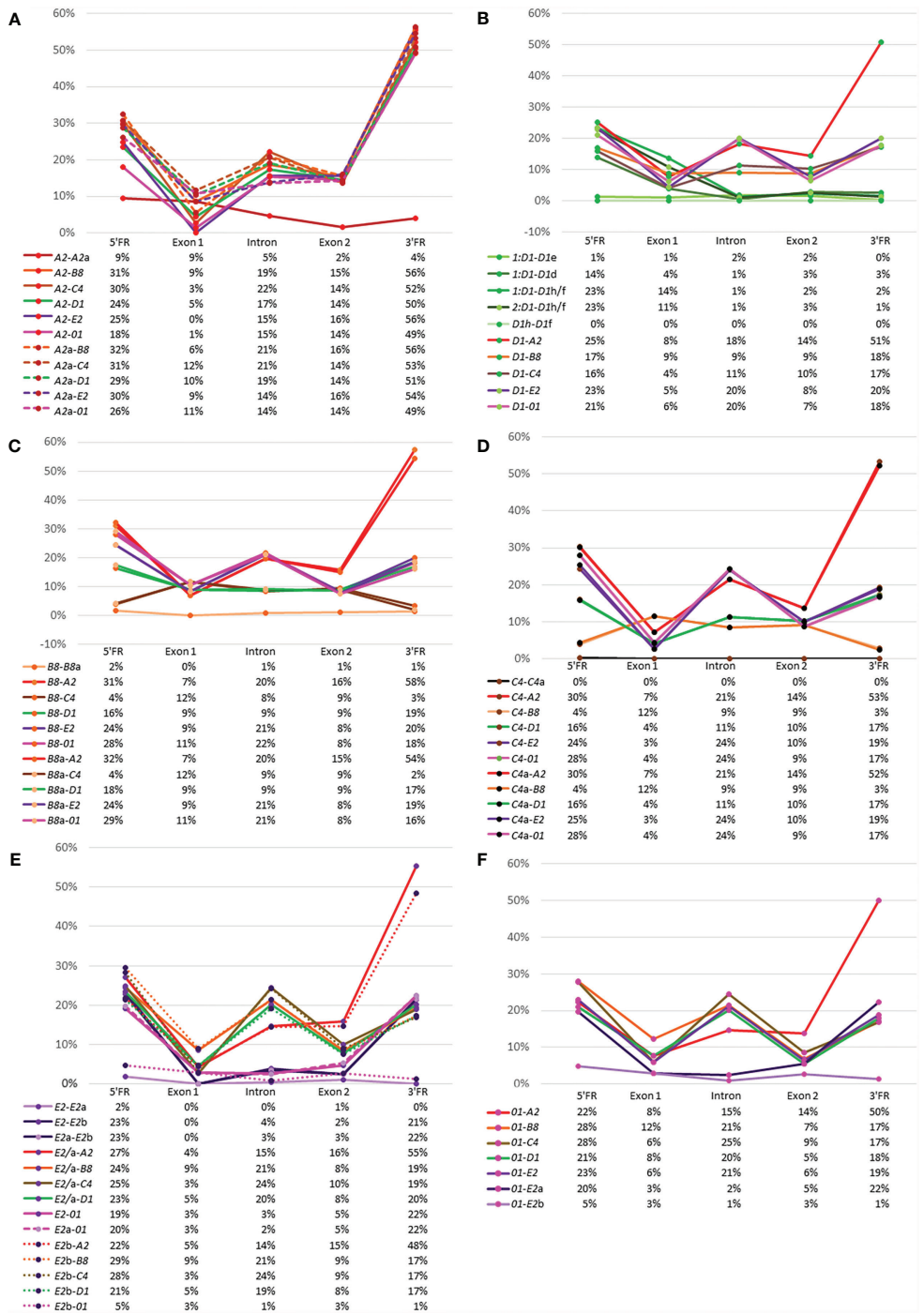


FIGURE 5 | Relatedness among the genes can be inferred from percent mismatch scores for the 5′FR, exon 1, the intron, exon 2, and the 3′FR. Pairwise comparisons among all genes are shown for the FRs, exons, and the intron. The X axis in the graphs indicates the calculated percent mismatch score for each pair of genes and the Y axis indicates the region in the gene. Solid lines, dashed lines, and dotted lines are used to identify the pairs of genes compared. The color of the line corresponds to the color of the genes shown in **Figure 1A** with the exception of the *D1* genes, which are all shown as green lines. Below each graph is a table that gives the percent mismatch scores for each region graphed above. **(A)** The *A2* genes *vs*. other *SpTrf* genes. **(B)** The average percent mismatch of *D1* genes *vs*. other *SpTrf* genes. **(C)** The *B8* genes *vs*. other *SpTrf* genes. **(D)** The *C4* genes *vs*. other *SpTrf* genes. **(E)** The *E2* genes *vs*. other *SpTrf* genes. **(F)** The *01* genes *vs*. other *SpTrf* genes. Percent mismatch [pairwise distance/Ln^2] was calculated from the pairwise distance matrix scores generated with MEGA7 using the PRANK codon alignment.

*E2*b gene were low for exon 1, the intron, and exon 2, higher mismatch scores were noted for both FRs (**Figure 5E**, dark purple lines). When the three *E2* genes were compared to the other *SpTrf* genes, all showed much higher percent mismatch scores for the intron and exon 2, except in the case of the *01* gene, which had low mismatch scores for exon 1, the intron, and exon 2 (**Figure 5E**, pink lines). Similar results were obtained when the regions of the *01* gene were compared to the other *SpTrf* genes (**Figure 5F**). The *01* gene had low percent mismatch scores at the 5′FR and the 3′FR against the same regions in the *E2*b gene but had much higher percent mismatches compared to the *E2* and *E2*a genes (**Figure 5F**, light purple *vs.* dark purple lines). These scores were comparable to scores for the 5FR′ and 3′FR of the *E2* genes and *01* genes against the 5′FR and 3′FR for the other *SpTrf* genes (**Figure 5F**). The percent mismatch scores were consistent with the clustering of the *01* and *E2* genes, specifically with the *E2*b gene, in the phylogenetic trees (**Figure 4**). Overall, these results indicated sequence similarity between the *D1, B8*, and *C4* genes in all regions, similarity between the *E2* and *01* genes, and indicated that the *A2* genes were equally dissimilar to the other *SpTrf* genes in these clusters.

A Modified Hypothesis for the Edges of the Segmental Duplications in the *SpTrf* Gene Clusters

Tandem segmental duplications have been noted in the *SpTrf* gene clusters in Locus 1 based on dot plot analysis, phylogenetic analysis of intergenic segments, and calculations of pairwise sequence diversity between pairs of genes (7, 30). Previous reports based on dot plots indicate that the edges of the segmental duplications are the GAT STRs that surround and are positioned near the 3′ end of the *D1* and *E2* genes [**Figure 6A**, red brackets (30)]. However, with the addition of the *SpTrf* genes in Locus 2 (Clusters 3 and 4), the placement of the edges of the segmental duplications did not match the previously published results for Cluster 1 (30). Dot plot analysis of Cluster 3 compared to itself indicated that the two genes, *C4* and *D1*f, plus their flanking regions were very similar, suggesting a 2.7 kb segmental duplication (**Figure 6B** and **Supplementary Figure S17**, offset diagonals) in agreement with a previous report for the *D1* genes in Cluster 1 (30). Dot plots for the *C4*a and *D1*h genes in Cluster 4 showed identical results (data not shown). However, unlike the previous report, the 5′ end of the *D1*f/h segmental duplications were located at the large GA STR island (**Figure 6B**; see also **Figure 1A**, STR 2) and the 3′ end was located at the short GA STR near the 3′FR of the *D1*f/h genes. Similarly, the *C4*/a segmental duplications of 2.8 kb were positioned between the short GA STR near the 3′ side of the *D1*f/h genes and the 3′ end of the duplications were positioned near the large GA STR islands (**Figure 6B**, brackets and offset diagonals; see also **Figure 1**, STR 3). In these segmental duplications the GAT STRs (**Figure 6B**, black triangles and associated dark gray bars) were located in the center of the offset diagonals and therefore in the center of the segmental duplication rather than at the edges. These results suggested that the GA STRs rather than the GAT STRs defined the edges of the segmental duplications in Locus 2.

The edges of the segmental duplications in Cluster 2 have been assumed to be the same as those in Cluster 1 based on the allelic status of these clusters (7). However, when dot plots were used to compare Locus 2 to Locus 1, a different outcome was identified relative to previous reports (7, 30). The dot plots of Cluster 3 compared to Clusters 1 or 2 indicated that the GA STRs were the most likely edges of the segmental duplications rather than the GAT STRs (**Figures 6C, D**). This redefined the edges of the segmental duplications for the *D1* genes in Locus 1 as GA STRs and indicated that they were the same size as reported previously (~4.5 kb). The new location of the duplications was a shift of 3 kb towards the end of the clusters in which the *A2* genes were positioned (**Figure 6A**, black brackets). The exception to this revised positioning of the segmental duplications in Locus 1 was the IGRs between *E2* and *D1*b in Cluster 1 and *E2*a and *D1*e in Cluster 2. The dot plot results indicated that the duplications terminated at the GAT STR located 5′ of the *D1*b and *D1*e genes (**Figures 6C, D** and **Supplementary Figure S17**), reducing the size of these particular duplications. To confirm the edges of the segmental duplications, alignments of the IGRs between linked genes was done using PRANK (IGRs were located between *B8*/a::*D1*y/d, the linked *D1* genes, *D1*b/e::*E2*/a, and *C4*/a::*D1*f/h) and percent mismatch scores were calculated. Results were ≤10% mismatch for the *B8*/a::*D1*y/d-IGRs, the *C4*/a::*D1*f/h-IGRs, and for all the IGRs between the linked *D1* genes (**Figure 7**, light blue and light purple). In comparison, the *D1*b/e::*E2*/a-IGRs in Locus 1 had ≥79% mismatch compared to the other IGRs indicating that they were not part of discernable segmental duplications (**Figure 7A**, red). Representative results for the percent mismatches between *C4*/a::*D1*f/h-IGRs and the other IGRs illustrated putative segmental duplications based on the results in the Locus 2 dot plots in which the edges of the duplication events were positioned at the GA STRs rather than the GAT STRs (**Figures 7B, C**, green *vs.* black triangles). These data suggested an alternative interpretation of the segmental duplications for this gene family and included the *B8* and *C4* genes in the duplication events with the *D1* genes, which had not been recognized previously.

The Intergenic Regions Show Isolated Regions of Sequence Similarity

Small Regions of Shared Sequence Similarity Exist Among the IGRs Between the *A2*/a, *01*, and *D1* Genes

While the results presented above suggest an evolutionary relationship between the *D1*, *B8*, and *C4* genes and between the *E2* and *01* genes, there was little to suggest any sequence similarity outside of the coding regions between these two subsets of segmental duplications or with the *A2* genes in this gene family. To understand the evolutionary relationship between these two subsets of *SpTrf* genes and the *A2* genes, a region of 3 kb upstream of the 5′FRs and downstream of the 3′FRs of the *A2*/a genes (**Figure 8A**, red brackets) were compared to the i) IGRs between the GA STR islands and *D1*h/f genes (*D1*f/h::GA-IGRs) and ii) the IGRs between the *E2*/a genes and *E2*b/*01* genes (*E2*/a::*E2*b/*01*-IGRs) (**Figure 8A** and **Supplementary Figures S18A–D**). Dot plot analysis identified

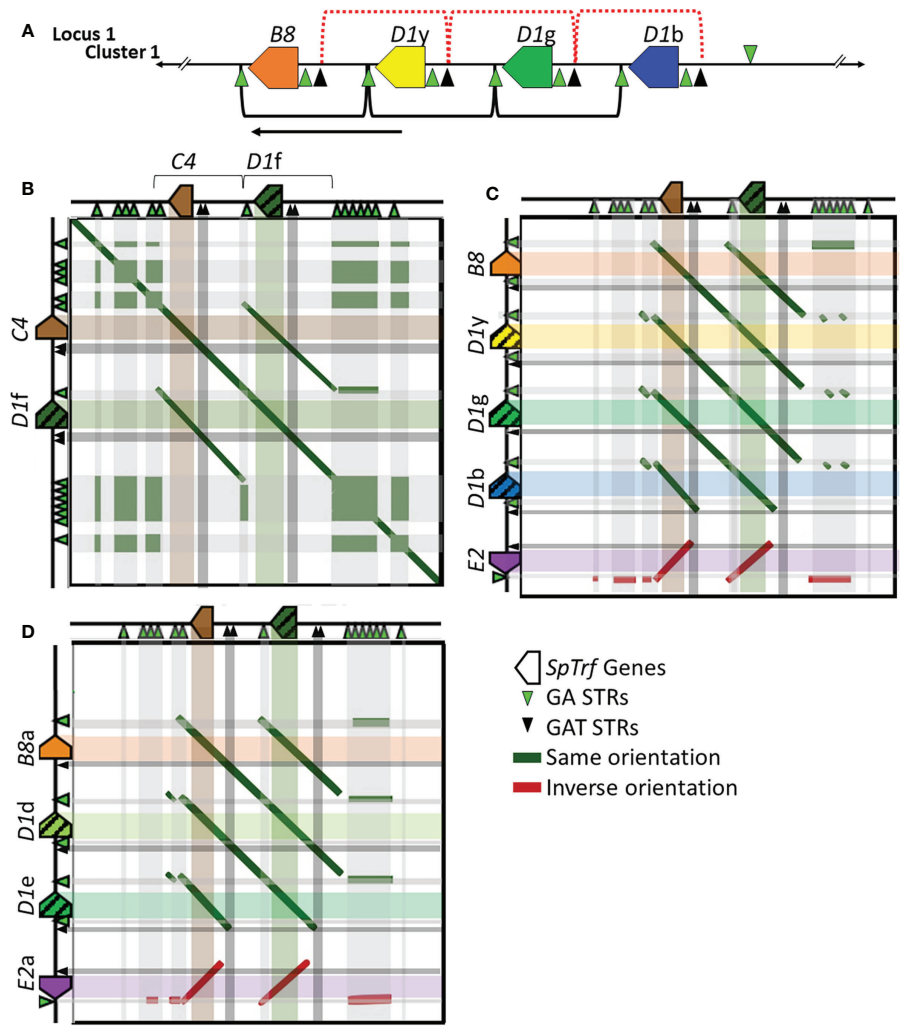


FIGURE 6 | Representative dot plots of Cluster 3 compared to other *SpTrf* clusters indicates that the GA STRs are the likely edges of the segmental duplications in both loci. **(A)** A portion of Cluster 1 shows the region with the segmental duplications and the *D1* genes. The previously reported *D1* segmental duplications are indicated at the top of the figure [red dotted brackets (30)] and the revision to the proposed segmental duplications are indicated on the bottom (black brackets). The arrow indicates the directional shift of the proposed edges of the segmental duplications. **(B–D)** Representative images of a gene cluster or portion of a gene cluster are located to the left and top of each dot plot. Colored polygons indicate genes and transcriptional direction. Green triangles represent GA STRs and black triangles represent GAT STRs. The central diagonal in **(A)** shows the main alignment of cluster 3 against itself, while lines that are offset from the central diagonal in all dot plots indicate the locations of repeats or highly similar regions. Diagonal dark green lines indicate similar regions in the same orientation whereas, dark red solid lines indicate regions of inverse orientation. The highlighted horizontal and vertical lines of multiple colors (matching to the genes at the top or side) are added to the dot plots to illustrate the location of matched sequences. Dark green areas indicate the locations of GA STRs and dark gray areas indicate the locations of the GAT STRs. **(B)** Cluster 3 *vs*. Cluster 3. **(C)** Cluster 3 *vs*. a subset of genes in Cluster 1. **(D)** Cluster 3 *vs*. a subset of genes in Cluster 2. YASS[8] was used to generate dot plots with standard parameters (scoring matrix = +5, -4, -3 -4: composition bias correction: gap costs = -16, -4: e-value threshold = 10: X-drop threshold = 30).

a 700 nt region in the 5′ end of the *A2*/a genes that contained two fragments (**Figure 8A**, red boxes 1 and 2) with sequence similarity to two separated regions in the *D1*f/h::GA-IGRs in which fragment 1 was positioned 1.4 kb from the 5′ end of the *D1*f/h genes (**Figure 8A**, green boxes 1 and 2). Fragment 2 was located 300 nt from the 5′ end of the *D1*f/h genes, similar to its location of 350 nt from the 5′ end of the *A2*/a genes. Fragment 1 was separated from fragment 2 by 730 nt in the *D1*f/h::GA-IGRs but was separated by only 30 nt in the 5′ end of the *A2*/a genes. Fragments 7 and 8 in the 5′ end of the *A2*/a genes were also identified in the *E2*/a::*E2*b/*01*-IGRs but were absent from the *D1*f/h::GA-IGRs (**Figure 8A**, red boxes 7 and 8). Fragments 7 and 8 were 3 kb from the 5′ end of the *E2*b/*01* genes and separated by 130 nt (**Figure 8A**, pink *vs*. red boxes 7 and 8). There were three regions of similarity between the *D1*f/h::GA-IGRs and the *E2*/a::*E2*b/*01*-IGRs (**Figure 8A**, green *vs*. pink boxes 3-5). Fragments 3-5 were larger than fragments 1 and 2 associated with the *A2*/a genes and together composed lengths of 1456 nt to 1483 nt. Fragments 3 and 4 were positioned next to each other in the *D1*f/h::GA-IGRs but were separated by 520 nt

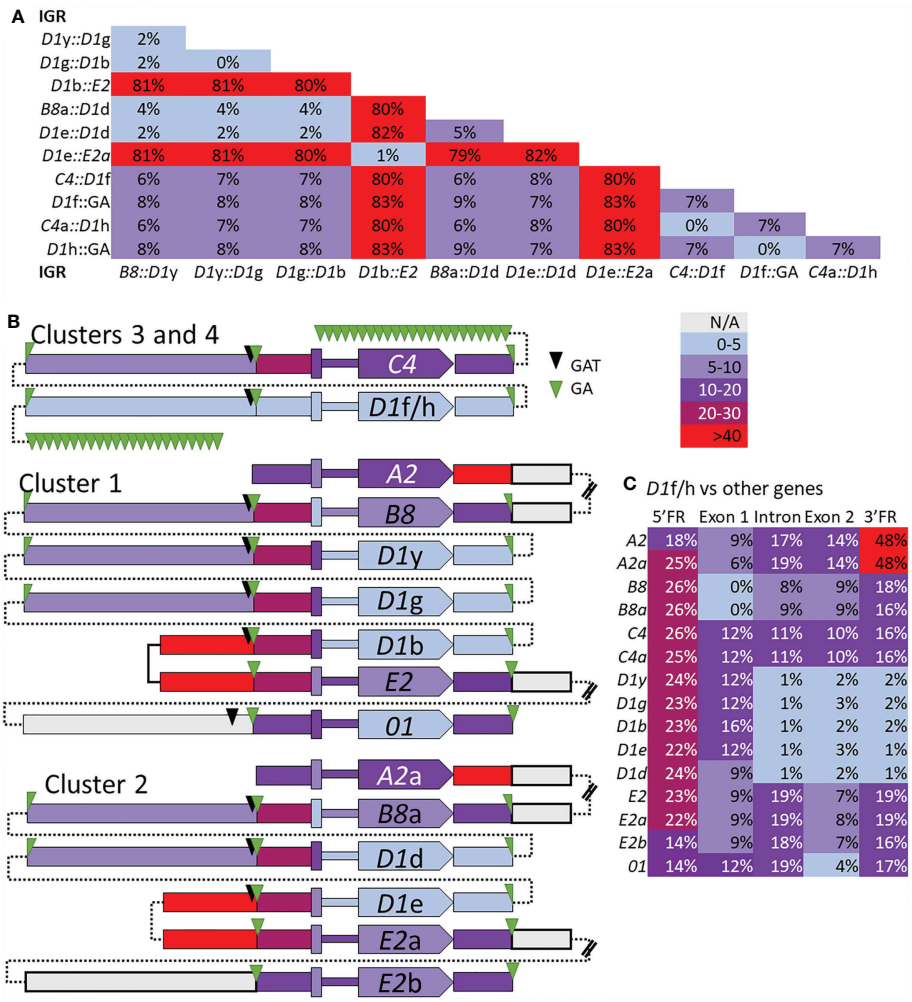


FIGURE 7 | The percent mismatch between regions of the *SpTrf* genes suggests that the segmental duplications include the *B8* and *C4* genes with the *D1* genes. Alignments of the *B8*/a::*D1*y/d-IGRs, the *D1* IGRs, the *D1*b/e::*E2*/a-IGRs, and the *C4*/a::*D1*h/f-IGRs, plus the alignment of the 5′FR, exon 1, the intron, exon 2, and the 3′FR for all *SpTrf* genes were done with PRANK. **(A)** The pair-wise percent mismatch scores for IGRs indicate the level of sequence similarity. Percent mismatches were calculated from pairwise diversity scores in MEGA7 and are indicated with the color gradient legend. There are no mismatch scores between 30-40%. **(B)** A graphical representation shows levels of sequence similarities among genes and IGRs based on percent mismatch scores against the *D1*f/h genes. All genes are oriented in the same direction as indicated by the pointed polygon labeled with the gene name. From left to right across the figure are blocks that represent the IGRs, GA and/or GAT STRs, the 5′FR, the exon 1, the intron (narrow region), the exon 2 with the gene name, and the 3′FR. The thin dotted lines indicate how the sequences are linked together in their respective clusters and do not indicate sequence. The double bars in some IGRs indicate sequence that was not analyzed and is not shown. Percent mismatches for all blocks are color coded based on the gradient key. **(C)** The percent mismatch values for the regions of all *SpTrf* genes compared to the *D1*f/h indicates regions of similarity and dissimilarity. Results are color coded according to the gradient key.

in the *E2*/a::*E2*b/*01*-IGRs (**Figure 8A**, green *vs.* pink boxes 3 and 4). Fragment 5 was 170 nt to 213 nt in length depending on the number of repeats in the GA/GAT STRs. This region was positioned within the *E2*/a::*E2*b/*01*-IGRs and matched to the GA/GAT STRs that made up the boundary of the 5′FR of the *D1*f/h genes. Fragment 5, which was associated with the *D1*f/h 5′FRs, also matched to the GA/GAT STRs that were located closer to the *E2*b/*01* genes and constituted the boundary of the 5′FRs. Only one region matched across all three regions (**Figure 8A**, indicated with an asterisk), which was fragment 2 or 7 in the 5′ end of the *A2*/a genes that was also identified within fragment 4 associated with the *E2*/a::*E2*b/*01*-IGRs and the *D1*f/h::GA-IGRs. No regions of similarity were identified to the 3′ side of the *E2*b/*01* genes compared to the other IGRs (not shown in **Figure 8**). However fragment 6 (**Figure 8A**, red box 6) was identified on the 3′ end of the *A2*a gene, which matched to a sequence located within the *E2*/a::*E2*b/*01*-IGRs and was positioned 730 nt from the 3′FRs of the *E2*/a genes. Fragment 6 was located 1350 nt from the 3′FR of the *A2*a gene and was inverted relative to fragment 6 associated with *E2*/a::*E2*b/*01*-IGRs. Fragment 6 was only identified in the 3′ end of the *A2*a gene and was missing from the 3′ end of the *A2* gene because this was a region of dissimilarity relative to the *A2* IGR (**Figure 2C**, red and white striped triangle). While the 5′ end of the *A2* gene

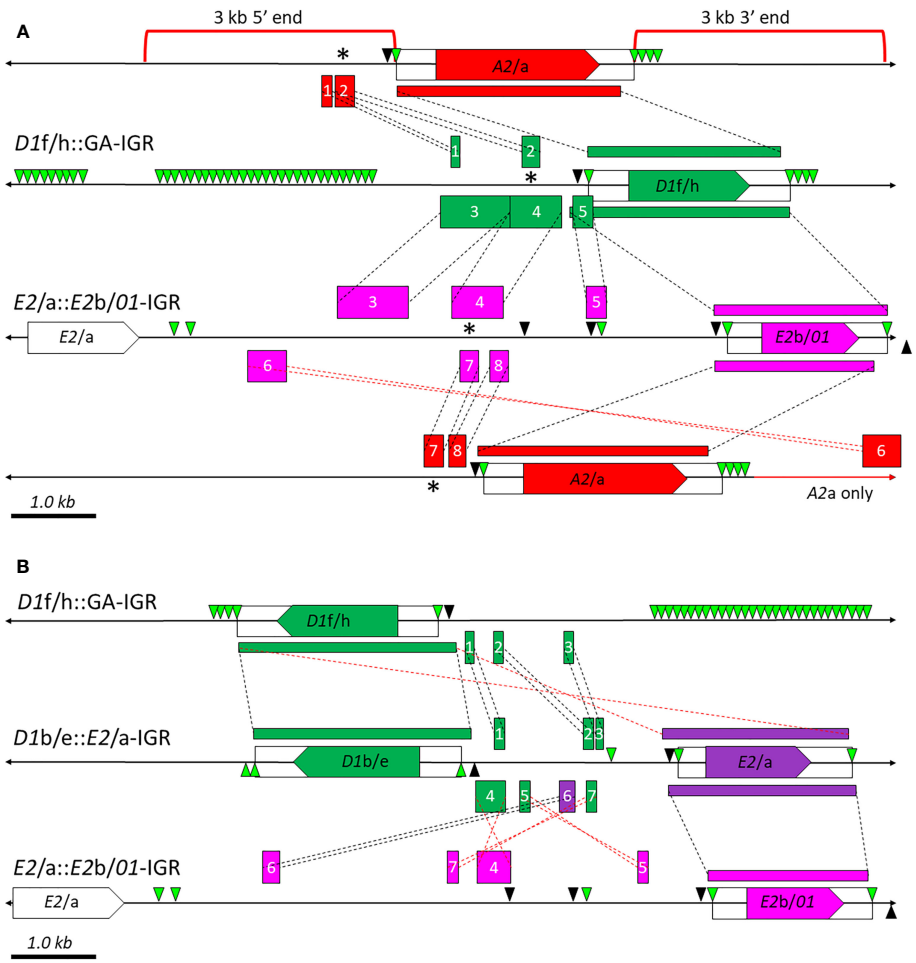


FIGURE 8 | Comparisons the IGRs between the *A2*, *E2*, *01*, and *D1*f/h genes identify short regions of similarity. **(A)** The *D1*f/h::GA-IGRs and *E2*/a::*E2*b/*01*-IGRs are compared to each other, and both are compared to the 5′ and 3′ ends of the *A2*/a genes (indicated by red brackets). **(B)** The *D1*f/h::GA-IGRs, *D1*d/e::*E2*/a-IGRs, and the *E2*/a::*E2*b/*01*-IGRs are compared. Genes are indicated by the polygon labeled with the gene name and are colored according to **Figure 1A** and are flanked by UTRs (open boxes). The genomic DNA is indicated by horizontal black lines that passes behind the genes and includes the IGRs and flanking regions. Genes without color were not included in the analysis and are shown for orientation and comparison to **Figure 1A**. GA (green triangles) and GAT (black triangles) STRs are indicated. The colored boxes above and below the black horizontal line indicate regions of similarity as identified from dot plots from the YASS genomic similarity search tool set to a threshold of e^{-20}. Areas of shared sequence among IGRs are numbered for clarity; see text for detailed description. Dotted lines connect the regions of similarity between IGRs including regions in the same (black lines) and inverted (red lines) orientation. This figure is drawn to scale. * indicates regions of similarity among all three alignments in **(A)**.

and the *E2*/a::*E2*b/*01*-IGRs were not identical to the *D1*f/h::GA-IGRs that were indicative of duplications, there were small fragments of shared sequence. These short fragments of sequence confirmed that there was sequence similarity outside of the coding regions of these genes that linked the *D1*, *B8*, and *C4* segmental duplications with the *E2* and *01* duplications and with the *A2* genes.

There Are Fragmented Regions of Shared Sequence Similarity in the *D1*b/e::*E2*/a-IGRs

The shared sequence fragments in the 5′ and 3′ ends of the *A2*/a genes, in the *E2*/a::*E2*b/*01*-IGRs, and in the *D1*f/h::GA-IGRs suggested that shared sequences may also be identified for the IGRs between the *E2*/a genes and the *D1*b/e genes (*D1*b/e::*E2*/a-IGRs). These IGRs were of interest because the *D1*b/e genes were missing the 5′ end of the proposed *D1*/*B8*/*C4* segmental duplications based on results of dot plot comparisons to Cluster 3 (**Figure 6**), and because these IGRs were short (3.4 kb) (**Figure 8B**) and located to the 5′ side of the *D1*b/e genes and the *E2*/a genes. To understand the complexity of these IGRs, the *D1*f/h::GA-IGRs and the *E2*/a::*E2*b/*01*-IGRs were compared to the *D1*b/e::*E2*/a-IGRs (**Figure 8B**; **Supplementary Figures S18E, F**). Results from the dot plots of the *D1*b/e::*E2*/a-IGRs indicated three short fragments of similarity, 1 - 3, that were present in the corresponding *D1*f/h::GA-IGRs (**Figure 8B**, green boxes 1 - 3). These fragments were in the same orientation in both loci relative to the local *D1* gene. There were four short fragments of similarity, 4 - 7, located in the *E2*/a::*E2*b/*01*-IGRs and the *D1*b/e::*E2*/a-IGRs

(**Figure 8B**, green and purple boxes 4 - 7). Of these four fragments, all but fragment 6 were in the same orientation as the local *D1*b/e genes, whereas fragment 6 was oriented the same orientation as the local *E2*b/*01* genes. This result, in addition to the dot plots (**Figure 6**) indicated that fragments 4, 5, and 7 were likely associated with the *D1* rather than the *E2* gene given that they were oriented in the same direction. However, the fragments 4, 5, and 7, which were in the same orientation as the *D1*b/e genes, were not positioned in the same order in the *E2*/a::*E2*b/*01*-IGRs indicating a possible sequence scrambling in this region. Taken together these data indicated that the regions between the *D1*b/e and the *E2*/a genes contained small fragments of sequence similarity in the IGRs of the *D1* genes and one small fragment that might be attributed to the *E2* genes. This was similar to the results for the *A2*/a analysis (**Figure 8A**). These results illustrated that, while the 5′ IGRs of these gene were not identical, there were short fragments of sequence similarity shared among them that would be consistent with genomic instability for both of the loci that harbor the *SpTrf* gene clusters. These shared regions may have implications not only to the relatedness among the genes but also among the IGRs.

DISCUSSION

A Hypothetical Evolutionary History of the *SpTrf* Gene Family in the Sequenced Sea Urchin Genome

The necessity for diverse and constantly diversifying genes in the face of a broad array of pathogens leads not only to the generation of complex immune systems but to complex immune gene families. Duplications, insertions, inversions, meiotic mispairing, unequal crossing over, and gene conversion all have the potential to result in large and diverse immune gene families encoding proteins that keep pace in the arms race with the pathogens (20, 23, 24, 37, 38, 87). Based on the sequence relationships among the genes in the four clusters including their FRs and IGRs, we propose a hypothetical evolutionary history of how the *SpTrf* gene clusters were generated. The LCA *SpTrf′* gene plus a portion of its 5′ and 3′ flanking regions is the starting sequence for this evolutionary history. *SpTrf′* underwent initial duplications and ectopic insertions into the same locus and into a different region of the genome to establish a second locus (**Figure 9A**). These two loci subsequently underwent gene diversification to generate the ancestral *D1′*, *E2′*, and the *A2* genes (**Figure 9B**). The two loci containing the ancestral *D1′* or *E2′* genes underwent independent secondary duplication events, generating several tandem genes of the same element pattern and forming the initial clusters (**Figure 9C**). These gene duplicates acquired internal SNPs and indels thereby continuing sequence diversification (**Figure 9D**). One outcome was the sequence variation among the *D1* genes and the appearance of the ancestral *B8*/*C4′* gene from *D1* duplications in Locus 2 (**Figures 9C, D**). The other outcome was the diversification of the *E2* genes to generate the *E2b* gene on Locus 1 (**Figures 9C, D**). Next, a larger duplication and ectopic insertion moved at least two *D1* genes plus the ancestral *B8*/*C4′* gene from Locus 2 into Locus 1 that was positioned between the *A2* and *E2* genes (**Figures 9D, E**). This may have been the ancestral change that resulted in genes facing in both directions in Locus 1 and which scrambled the IGR sequences between the *D1* and the *E2* genes. The mismatch in the number of *D1* genes between Clusters 1 and 2 in Locus 1 is likely due to tertiary duplications that generated the *D1*y/g genes, which may have occurred either by a direct duplication of the *D1*y/g genes in Cluster 1 (**Figure 9F**) (30), or by an ectopic insertion from the allele in Cluster 2 (not shown). Finally, the individual *SpTrf* genes underwent further internal indels and SNPs generating the individual sequence variation among the genes, including the generation of the *01* gene from the *E2*b gene and the *B8* and *C4* genes from the *B8/C4′* ancestor. The final outcome is the extant clusters and loci in the sequenced sea urchin genome (**Figure 9G**).

Supporting Evidence for the Evolutionary History of the Extant *SpTrf* Family

The evolutionary history of the *SpTrf* gene family is based on the results presented herein. We speculate that the sequence of the LCA *SpTrf′* gene had the majority of elements and the maximum number of repeats in exon 2 (**Figure 3C**), which subsequently underwent at least two duplications and ectopic insertion events (**Figure 9A**). This is based on alignments of the IGRs of the extant genes, which reveal a number of small regions of sequence similarity across all extant *SpTrf* genes reported here. We also hypothesize that an *SpTrf′* gene with the maximum number of repeats in exon 2 would be the most parsimonious candidate gene sequence to generate other *SpTrf* genes, which are short genes with fewer elements, through deletions rather than vice versa through element or repeat duplication and diversification. The *A2* genes are an exception to this as previous research has proposed that *A2* genes have undergone a large duplication event in exon 2 that increased their size and gave them the designation of long genes (29). We hypothesize that the *A2* genes underwent a separate evolutionary history compared to the *E2′* and *D1′* genes after the duplication and ectopic insertions of the *SpTrf′* (**Figures 9A, B**). The separate evolutionary history of the *A2*/a genes is based on the early branching position of the *A2*/a genes in the phylogenetic trees that infers a later divergence of the short *SpTrf* genes, and is based on the distant location of the *A2*/a genes in Locus 1 that are separated by non-conserved IGRs. This notion is consistent with a previous report speculating that long genes have unique type 1 repeats (see **Figure 3C**, type 1 repeats are shown as red rectangles) that underwent a separate evolutionary history from the type 1 repeats in the short *SpTrf* genes (29).

The similarities between the *E2*/a::*E2*b/*01*-IGRs and the *D1*f/h::GA-IGRs support the idea of a shared evolutionary history among the genes, which extends beyond the similarities of the coding regions and into the 3′ ends of the genes. The *E2*/a::*E2*b/*01*-IGRs in Locus 1 contain large regions that match to sequences in the *D1*f/h::GA-IGRs in Locus 2 that are also present in most of the *D1* segmental duplications. These matching regions are dispersed within the large *E2*/a::*E2*b/*01*-IGRs but are relatively

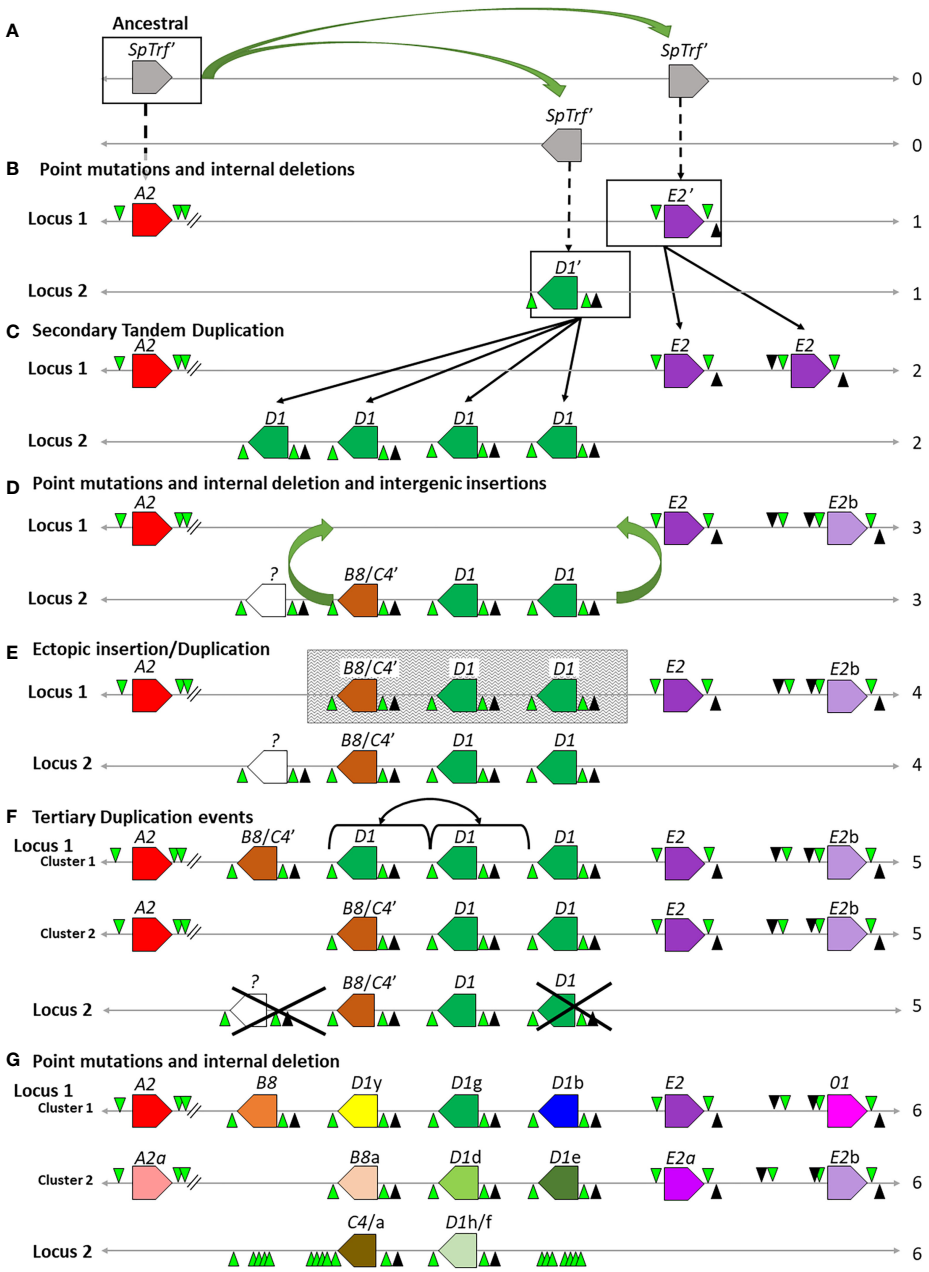


FIGURE 9 | A model for the theoretical evolutionary history of the *SpTrf* gene clusters in the sequenced genome based on gene duplications, ectopic insertions, and deletions. Each step in this theoretical evolutionary history of the gene clusters is indicated on the right with numbers and labeled on the left **(A–G)**. Genes and their direction are indicated by colored polygons and labeled with the gene name. The prime (') associated with a gene name indicates a hypothetical LCA version of the gene. Gene polygons without color indicate genes that are proposed to exist but whose element pattern cannot be determined. Variations in or changes to gene colors indicate internal point mutations/insertions/deletions during the lineage of a particular gene. GA (green triangles) and GAT (black triangles) STRs are shown. The horizontal gray line indicates the IGRs that flank the genes. Open boxes surrounding the genes in **(A, B)** indicate edges of proposed duplication regions. Curved green arrows indicate duplications and ectopic insertions, dotted black arrows indicated the transition from an ancestral *SpTrf* gene to specific *SpTrf* gene lineages, straight black arrows indicate duplications of genes over time. The shaded box shows the ectopically inserted region in Locus 1. Brackets connected with double ended arrows indicate a recent duplication event. Large black Xs indicate gene deletions. This figure is not drawn to scale..

contiguous in the *D1*f/h::GA-IGRs. This suggests that the *E2*/a::*E2*b/*01*-IGRs may have originally been similar in size to the *D1*f/h::GA-IGRs and underwent a number of insertion events to separate the regions of sequence similarity and elongate the IGRs to their current size (**Figures 9C, D**). On the other hand, when the *D1*f/h::GA-IGRs are compared to the *D1*b/e::*E2*/a-IGRs only short, fragmented regions of sequence similarity are identified. These short regions may have been the outcome of the proposed

ectopic insertion of the *D1*/*B8*/*C4* region from Locus 2 into Locus 1 (see below; **Figures 9D, E**). This evolutionary history suggests that both the *D1* and the *E2* genes were both products of the *SpTrf'* ancestral gene duplication that subsequently underwent separate evolutionary histories to generate the two subsets of *E2*/*01* and *D1*/*B8*/*C4* genes (**Figure 9**).

The sequence diversification of the *D1* genes, which are present in segmental duplications, are based on sequence analysis of the *D1* genes and their flanking regions. In agreement with Miller et al. (30), the *D1* genes appear to be a product of multiple recent duplication events that is also supported by our phylogenetic analysis and percent mismatch scores, which includes similarities among the FRs. However, based on our analyses, we hypothesize that the ancestral *D1′* gene was most similar to the *D1* genes in Cluster 1 plus *D1*e in Cluster 2 (**Figures 9C–G**) because these genes are more similar to each other than to the remaining *D1* genes in either locus. This result is also consistent with purifying selection detected for the *D1*y/b/e genes and for diversifying selection for the *D1*f/h genes. Although the identity between *D1*f and *D1*h could be based on their location in Locus 2, a more in-depth analysis suggests a specific evolutionary relationship among the *D1* genes in the two loci, which is based on two levels of results. First is a hypothesized evolutionary relationship among the *D1* genes with the *B8* and *C4* genes. This is based on the sequence similarity among these genes along with the updated edges of the *D1* segmental duplications to include the *C4* and *B8* genes. Both the *B8* and *C4* genes may have once initially been a product of a *D1′* gene that underwent diversification events in Locus 2 to generate a descendant LCA *B8*/*C4′* gene (**Figures 9C, D**), along with duplications of an unknown number of additional *D1* genes, that would later go on to become the extant *C4* and *B8* genes. Although the number of duplicated *D1* genes that may have been present in Locus 2 is unknown, the large islands of GA STRs associated with this gene cluster may be the remnants of gene deletions (7). Secondly, there are indications that the *B8* genes and several *D1* genes in Locus 1 may have been the product of a duplication and ectopic insertion event from Locus 2 (**Figures 9D, E**). This idea is supported by the sequence similarity between the *B8* and *C4* genes, which are located in allelic positions in the two extant loci. A recent evolutionary history between the *B8* and *C4* genes is supported by phylogenetic analysis, percent mismatch scores, and dot plot analysis. The duplication of the *D1′* and *B8*/*C4′* genes in Locus 2 and the location of their insertion in Locus 1 (**Figures 9D, E**) is supported by the IGR sequences on either side of the *B8* and *D1* genes, which are either highly dissimilar (*A2*/a::*B8*/a-IGRs) or show signatures of sequence scrambling (*D1*b/e::*E2*/a-IGRs) (**Figures 9D, E**). The outcome of the ectopic insertion is a heterogeneous cluster of genes in Locus 1 the include both *D1* derived genes and *E2* derived genes that are present in opposite orientations (**Figure 9F**).

The appearance of the *E2* and *01* genes is proposed to have originated with the *E2′* gene (**Figure 9B**). *E2′* initially underwent a tandem duplication to form two *E2* genes in Locus 1 (**Figures 9B, C**). This was followed by sequence diversification of one of the *E2* genes into *E2*a and *E2*b in Cluster 1 and Cluster 2, respectively (**Figures 9C, D**). The *E2*b allele in Cluster 1 subsequently acquired multiple deletions that resulted in the *01* gene (**Figure 9G**) including a large deletion in exon 2 that maintained the reading frame either fortuitously or through unknown repair mechanisms (88). The evolutionary relationship between the *E2* and *01* genes is noteworthy because the sequence similarity between the *01* gene and the three *E2* genes has not been reported previously.

CONCLUSION

Overall, the evolutionary history of this gene family suggests a number of duplications, deletions, insertions, conversions, and point mutations, all of which lead to the distinct clustering and sequence similarity among the members of this gene family (**Figure 9**). It must be noted, however, that this hypothetical evolutionary history is based on genes from a single sea urchin and that different sea urchins have been proposed to contain different repertoires of this gene family (8, 12, 19). Variations among *SpTrf* gene repertoires can be considered as a population level immunological benefit in an environment with many potential pathogens. Additional gene sequence data and cluster structure from other individual sea urchins will either clarify and verify this history or will expand the numbers of genes and their sequence variations to further illuminate the evolution of this gene family.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MABH conceived of the project, generated the data, and wrote the paper. LCS acquired the funding and wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by funding from the Wilber V. Harlan Trust through the Department of Biological Sciences at GWU to MABH, and from the US National Science Foundation (IOS-1146124, IOS-1550474, IOS-1855747) to LCS.

ACKNOWLEDGMENTS

The authors would like to acknowledge the laboratory assistance from Hung-Yen Chou and Caroline Reynolds for sea urchin care,

Drs. Katherine Buckley and Matan Oren for bioinformatics advice, and Drs. Leon Grayfer and Damian O'Halloran for improvements to the manuscript. We are indebted to the two reviewers whos thoughtful and expertise comments resulted in a significantly improved paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.744783/full#supplementary-material

REFERENCES

1. Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen AH, Terwilliger DP, et al. The Immune Gene Repertoire Encoded in the Purple Sea Urchin Genome. *Dev Biol* (2006) 300(1):349–65. doi: 10.1016/j.ydbio.2006.08.065
2. Rast JP, Smith LC, Loza-Coll M, Hibino T, Litman GW. Genomic Insights Into the Immune System of the Sea Urchin. *Science* (2006) 314(5801):952–6. doi: 10.1126/science.1134301
3. Buckley KM, Rast JP. Dynamic Evolution of Toll-Like Receptor Multigene Families in Echinoderms. *Front Immunol* (2012) 3:136. doi: 10.3389/fimmu.2012.00136
4. Pancer Z. Dynamic Expression of Multiple Scavenger Receptor Cysteine-Rich Genes in Coelomocytes of the Purple Sea Urchin. *Proc Natl Acad Sci USA* (2000) 97(24):13156–61. doi: 10.1073/pnas.230096397
5. Pancer Z. Individual-Specific Repertoires of Immune Cells SRCR Receptors in the Purple Sea Urchin (*S. purpuratus*). *Adv Exp Med Biol* (2001) 484:31–40. doi: 10.1007/978-1-4615-1291-2_3
6. Buckley KM, Ho ECH, Hibino T, Schrankel CS, Schuh NW, Wang G, et al. IL17 Factors Are Early Regulators in the Gut Epithelium During Inflammatory Response to *Vibrio* in the Sea Urchin Larva. *eLife* (2017) 6: e23481. doi: 10.7554/eLife.23481
7. Oren M, Barela Hudgell MA, D'Allura B, Agronin J, Gross A, Podini D, et al. Short Tandem Repeats, Segmental Duplications, Gene Deletion, and Genomic Instability in a Rapidly Diversified Immune Gene Family. *BMC Genomics* (2016) 17(1):900. doi: 10.1186/s12864-016-3241-x
8. Oren M, Rosental B, Hawley TS, Kim GY, Agronin J, Reynolds CR, et al. Individual Sea Urchin Coelomocytes Undergo Somatic Immune Gene Diversification. *Front Immunol* (2019) 10:1298. doi: 10.3389/fimmu.2019.01298
9. Buckley KM, Rast JP. Diversity of Animal Immune Receptors and the Origins of Recognition Complexity in the Deuterostomes. *Dev Comp Immunol* (2015) 49:179–89. doi: 10.1016/j.dci.2014.10.013
10. Nair SV, Del Valle H, Gross PS, Terwilliger DP, Smith LC. Macroarray Analysis of Coelomocyte Gene Expression in Response to LPS in the Sea Urchin. Identification of Unexpected Immune Diversity in an Invertebrate. *Physiol Genomics* (2005) 22(1):33–47. doi: 10.1152/physiolgenomics.00052.2005
11. Terwilliger DP, Buckley KM, Brockton V, Ritter NJ, Smith LC. Distinctive Expression Patterns of *185/333* Genes in the Purple Sea Urchin, *Strongylocentrotus purpuratus*: An Unexpectedly Diverse Family of Transcripts in Response to LPS, β-1,3-Glucan, and dsRNA. *BMC Mol Biol* (2007) 8:16. doi: 10.1186/1471-2199-8-16
12. Terwilliger DP, Buckley KM, Mehta D, Moorjani PG, Smith LC. Unexpected Diversity Displayed in cDNAs Expressed by the Immune Cells of the Purple Sea Urchin, *Strongylocentrotus purpuratus*. *Physiol Genomics* (2006) 26 (2):134–44. doi: 10.1152/physiolgenomics.00011.2006
13. Brockton V, Henson JH, Raftos DA, Majeske AJ, Kim Y, Smith LC. Localization and Diversity of 185/333 Proteins From the Purple Sea Urchin – Unexpected Protein-Size Range and Protein Expression in a New Coelomocyte Type. *J Cell Sci* (2008) 121(Pt 3):33–348. doi: 10.1242/jcs.012096
14. Golconda P, Buckley KM, Reynolds CR, Romanello JP, Smith LC. The Axial Organ and the Pharynx Are Sites of Hematopoiesis in the Sea Urchin. *Front Immunol* (2019) 10:870. doi: 10.3389/fimmu.2019.00870
15. Ho ECH, Buckley KM, Schrankel CS, Schuh NW, Hibino T, Solek CM, et al. Perturbation of Gut Bacteria Induces a Coordinated Cellular Immune Response in the Purple Sea Urchin Larva. *Immunol Cell Biol* (2017) 95 (7):647. doi: 10.1038/icb.2017.40
16. Chou H-Y, Lun CM, Smith LC. SpTransformer Proteins From the Purple Sea Urchin Opsonize Bacteria, Augment Phagocytosis, and Retard Bacterial Growth. *PloS One* (2018) 13(5):e0196890. doi: 10.1371/journal.pone.0196890
17. Lun CM, Bishop BM, Smith LC. Multitasking Immune Sp185/333 Protein, rSpTransformer-E1, and Its Recombinant Fragments Undergo Secondary Structural Transformation Upon Binding Targets. *J Immunol* (2017) 198 (7):2957–66. doi: 10.4049/jimmunol.1601795
18. Lun CM, Schrankel CS, Chou H, Sacchi S, Smith LC. A Recombinant Sp185/333 Protein From the Purple Sea Urchin has Multitasking Binding Activities Towards Certain Microbes and PAMPs. *Immunobiology* (2016) 221(8):889–903. doi: 10.1016/j.imbio.2016.03.006
19. Buckley KM, Smith LC. Extraordinary Diversity Among Members of the Large Gene Family, *185/333*, From the Purple Sea Urchin, *Strongylocentrotus purpuratus*. *BMC Mol Biol* (2007) 8:68. doi: 10.1186/1471-2199-8-68
20. Duggal NK, Emerman M. Evolutionary Conflicts Between Viruses and Restriction Factors Shape Immunity. *Nat Rev Immunol* (2012) 12:687–95. doi: 10.1038/nri3295
21. Hughes AL, Yeager M. Natural Selection at Major Histocompatibility Complex Loci of Vertebrates. *Annu Rev Genet* (1998) 32:414–35. doi: 10.1146/annurev.genet.32.1.415
22. Taketa DA, De Tomaso AW. *Botryllus schlosseri* Allorecognition: Tackling the Enigma. *Dev Comp Immunol* (2015) 48(1):654–65. doi: 10.1016/j.dci.2014.03.014
23. Joshi RK, Nayak S. Perspectives of Genomic Diversification and Molecular Recombination Towards *R*-Gene Evolution in Plants. *Physiol Mol Biol Plants* (2013) 19(1):1–9. doi: 10.1007/s12298-012-0138-2
24. Oren M, Barela Hudgell MA, Golconda P, Lun CM, Smith LC. Genomic Instability and Shared Mechanisms for Gene Diversification in Two Distant Immune Gene Families: The Plant *NBS-LRR* Genes and the Echinoid *185/333* Genes. In: Malagoli D, editor. *The Evolution of the Immune System: Conservation and Diversification*. Elsevier Inc. Academic Press, London (2016). p. 295–310.
25. Meyers BC, Kaushik S, Nandety RS. Evolving Disease Resistance Genes. *Curr Opin Plant Biol* (2005) 8(2):129–34. doi: 10.1016/j.pbi.2005.01.002
26. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. Genome-Wide Analysis of NBS-LRR-Encoding Genes in Arabidopsis. *Plant Cell* (2003) 15:809–34. doi: 10.1105/tpc.009308
27. Leister D. Tandem and Segmental Gene Duplication and Recombination in the Evolution of Plant Disease Resistance Genes. *Trends Genet* (2004) 20 (3):116–22. doi: 10.1016/j.tig.2004.01.007
28. Johnson RD, Jasin M. Sister Chromatid Gene Conversion is a Prominent Double-Strand Break Repair Pathway in Mammalian Cells. *EMBO J* (2000) 19 (13):3398–407. doi: 10.1093/emboj/19.13.3398
29. Buckley KM, Munshaw S, Kepler TB, Smith LC. The *185/333* Gene Family is a Rapidly Diversifying Host-Defense Gene Cluster in the Purple Sea Urchin *Strongylocentrotus purpuratus*. *J Mol Biol* (2008) 379(4):912–28. doi: 10.1016/j.jmb.2008.04.037
30. Miller CA, Buckley KM, Easley RL, Smith LC. An *Sp185/333* Gene Cluster From the Purple Sea Urchin and Putative Microsatellite-Mediated Gene Diversification. *BMC Genomics* (2010) 11(1):575. doi: 10.1186/1471-2164-11-575
31. Tang H. Genome Assembly, Rearrangement, and Repeats. *Chem Rev* (2007) 107(8):3391–406. doi: 10.1021/cr0683008
32. Gemayel R, Vinces M, Legendre M, Verstrepen K. Variable Tandem Repeats Accelerate Evolution of Coding and Regulatory Sequences. *Annu Rev Genet* (2010) 44:445–77. doi: 10.1146/annurev-genet-072610-155046
33. Pearson CE, Edamura KN, Cleary JD. Repeat Instability: Mechanisms of Dynamic Mutations. *Nat Rev Genet* (2005) 6(10):729–42. doi: 10.1038/nrg1689
34. Gemayel R, Cho J, Boeynaems S, Verstrepen KJ. Beyond Junk-Variable Tandem Repeats as Facilitators of Rapid Evolution of Regulatory and Coding Sequences. *Genes (Basel)* (2012) 3(3):461–80. doi: 10.3390/genes3030461

35. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic Tandem Repeats Generate Functional Variability. *Nat Genet* (2005) 37(9):986–90. doi: 10.1038/ng1618
36. Pâques F, Leung W, Haber J. Expansions and Contractions in a Tandem Repeat Induced by Double-Strand Break Repair. *Mol Cell Biol* (1998) 18:2045–54. doi: 10.1128/MCB.18.4.2045
37. Smith LC, Coscia MR. Tuning the Host-Pathogen Relationship Through Evolution With a Special Focus on the Echinoid Sp185/333 System. *Invertebr Surviv J* (2016) 13:355–73.
38. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary Insights Into Host-Pathogen Interactions From Mammalian Sequence Data. *Nat Rev Genet* (2015) 16:224–36. doi: 10.1038/nrg3905
39. Van Valen L. A New Evolutionary Theory. *Evol Theory* (1973) 1:1–30.
40. Carroll L. *Through the Looking-Glass, and What Alice Found There*. Macmillan and Co. London (1871).
41. Dawkins R, Krebs JR. Arms Races Between and Within Species. *Proc R Soc London - Biol Sci* (1979) 205(1161):489–511. doi: 10.1098/rspb.1979.0081
42. Decaestecker E, King K. Red Queen Dynamics. In: Wertheim B, editor. *Encyclopedia of Ecology*. Elsevier, Amsterdam. (2019). 3:185–92.
43. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. Dynamic Evolution of the Innate Immune System in Drosophila. *Nat Genet* (2007) 39(12):1461–8. doi: 10.1038/ng.2007.60
44. McTaggart SJ, Obbard DJ, Conlon C, Little TJ. Immune Genes Undergo More Adaptive Evolution Than Non-Immune System Genes in *Daphnia pulex*. *BMC Evol Biol* (2012) 12:63. doi: 10.1186/1471-2148-12-63
45. Obbard DJ, Welch JJ, Kim KW, Jiggins FM. Quantifying Adaptive Evolution in the Drosophila Immune System. *PloS Genet* (2009) 5(10):e1000698. doi: 10.1371/journal.pgen.1000698
46. Uhrberg M. The KIR Gene Family: Life in the Fast Lane of Evolution. *Eur J Immunol* (2005) 35(1):10–5. doi: 10.1002/eji.200425743
47. Adema CM, Hertel LA, Miller RD, Loker ES. A Family of Fibrinogen-Related Proteins That Precipitates Parasite-Derived Molecules is Produced by an Invertebrate After Infection. *Proc Natl Acad Sci USA* (1997) 94(16):8691–6. doi: 10.1073/pnas.94.16.8691
48. Liberti A, Leigh B, De Santis R, Pinto MR, Cannon JP, Dishaw LJ, et al. An Immune Effector System in the Protochordate Gut Sheds Light on Fundamental Aspects of Vertebrate Immunity. *Results Probl Cell Differ* (2015) 57:159–73. doi: 10.1007/978-3-319-20819-0_7
49. Litman GW, Cannon JP, Dishaw LJ. Reconstructing Immune Phylogeny: New Perspectives. *Nat Rev Immunol* (2005) 5(11):866–79. doi: 10.1038/nri1712
50. Yuen B, Bayes JM, Degnan SM. The Characterization of Sponge NLRs Provides Insight Into the Origin and Evolution of This Innate Immune Gene Family in Animals. *Mol Biol Evol* (2014) 31(1):106–20. doi: 10.1093/molbev/mst174
51. Thomas JH, Robertson HM. The Caenorhabditis Chemoreceptor Gene Families. *BMC Biol* (2008) 6:42. doi: 10.1186/1741-7007-6-42
52. Malnic B, Godfrey PA, Buck LB. The Human Olfactory Receptor Gene Family. *Proc Natl Acad Sci USA* (2004) 101(8):2584–9. doi: 10.1073/pnas.0307882100
53. Niimura Y, Nei M. Evolutionary Dynamics of Olfactory and Other Chemosensory Receptor Genes in Vertebrates. *J Hum Genet* (2006) 51 (6):505–17. doi: 10.1007/s10038-006-0391-8
54. Go Y. Lineage-Specific Expansions and Contractions of the Bitter Taste Receptor Gene Repertoire in Vertebrates. *Mol Biol Evol* (2006) 23(5):964–72. doi: 10.1093/molbev/msj106
55. Bachmanov A, Bosak N, Lin C, Matsumoto I, Ohmoto M, Reed D, et al. Genetics of Taste Receptors. *Curr Pharm Des* (2014) 20(16):2669–83. doi: 10.2174/13816128113199990566
56. Cameron RA, Mahairas G, Rast JP, Martinez P, Biondi TR, Swartzell S, et al. A Sea Urchin Genome Project: Sequence Scan, Virtual Map, and Additional Resources. *Proc Natl Acad Sci USA* (2000) 97(17):9514–8. doi: 10.1073/pnas.160261897
57. Rebeiz M, Posakony JW. GenePalette: A Universal Software Tool for Genome Sequence Visualization and Analysis. *Dev Biol* (2004) 271(2):431–8. doi: 10.1016/j.ydbio.2004.04.011
58. Hall TA. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp Ser* (1999) 41:95–8.
59. Buckley KM, Florea LD, Smith LC. A Method for Identifying Alternative or Cryptic Donor Splice Sites Within Gene and mRNA Sequences. Comparisons Among Sequences From Vertebrates, Echinoderms and Other Groups. *BMC Genomics* (2009) 10:318. doi: 10.1186/1471-2164-10-318
60. Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. GUIDANCE: A Web Server for Assessing Alignment Confidence Scores. *Nucleic Acids Res* (2010) 38(Web Server issue):W23–28. doi: 10.1093/nar/gkq443
61. Landan G, Graur D. Local Reliability Measures From Sets of Co-Optimal Multiple Sequence Alignments. *Pacific Symp Biocomput 2008 PSB* (2008) 2008:2008:15–24.
62. Sela I, Ashkenazy H, Katoh K, Pupko T. GUIDANCE2: Accurate Detection of Unreliable Alignment Regions Accounting for the Uncertainty of Multiple Parameters. *Nucleic Acids Res* (2015) 43(W1):W7–14. doi: 10.1093/nar/gkv318
63. Löytynoja A, Goldman N. Phylogeny-Aware Gap Placement Prevents Errors in Sequence Alignment and Evolutionary Analysis. *Science* (2008) 320 (5883):163–1635. doi: 10.1126/science.1158395
64. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0. *Mol Biol Evol* (2016) 33(7):1870–4. doi: 10.1093/molbev/msw054
65. Roth MO, Wilkins AG, Cooke GM, Raftos DA, Nair SV. Characterization of the Highly Variable Immune Response Gene Family, *He185/333*, in the Sea Urchin, *Heliocidaris erythrogramma*. *PloS One* (2014) 9(10):e62079. doi: 10.1371/journal.pone.0062079
66. Davidson PL, Guo H, Wang L, Berrio A, Zhang H, Chang Y, et al. Chromosomal-Level Genome Assembly of the Sea Urchin *Lytechinus variegatus* Substantially Improves Functional Genomic Analyses. *Genome Biol Evol* (2020) 12(7):1080–6. doi: 10.1093/gbe/evaa101
67. Noé L, Kucherov G. YASS: Enhancing the Sensitivity of DNA Similarity Search. *Nucleic Acids Res* (2005) 33:W540–3. doi: 10.1093/nar/gki478
68. Korber B. HIV Signature and Sequence Variation Analysis. In: Rodrigo AG, Learn GH, editors. *Computational Analysis of HIV Molecular Sequences*. Dordrecht, Netherlands: Kluwer Academic Publishers (2000). p. 55–72.
69. Kosakovsky Pond SL, Frost SDW. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol Biol Evol* (2005) 22(5):1208–22. doi: 10.1093/molbev/msi105
70. Kosakovsky Pond SL, Frost SDW. Datamonkey: Rapid Detection of Selective Pressure on Individual Sites of Codon Alignments. *Bioinformatics* (2005) 21 (10):2531–3. doi: 10.1093/bioinformatics/bti320
71. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. Datamonkey 2010: A Suite of Phylogenetic Analysis Tools for Evolutionary Biology. *Bioinformatics* (2010) 26(19):2455–7. doi: 10.1093/molbev/msx335
72. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol Biol Evol* (2018) 35(3):773–7. doi: 10.1093/bioinformatics/btq429
73. Jukes TH, Cantor CR. Evolution of Protein Molecules. In: Munro HN, editor. *Mammalian Protein Metabolism*. Academic Press, New York (1969). p. 21–132.
74. Ota T, Nei M. Variance and Covariances of the Numbers of Synonymous and Nonsynonymous Substitutions Per Site. *Mol Biol Evol* (1994) 11(4):613–9. doi: 10.1093/oxfordjournals.molbev.a040140
75. Nei M, Gojoborit T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol Biol Evol* (1986) 3(5):418–26. doi: 10.1093/oxfordjournals.molbev.a040410
76. Ngoc LV, Kassavetis GA, Kadonaga JT. The RNA Polymerase II Core Promoter in Drosophila. *Genetics* (2019) 212(1):13–24. doi: 10.1534/genetics.119.302021
77. Gershenzon NI, Trifonov EN, Ioshikhes IP. The Features of Drosophila Core Promoters Revealed by Statistical Analysis. *BMC Genomics* (2006) 7:161. doi: 10.1186/1471-2164-7-161
78. Smale ST, Baltimore D. The "Initiator" as a Transcription Control Element. *Cell* (1989) 57(1):103–13. doi: 10.1016/0092-8674(89)90176-1
79. Lažetić V, Troemel ER. Conservation Lost: Host-Pathogen Battles Drive Diversification and Expansion of Gene Families. *Fed Eur Biochem Soc J* (2021) 288(18):5289–99. doi: 10.1111/febs.15627
80. Yakovenko I, Donnyo A, Ioscovich O, Rosental B, Oren M. The Diverse Transformer (Trf) Protein Family in the Sea Urchin *Paracentrotus lividus* Acts

Through a Collaboration Between Cellular and Humoral Immune Effector Arms. *Int J Mol Sci* (2021) 22(13):6639.
81. Smith LC, Lun CM. The *SpTransformer* Gene Family (Formerly *Sp185/333*) in the Purple Sea Urchin and the Functional Diversity of the Anti-Pathogen rSp Transformer-E1 Protein. *Front Immunol* (2017) 8:725. doi: 10.3389/fimmu.2017.00725
82. Yinan W, Jun D, Yang LIU, Xuewei LIU, Yaqing C. Isolation of Immune-Relating 185/333-1 Gene From Sea Urchin (*Strongylocentrotus intermedius*) and Its Expression Analysis. *J Ocean Univ China* (2016) 15(1):163–70. doi: 10.1007/s11802-016-2707-4
83. Mongiardino Koch N, Coppard SE, Lessios HA, Briggs DEG, Mooi R, Rouse GW. A Phylogenomic Resolution of the Sea Urchin Tree of Life. *BMC Evol Biol* (2018) 18(1):189. doi: 10.1186/s12862-018-1300-4
84. Koch NM, Thompson JR. A Total-Evidence Dated Phylogeny of Echinoids and the Evolution of Body Size Across Adaptive Landscape. *bioRxiv* (2021) 70 (3):421–39. doi: 10.1101/2020.02.13.947796
85. Kober KM, Bernardi G. Phylogenomics of Strongylocentrotid Sea Urchins. *BMC Evol Biol* (2013) 13:88. doi: 10.1186/1471-2148-13-88
86. Kenny NJ, Sin YW, Hayward A, Paps J, Chu KH, Hui JHL. The Phylogenetic Utility and Functional Constraint of microRNA Flanking Sequences. *Proc R Soc B Biol Sci* (2015) 282(1803):20142983. doi: 10.1098/rspb.2014.2983
87. Smith LC. Diversification of Innate Immune Genes: Lessons From the Purple Sea Urchin. *Dis Model Mech* (2010) 3(5–6):274–9. doi: 10.1242/dmm.004697
88. Smith LC. Innate Immune Complexity in the Purple Sea Urchin: Diversity of the Sp185/333 System. *Front Immunol* (2012) 3:70. doi: 10.3389/fimmu.2012.00070

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

REVIEW
published: 17 November 2021
doi: 10.3389/fimmu.2021.784484

Structural Evolution of TIR-Domain Signalosomes

*Surekha Nimma, Weixi Gu, Natsumi Maruta, Yan Li, Mengqi Pan, Forhad Karim Saikot, Bryan Y. J. Lim, Helen Ying McGuinness, Zannati Ferdous Zaoti, Sulin Li, Sneha Desa, Mohammad Kawsar Manik, Jeffrey D. Nanson and Bostjan Kobe**

School of Chemistry and Molecular Biosciences, Institute for Molecular Bioscience and Australian Infectious Diseases Research Centre, University of Queensland, Brisbane, QLD, Australia

TIR (Toll/interleukin-1 receptor/resistance protein) domains are cytoplasmic domains widely found in animals and plants, where they are essential components of the innate immune system. A key feature of TIR-domain function in signaling is weak and transient self-association and association with other TIR domains. An additional new role of TIR domains as catalytic enzymes has been established with the recent discovery of NAD^+-nucleosidase activity by several TIR domains, mostly involved in cell-death pathways. Although self-association of TIR domains is necessary in both cases, the functional specificity of TIR domains is related in part to the nature of the TIR : TIR interactions in the respective signalosomes. Here, we review the well-studied TIR domain-containing proteins involved in eukaryotic immunity, focusing on the structures, interactions and their corresponding functional roles. Structurally, the signalosomes fall into two separate groups, the scaffold and enzyme TIR-domain assemblies, both of which feature open-ended complexes with two strands of TIR domains, but differ in the orientation of the two strands. We compare and contrast how TIR domains assemble and signal through distinct scaffolding and enzymatic roles, ultimately leading to distinct cellular innate-immunity and cell-death outcomes.

Keywords: protein structure, protein-protein interactions, axon degeneration, cell-death signaling, signaling by cooperative assembly formation (SCAF), innate immunity, plant disease resistance, toll/interleukin-1 receptor/resistance protein

OPEN ACCESS

Edited by:
Pedro Jose Esteves, Centro de Investigacao em Biodiversidade e Recursos Geneticos (CIBIO-InBIO), Portugal

Reviewed by:
Iva Hafner Bratkovic, National Institute of Chemistry, Slovenia
Takaki Maekawa, Max Planck Institute for Plant Breeding Research, Germany

***Correspondence:**
Bostjan Kobe
b.kobe@uq.edu.au

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *27 September 2021*
Accepted: *25 October 2021*
Published: *17 November 2021*

Citation:
Nimma S, Gu W, Maruta N, Li Y, Pan M, Saikot FK, Lim BYJ, McGuinness HY, Zaoti ZF, Li S, Desa S, Manik MK, Nanson JD and Kobe B (2021) Structural Evolution of TIR-Domain Signalosomes. Front. Immunol. 12:784484. doi: 10.3389/fimmu.2021.784484

INTRODUCTION

TIR (Toll/interleukin-1 receptor/resistance protein) domains are cytoplasmic domains found in both eukaryotic and prokaryotic proteins that are involved in innate-immunity and cell-death pathways. They consist of 135–160 residues and typically display a five-stranded parallel β-sheet (strands βA–βE) surrounded by five α-helices (αA–αE) (**Figure 1**) (1, 2). TIR-domain functions are governed by weak and transient interactions. They predominantly function through homotypic interactions, including self-association or association with other TIR domains, to create scaffolds that facilitate signal transduction, leading to immune and cell-death responses (3, 4). The mechanism of signaling employed has been described as SCAF (signaling by cooperative assembly formation) (5–7). SCAF involves the assembly of higher-order complexes - signalosomes or "supramolecular organizing centers" (SMOCs) (8). In the case of SCAF, receptor activation through activating ligand binding induces receptor oligomerization, which in turn

nucleates recruitment and oligomerization of adaptor proteins, and subsequently the recruitment and oligomerization of effector enzymes that can be activated through proximity-induced mechanisms in the resulting signalosome.

In mammals, TIR domains are found in Toll-like receptors (TLRs), interleukin-1 receptors (IL-1Rs) and cytoplasmic adaptor proteins, such as MyD88 (myeloid differentiation primary response gene 88) and MAL (MyD88 adaptor-like protein) (**Figures 1**, **2**) (2). TLRs and IL-1Rs are pattern recognition receptors (PRRs) that recognize evolutionarily conserved pathogen-associated molecular patterns (PAMPs) and endogenous danger-associated molecular patterns (DAMPs) released by dying or damaged cells (3, 4). Upon activation, TLR TIR domains dimerize, creating an intracellular TIR-domain signaling scaffold, which then recruits TIR domain-containing adaptor proteins that activate further downstream signaling (e.g. recruiting IL-1R-associated kinases (IRAKs) and activating the transcription factor NF-κB [nuclear factor kappa-light-chain-enhancer of activated B cells]), to induce inflammatory responses through the production of proinflammatory cytokines and programmed cell-death responses (3, 4).

In contrast to the scaffolding protein-protein interaction function described above, the TIR domain-containing protein SARM1 (sterile alpha and TIR motif containing 1) has been found to cleave NAD^+ (nicotinamide adenine dinucleotide) and $NADP^+$ (nicotinamide adenine dinucleotide phosphate) and initiate axon degeneration, in a process also dependent on the self-assembly of SARM1 TIR domains (**Figure 2**) (9–11). The dual role of TIR domains, serving as scaffolds or as enzymes, relies in part on the different assembly mechanisms of the proteins, leading to functional specificity.

Similar to mammals, plants also have a complex immune system to protect themselves from pathogen invasion. To restrict pathogen infection, the intracellular innate immune receptors called NLRs (nucleotide-binding domain (NBD)/leucine-rich repeat (LRR) receptors) directly or indirectly detect pathogen effector proteins *via* their C-terminal LRR domains and activate defense responses, including localized programmed cell death, in

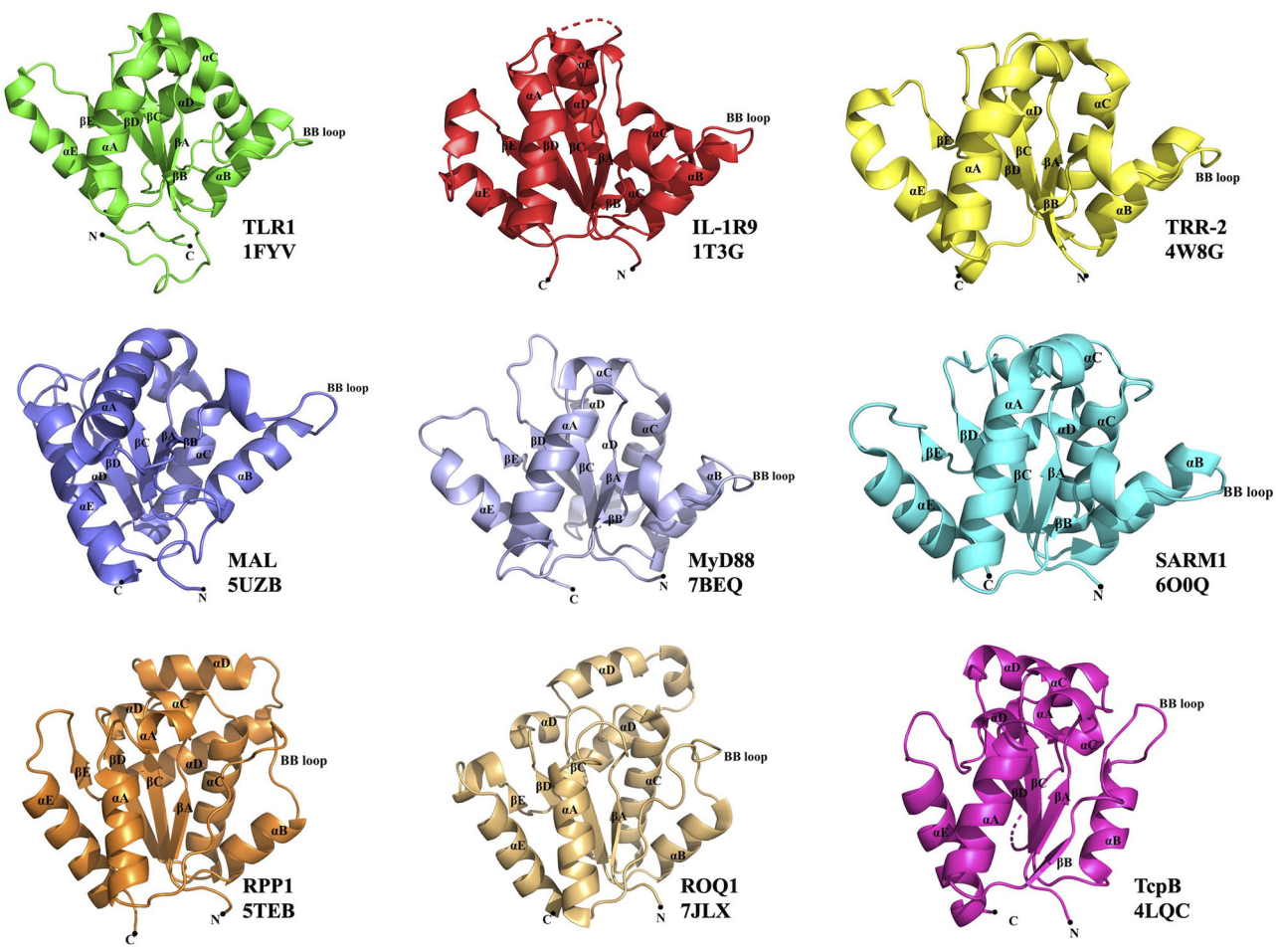


FIGURE 1 | Representative TIR domain structures. TIR domains of the human (TLR1, IL-1R9, MAL, MyD88 and SARM1), lower metazoan *Hydra magnipapillata* (TRR-2), plant (RPP1 and ROQ1) and bacterial (TcpB) proteins with their corresponding PDB IDs are shown. All the TIR domains show a central core of five β-strands (βA–βE) surrounded by five α-helices (αA–αE). The functionally important BB-loop in each TIR is labeled.

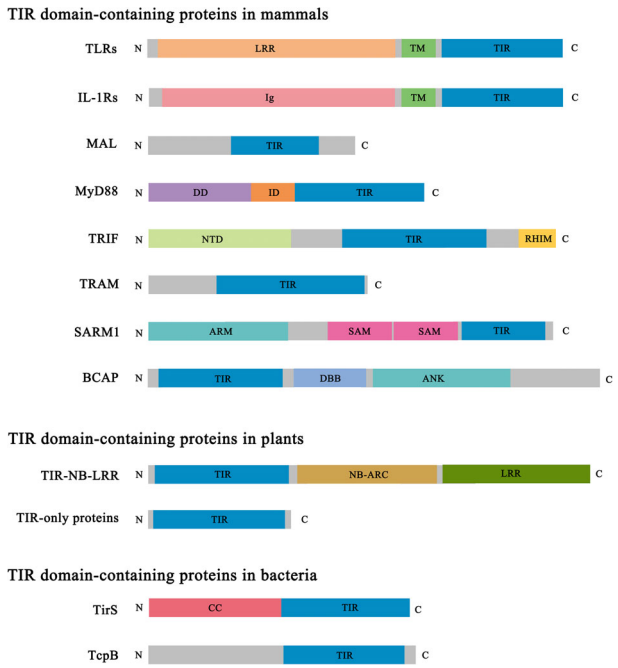


FIGURE 2 | Domain architecture of representative TIR domain-containing proteins. TIR, Toll/interleukin-1 receptor/resistance protein; LRR, leucine-rich repeat; TM, transmembrane; Ig, immunoglobulin; DD, death domain; ID, intermediate domain; NTD, N-terminal domain; RHIM, RIP (receptor-interacting protein) homotypic interaction motif; ARM, armadillo-repeat motif; SAM, sterile alpha motif; DBB, Dof (Drosophila downstream of fibroblast growth factor receptor)/BCAP/BANK (B cell scaffold protein with ankyrin repeats); ANK, ankyrin repeat; NB-ARC, nucleotide-binding adaptor shared by APAF-1 (apoptotic protease-activating factor 1), R proteins and CED-4 (cell death protein 4); CC, coiled coil.

a process termed hypersensitive response (HR) (**Figure 2**) (12). The central NBD (usually called NB-ARC – see **Figure 2**) is important for the oligomerization of these proteins into a signalosome called the "resistosome" upon activation. TIR domains are found at the N-termini of a large group of NLRs (TIR domain-containing NLRs, TNLs), as well as TIR-only proteins, and truncated NLRs lacking LRR domains (**Figure 2**) (12, 13). It is unclear if some plant TIR domains serve scaffolding functions like the canonical mammalian TIR domains; however, plant TIR domains self-associate and display the enzymatic function of cleaving NAD^+, sharing similarities with the function of SARM1 (10, 14). In agreement, plant TIR domains exhibit similar assembly mechanisms as SARM1 (10, 15, 16).

TIR domains are also found in bacteria and archaea. Structures and functions of bacterial and archaeal TIR domains and the corresponding proteins are poorly characterized, compared to their animal and plant counterparts, and are not the focus of the current review. Bacterial TIR domains are found in a wide range of domain architectures and domain types, which indicates diverse functional roles (**Figure 2**) (17). Some bacterial TIR domains, including TcpC (*Escherichia coli*), TirS (*Staphylococcus aureus*), PumA (*Pseudomonas aeruginosa*) and TcpB (*Brucella melitensis*), are linked to bacterial pathogenicity (18, 19). TirS and TcpC possess NAD^+-nucleosidase activity (20), but so do bacterial TIR domains from non-pathogenic bacteria, suggesting roles in bacterial physiology (20, 21). Recent studies have also shown that the NAD^+-nucleosidase activity of the bacterial TIR domains is linked to bacterial antiviral defenses (22–24).

Here, we review structural information available for various TIR domains, focusing on the nature of the different assemblies they form and their corresponding functional roles. Based on the available structural and functional evidence, we observe a correlation of TIR-domain assembly with their specific functional roles. In this respect, they fall into two different groups, the "scaffold" assemblies involved in innate-immunity signaling, and "enzyme" assemblies leading to NAD^+ cleavage associated with cell-death signaling.

SCAFFOLD TIR-DOMAIN ASSEMBLIES

This group comprises the TIR domains that undergo self-association to form a scaffold, which facilitates nucleation-controlled cooperative recruitment of other TIR domain-containing proteins and signal transduction (employing a SCAF mechanism) (5–7, 25). The group includes TIR domains from mammalian membrane receptors (TLRs and IL-1Rs), as well as those from the cytoplasmic adaptor proteins. A number of crystal structures of TIR domains from this group have been determined, but the structural basis of their self-assembly only became clear through the structural studies of higher-order structures reconstituted for the adaptors MAL and MyD88 (26, 27).

TLRs and IL-1Rs

Ten TLRs are present in humans and found either on the cell surface (TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10) or in intracellular endosomal compartments (TLR3, TLR4, TLR7, TLR8 and TLR9) (28). TLRs are characterized by an extra-cytoplasmic LRR domain, a transmembrane domain, and an intracellular TIR domain (**Figure 2**). When activated by their ligands, they mostly function as homodimers, but TLR2 functions as a heterodimer with either TLR1 or TLR6. A considerable amount of structural information is available on their LRR domains (29, 30), and even on the full-length TLR3 and TLR7 in complex with the membrane chaperone UNC93B1 involved in TLR trafficking (although the TIR domains could not be visualized in this case) (31). Crystal structures of the TIR domains of TLR1 (1), TLR2 (1, 32), TLR6 (33) and TLR10 (34) have been reported. These TIR domains, when expressed as separate proteins, are all monomeric in solution under the conditions tested. The characteristic BB-loop (connecting the βB strand and the αB helix) has been shown to play an important role in signaling, as a naturally occurring mutation (P712H) in this loop in TLR4 makes it non-responsive to the PAMP lipopolysaccharide (LPS) (35). Other mutations in this loop were also shown to abolish signaling in numerous TIR

domains, including TLR4 (36) and TLR7 (37). Downstream signaling involves the recruitment of the cytoplasmic adaptors MyD88 (for all TLRs except TLR3), MAL (as a bridging adaptor for MyD88 in the case of TLR4 and TLR1/2/6), TRIF (TIR domain-containing adaptor protein-inducing interferon β; for TLR3) and TRAM (TRIF-related adaptor molecule; as a bridging adaptor for TRIF in the case of TLR4).

There are also ten IL-1Rs (IL-1R1 to IL-1R10) in humans, characterized by extracellular immunoglobulin domains, a transmembrane domain, and an intracellular TIR domain (except for IL-1R2) (38) (**Figure 2**). IL-1Rs recognize and bind specific IL-1 family cytokines, which leads to recruitment of an accessory receptor chain. The downstream signaling mechanism is similar to that of TLRs, involving recruitment of MyD88, IRAKs and activating NF-κB to induce inflammatory responses (39). The IL-1R TIR domains are structurally less well characterized, with the majority of the available structural data limited to the complexes involving the ectodomains and their corresponding ligands (40). The only available structure of a TIR domain corresponds to that from IL-1R9 (IL-1RAPL1, IL-1R accessory protein like 1) (41). However, IL-1R9 is not a classical signaling IL-1R; it does not activate NF-κB, but has been reported to be involved in trans-synaptic signaling (42).

Cytoplasmic TLR Adaptor Proteins

TLRs and IL-1Rs require adaptor proteins for signaling. Six adaptor proteins have been identified in humans: MyD88, MAL, TRIF, TRAM, SARM1 and BCAP (B-cell adaptor for phosphoinositide 3-kinase) (28) (**Figure 2**). MyD88, MAL, TRIF and TRAM are the principal signaling adaptors in this pathway (43), while SARM1 and BCAP have been reported to negatively regulate TLR signaling (44–46). Crystal and/or NMR structures are available for the TIR domains from all these proteins (2, 5, 10).

Structures of Higher-Order Assemblies of MAL and MyD88 TIR Domains

Reconstitution of higher-order assemblies of MAL and MyD88 TIR domains yielded filamentous and micro-crystalline complexes, respectively (26, 27). These reconstitution experiments correlated with the functional signaling pathway, as the TLR4 TIR domain seeded the assembly of MAL TIR domains, while MAL TIR domains seeded the assembly of MyD88 TIR domains. Structure determination of these higher-order assemblies, combined with mutagenesis and signaling assays, has provided clarity on the biologically relevant association of scaffold TIR domains (26, 27). These structures feature two parallel strands of TIR domains, each showcasing a head-to-tail arrangement of TIR domains, held together through a BB-loop-mediated "BE" intrastrand interface. The two strands are offset and held together through a "BCD" interstrand interface (**Figures 3**, **4**).

The structure of the filament formed by MAL TIR domains was determined by helical reconstruction cryo-electron microscopy (cryo-EM) (26). It revealed a hollow tube consisting of 12 protofilaments, with each protofilament corresponding to a two-stranded assembly of TIR domains described above. The intrastrand BE interface includes the area around the BB-loop of one subunit and the EE surface (the βD and βE strands and the αE helix) of the interacting subunit (**Figure 3A**). The interstrand interaction connecting the two strands involves the residues on the BC surface (αB and αC helices) of one subunit from strand 1 and the CD surface (αD helix and the CD loop) of another subunit from strand 2 (**Figure 3A**). Structure-based mutagenesis studies confirmed that key residues in both interfaces (e.g., P125A in the intrastrand interface; L162A, L165A, W156A, Y159A and F193A in the interstrand interface) are necessary for the activation of downstream signaling through NF-κB, whereas mutations of residues mediating interactions between protofilaments were not found to have consequences for function (26).

The structure of the MAL TIR domain-nucleated microcrystals of MyD88 TIR domain was determined using two approaches, microcrystal electron diffraction (microED) and serial femtosecond crystallography (SFX) with an X-ray free electron laser (X-FEL) source (both yielding nearly identical structures) (27). The structure shows that the TIR domains of MyD88 assemble in a fashion analogous to MAL TIR-domain protofilaments (**Figure 3A**). The interfaces responsible for this association were again found to be relevant to signaling, with the mutations R196A, W284A, I253D and R288A in the intrastrand interface, and mutations K238A, L241A, F270A and F270E in the interstrand interface abolishing TLR4-induced NF-κB activation (27). The analogous structural arrangement of TIR domains in the MAL and MyD88 higher-order assemblies suggests a hierarchical, nucleation-controlled and cooperative mechanism for TLR signal transduction, in which the receptor and adaptor TIR domains assemble *via* the inter- and intrastrand interactions observed in the MyD88 and MAL TIR-domain higher-order assemblies, leading to formation of a TIR-domain signalosome. This in turn promotes clustering of the MyD88 death domains (DDs) to form a signalosome termed the "Myddosome" (47), recruiting and activating IRAKs (26, 27), thereby facilitating signaling through a SCAF mechanism.

Several mutations in the TIR domains of TLRs and adaptors are associated with disease. In MyD88, the R196C polymorphism, which is associated with susceptibility to pyogenic bacterial infection during childhood (48), maps to the intrastrand interface of the TIR-domain signalosome. Similarly, the L252P gain-of-function variant of MyD88, which is found in diffuse large B cell lymphoma and promotes tumor survival through enhanced NF-κB activation (49), also maps to the intrastrand interface. The corresponding mutant forms extremely stable oligomers, compared to the wild-type protein, explaining the molecular basis of its phenotype (50).

Interestingly, in one of the crystal structures of the TRR-2 TIR domain from the lower metazoan *Hydra magnipapillata* (PDB ID 4W8G), a parallel two-strand arrangement analogous to MAL and MyD88 TIR-domain signalosomes is observed (**Figures 1**, **3A**). These signalosomes are therefore likely to be structurally conserved in a range of eukaryotes.

The interactions observed in the MyD88 and MAL TIR-domain signalosomes were not captured in the crystal structures of any of the mammalian TLR or adaptor TIR domains. However, surfaces equivalent to the ones mediating interstrand

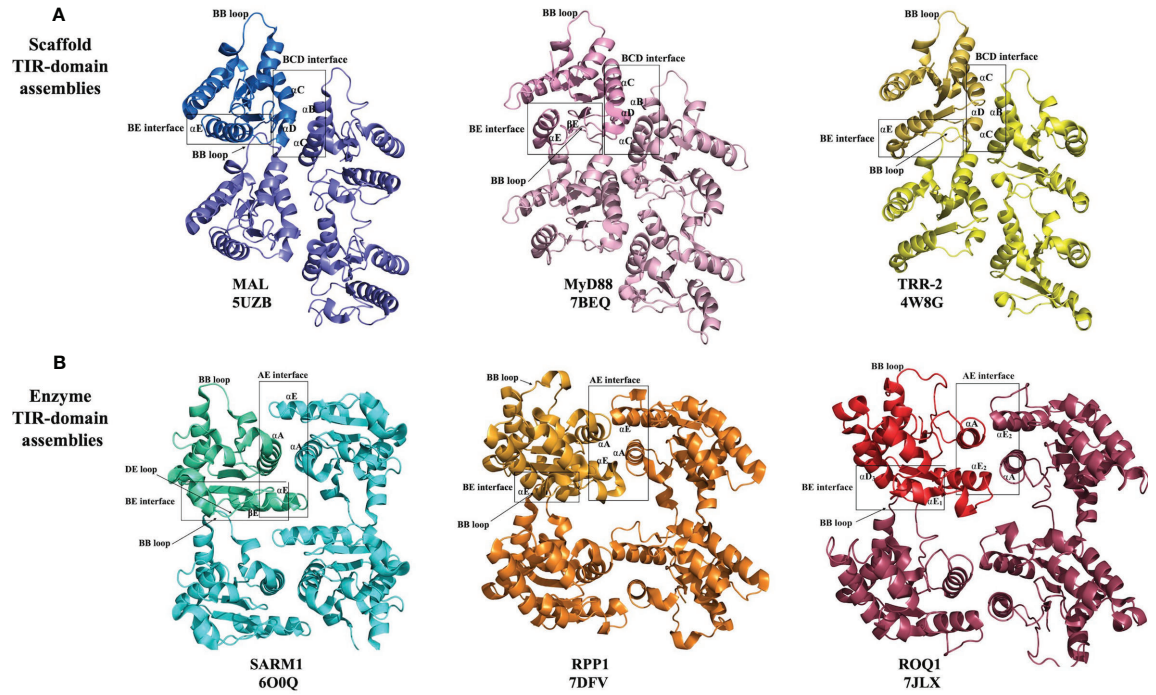


FIGURE 3 | Structural basis of eukaryotic TIR-domain assembly formation. **(A)** Scaffold TIR-domain assemblies, held together by the BE and BCD interfaces, represented by the TIR domains from MAL, MyD88 and TRR-2. **(B)** Enzyme TIR-domain assemblies held together by the BE and AE interfaces, represented by the TIR domains from SARM1, RPP1 and ROQ1.

interactions (involving residues in αB, αC and αD helices and the BB-loop; "BCD surface") are found to mediate symmetric dimer formation in the crystals of a number of TLR and IL-1R TIR domains (TLR1, TLR2, TLR6, TLR10 and IL-1R9) (5). While it is possible that functional surfaces simply find alternative, biologically irrelevant binding partners when forming crystals, it could be that such symmetric interaction contributes to the regulation of signaling by representing an inactive, auto-inhibited state, especially in the case of TLRs that always exist as dimers; such an interaction would presumably be broken when the receptor is activated by ligand binding and the signalosome formation would be allowed.

In summary, the scaffold TIR-domain signalosomes correspond to parallel two-stranded open-ended assemblies of TIR domains, held together by two asymmetric interactions, the intrastrand and interstrand interactions, mediated by the BE and BCD interfaces, respectively (**Figures 3**, **4**). The nucleation-controlled cooperative assembly of these signalosomes is responsible for the SCAF mechanism of signaling.

ENZYME TIR-DOMAIN ASSEMBLIES

This group comprises the TIR domains that undergo self-association, which facilitates NAD^+-nucleosidase activity and eventually cell death. It includes TIR domains from the mammalian protein SARM1 and plant TIR domain-containing proteins. While some bacterial and archaeal TIR domains have also been shown to have NAD^+-nucleosidase activity, the structural basis of their self-association and enzymatic activity is not well characterized and may be different from their eukaryotic counterparts discussed in this group (10, 20). On the other hand, the TIR domains from SARM1 and plant NLRs assemble in an analogous fashion, forming two-stranded assemblies different from that of the scaffold TIR-domain assemblies (**Figures 3**, **4**). The association of TIR domains in each individual strand resembles the one observed in scaffold TIR-domain signalosomes, featuring a head-to-tail arrangement mediated by the BB-loop-containing BE interface. However, the two strands in enzyme TIR-domain assemblies are associated in an antiparallel, rather than parallel, fashion. The interstrand interface corresponds to a symmetric "AE interface", involving the αA and αE helices. This structural information is based on three key structures: the crystal structure of the SARM1 TIR domain (10), and the cryo-EM structures of activated resistosome complexes of the TNLs ROQ1 and RPP1 (15, 16).

SARM1

While SARM1 has a number of suggested roles in the regulation of innate immunity, the central function appears to be to serve as the executioner of Wallerian or programmed axon degeneration, a highly conserved pathway of injury-induced axon degeneration (51–54). SARM1 facilitates rapid depletion of NAD^+ in response to axon injury, leading to subsequent axon demise (10, 11, 55).

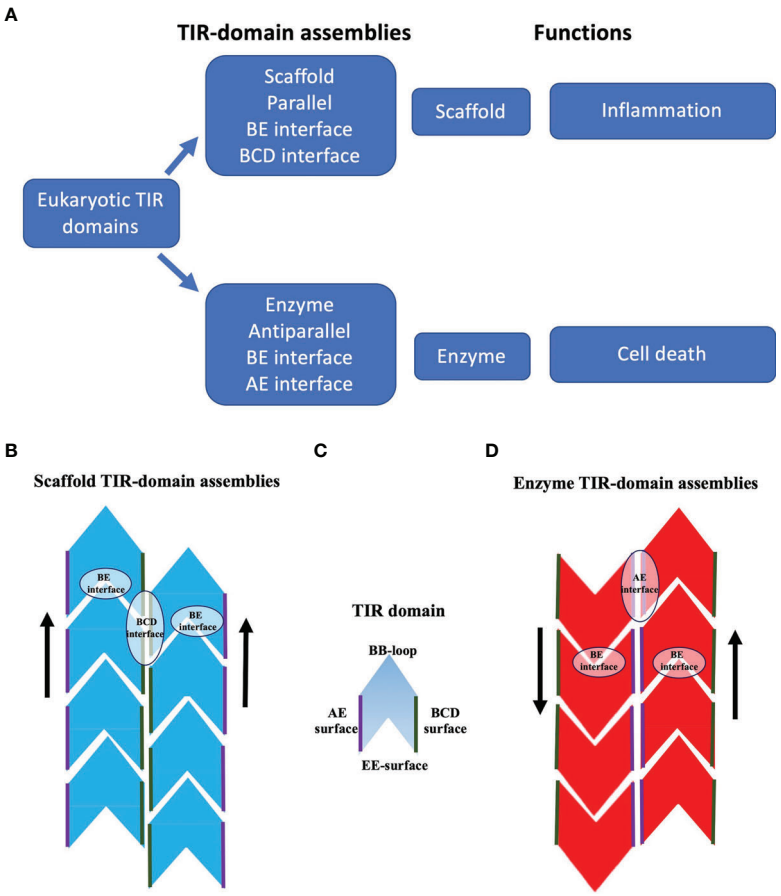


FIGURE 4 | Schematic representation of TIR-domain assemblies. **(A)** Overview of TIR-domain assemblies. **(B)** Scaffold TIR-domain assemblies, represented by the structures formed by the TIR domains from human MAL and MyD88. These parallel two-stranded assemblies are held together by the intrastrand BE interfaces, and the interstrand BCD interfaces. **(C)** Schematic diagram of a single TIR domain, to highlight the BB-loop, BCD, EE and AE surfaces (the diagram does not differentiate the structures of TIR domains in monomeric and signalosome forms). **(D)** Enzyme TIR-domain assemblies, represented by the structures formed by the TIR domains from human SARM1 and the plant NLRs ROQ1 and RPP1. These antiparallel two-stranded assemblies are held together by the intrastrand BE interfaces and the interstrand AE interfaces.

SARM1 deletion mutants lacking the TIR domain show a dominant-negative phenotype, delaying axon degeneration (52). In addition to its C-terminal TIR domain, SARM1 contains an N-terminal ARM (armadillo repeat motif) domain and central tandem SAM (sterile alpha motif) domains. SAM domains form an octameric ring, and the formation of such an oligomeric state is essential for axon degeneration (10, 52, 56). The SARM1 TIR domain has intrinsic NAD^+-nucleosidase activity, cleaving NAD^+ into nicotinamide and either ADPR (ADP-ribose) or cyclic ADPR (cADPR) (10, 11, 57). The active site has similarities to that of the NAD^+ glycohydrolase CD38, including a catalytic glutamate residue (10, 11). Mutation of this residue is sufficient to abolish NAD^+-nucleosidase activity, indicating similar mechanisms of NAD^+-nucleosidase activity among these enzymes (10, 11). Self-association of SARM1 TIR domains is essential for enzyme activity (10, 11). The crystal structure of the SARM1 TIR domain revealed the two-stranded antiparallel arrangement described above, and mutational analysis of the BE (D594A, E596K, and G601P) and AE interfaces (L579A and H685A) revealed that these interactions are functionally important for stabilization of the active conformation of the TIR domains (10).

In the inactive SARM1 octamer, interaction of the ARM and TIR domains is responsible for preventing TIR-domain self-association, directly inhibiting NAD^+-nucleosidase activity (58–60). The ARM-TIR inhibitory interaction is regulated by the cellular ratios of NAD^+ and its metabolites NMN (nicotinamide mononucleotide) and NaMN (nicotinic acid mononucleotide), through binding to an allosteric site in the ARM domain; NMN as an activator, and NAD^+ and NaMN as inhibitors (59–63). Injury-associated increase in NMN results in the release of this ARM-TIR autoinhibition and subsequent assembly of TIR domains (59).

Plant TIR Domains

TIR domains from several plant NLRs and TIR-only plant proteins can also cleave NAD^+, dependent on the TIR domain

self-association and the conserved glutamate residue at the catalytic site; this activity is essential for cell-death signaling (10, 14). Crystal structures of numerous plant TIR domains [including those from RPS4 and RPS4-RRS1 heterodimer, AtTIR, SNC1, RPP1 and RPV1 (64–67)] feature a symmetrical AE interface analogous to the one observed in SARM1, and mutagenesis of residues in this interface was shown to abolish cell-death signaling in plants (10, 65–68). The AE interface often features conserved residues, such as the SH (serine-histidine) motif in the αA helices of the TIR domains, which allows stacking and hydrogen-bonding interactions across the interface (65). Charged residues surrounding the SH motif further stabilize the AE interface (67). However, the structure of a functional TIR-domain assembly only became clear based on the cryo-EM structures of ROQ1 and RPP1 NLRs (15, 16).

Activated ROQ1 and RPP1 form tetrameric structures, largely through the interactions of their NBDs (15, 16). This NBD-mediated tetramer exhibits four-fold symmetry; however, the TIR domains arranged on top of this tetramer exhibit only two-fold symmetry, featuring a dimer of dimers. The arrangement of TIR domains is identical to the arrangement of SARM1 TIR domains observed in the corresponding crystals. Therefore, each dimer features the symmetrical AE interface, but the two dimers are held together through BE interfaces (**Figure 3B**). Such a BE interface was never observed in any of the crystal structures of plant TIR domains; however, the DE surface mediating this interaction has been implicated previously in the function of these domains through site-directed mutagenesis (67–69).

The BE interface-mediated interaction is crucial for NAD^+-nucleosidase activity, because in the absence of this interaction, the positively charged lysine and arginine residues in the BB-loop block NAD^+ binding to the active site. In the presence of BE-interface interaction, the BB-loop moves under the DE surface and allows access by the substrate to the NAD^+-binding site. Jointly, the AE and BE interface-mediated interactions facilitate the appropriate conformation of the active site to enable enzymatic activity (16). The oligomerization of plant NLRs, driven largely by the NBDs, is presumed to nucleate assembly of the TIR domains.

In summary, the enzyme TIR-domain signalosomes involve anti parallel head-to-tail asymmetric intrastrand and symmetric interstrand interactions, mediated by the BE and AE interfaces, respectively (**Figures 3**, **4**). Such an assembly is required for configuring the NAD^+-nucleosidase active site, and consequently enzymatic activity and biological function (axon degeneration or cell death).

SIGNALING BY COOPERATIVE ASSEMBLY FORMATION

Both the scaffold and enzyme types of TIR-domain signalosomes function through a signaling mechanism termed SCAF (5–7). Compared to a more gradual signal amplification in a classical signaling pathway, SCAF enables a rapid and strong response to minute amounts of stimulus, resembling a switch; it is therefore well suited to innate-immunity and cell-death pathways. An important aspect of regulation involves the nucleation barriers to oligomerization (25). In the case of scaffold TIR-domain assemblies, the effector enzymes correspond to proteins kinases (IRAKs), which can activate themselves through phosphorylation when brought together. In the case of enzyme TIR-domain assemblies, the TIR domains themselves serve as effector enzymes, as they require self-association for activity. In both scaffold and enzyme groups, other domains play key roles to create a functional pathway, for example LRR domains as receptors (also called sensors) in TLRs and plant NLRs, ARM domains as receptors in SARM1, and DDs and NBDs as adaptors in TLR and plant NLR pathways. The presence of more than one self-associating domain provides further opportunities for regulation of the system (e.g. auto-inhibition between the DD and TIR domains in MyD88), which ultimately creates a large range of concentrations where the system is poised for activation (50, 70, 71).

CONCLUSIONS AND FUTURE DIRECTIONS

TIR domains are found in animals, plants, bacteria and archaea, and often have functions associated with innate immunity and cell death. Self-association of TIR domains is key to their function, but weak affinities prevent association until aided by activating ligands, other domains in TIR domain-containing proteins, or adaptor proteins. The differences in the structures of scaffold and enzyme TIR-domain assemblies correlate with their distinct functional roles. Scaffold TIR-domain signalosomes, represented by MAL and MyD88 TIR-domain complexes, correspond to parallel two-stranded assemblies that are used in nucleation-controlled pathways that are activated by PAMPs or DAMPs and result in the activation of protein kinases (such as IRAKs). Enzyme TIR-domain signalosomes, represented by TIR-domain complexes from SARM1 and plant NLRs, correspond to antiparallel two-stranded assemblies with NAD^+-nucleosidase activity that form in response to NMN or plant pathogen effector protein binding, respectively. Both groups of assemblies feature an analogous head-to-tail arrangement of TIR domains in individual strands, mediated by the BE interface, but differ in their interstrand association. In line with their analogous interstrand interactions, the functionally important BB-loop adopts a similar conformation in the assembled state in both cases.

Our classification into the two groups of TIR-domain signalosomes is based on a limited number of signalosome structures, and further work will be required to establish how generally applicable the current models are to TIR-domain signaling, and what variations exist in different pathways. However, in support of our general conclusions, molecular, mutational, and functional data on a large variety of TIR domains are consistent with our proposed models. For example, the TIR domain of the TLR adaptor TRAM only has limited sequence identity to those from MAL and MyD88, yet it is compatible with filaments with the same arrangement of TIR domains as seen in the

latter two cases (26). Similarly, the *Arabidopsis* plant TIR-only immune receptor RBA1 and the TIR-only protein BdTIR of unknown function from the monocot plant *Brachypodium distachyon* display self-association-dependent NAD^+-nucleosidase activity (14). Nevertheless, further studies on different TIR domain-containing proteins will be required to understand the breadth of similarities and differences in their molecular basis of function, in particular the less characterized proteins from bacteria and archaea. Another important but poorly characterized aspect involves the molecular details of the reaction catalyzed by different enzyme TIR domains. While all characterized enzyme TIR domains cleave the nicotinamide moiety off the substrate NAD^+ molecule, many produce a cyclic variant of the remaining ADPR, rather than just the linear ADPR molecule. Human SARM1 produces a small proportion of the classical cADPR, but several bacterial and plant TIR domains instead produce variants with a different cyclic linkage (11, 14, 20). The chemical structures of these variants remain uncharacterized, as do their roles in the corresponding biological pathways.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by funding from the National Health and Medical Research Council (NHMRC) (Project Grant 1160570 to BK) and the Australian Research Council (ARC) (Discovery Project DP190102526 and Laureate Fellowship FL180100109 to BK).

ACKNOWLEDGMENTS

The authors thank Dr Gayle Petersen for proofreading, editing, and constructive criticism of the manuscript.

REFERENCES

1. Xu Y, Tao X, Shen B, Horng T, Medzhitov R, Manley JL, et al. Structural Basis for Signal Transduction by the Toll/interleukin-1 Receptor Domains. *Nature* (2000) 408:111–5. doi: 10.1038/35040600
2. Ve T, Williams SJ, Kobe B. Structure and Function of Toll/interleukin-1 Receptor/Resistance Protein (TIR) Domains. *Apoptosis* (2015) 20:250–61. doi: 10.1007/s10495-014-1064-2
3. Dinarello CA. Interleukin-1 in the Pathogenesis and Treatment of Inflammatory Diseases. *Blood* (2011) 117:3720–32. doi: 10.1182/blood-2010-07-273417
4. Kawai T, Akira S. The Role of Pattern-Recognition Receptors in Innate Immunity: Update on Toll-Like Receptors. *Nat Immunol* (2010) 11:373–84. doi: 10.1038/ni.1863
5. Nimma S, Ve T, Williams SJ, Kobe B. Towards the Structure of the TIR-Domain Signalosome. *Curr Opin Struct Biol* (2017) 43:122–30. doi: 10.1016/j.sbi.2016.12.014
6. Vajjhala PR, Ve T, Bentham A, Stacey KJ, Kobe B. The Molecular Mechanisms of Signaling by Cooperative Assembly Formation in Innate Immunity Pathways. *Mol Immunol* (2017) 86:23–37. doi: 10.1016/j.molimm.2017.02.012
7. Nanson JD, Kobe B, Ve T. Death TIR, and RHIM: Self-Assembling Domains Involved in Innate Immunity and Cell-Death Signaling. *J Leukoc Biol* (2019) 105:363–75. doi: 10.1002/JLB.MR0318-123R
8. Kagan JC, Magupalli VG, Wu H. SMOCs: Supramolecular Organizing Centres That Control Innate Immunity. *Nat Rev Immunol* (2014) 14:821–6. doi: 10.1038/nri3757
9. Zhao ZY, Xie XJ, Li WH, Liu J, Chen Z, Zhang B, et al. A Cell-Permeant Mimetic of NMN Activates SARM1 to Produce Cyclic ADP-Ribose and Induce Non-Apoptotic Cell Death. *iScience* (2019) 15:452–66. doi: 10.1016/j.isci.2019.05.001
10. Horsefield S, Burdett H, Zhang X, Manik MK, Shi Y, Chen J, et al. NAD(+) Cleavage Activity by Animal and Plant TIR Domains in Cell Death Pathways. *Science* (2019) 365:793–9. doi: 10.1126/science.aax1911
11. Essuman K, Summers DW, Sasaki Y, Mao X, DiAntonio A, Milbrandt J. The SARM1 Toll/Interleukin-1 Receptor Domain Possesses Intrinsic NAD+ Cleavage Activity That Promotes Pathological Axonal Degeneration. *Neuron* (2017) 93:1334–43.e1335. doi: 10.1016/j.neuron.2017.02.022
12. Tamborski J, Krasileva KV. Evolution of Plant NLRs: From Natural History to Precise Modifications. *Annu Rev Plant Biol* (2020) 71:355–78. doi: 10.1146/annurev-arplant-081519-035901
13. Meyers BC, Morgante M, Michelmore RW. TIR-X and TIR-NBS Proteins: Two New Families Related to Disease Resistance TIR-NBS-LRR Proteins Encoded in Arabidopsis and Other Plant Genomes. *Plant J* (2002) 32:77–92. doi: 10.1046/j.1365-313X.2002.01404.x
14. Wan L, Essuman K, Anderson RG, Sasaki Y, Monteiro F, Chung E-H, et al. TIR Domains of Plant Immune Receptors Are NAD+-Cleaving Enzymes That Promote Cell Death. *Science* (2019) 365:799–803. doi: 10.1126/science.aax1771
15. Ma S, Lapin D, Liu L, Sun Y, Song W, Zhang X, et al. Direct Pathogen-Induced Assembly of an NLR Immune Receptor Complex to Form a Holoenzyme. *Science* (2020) 370(6521):eabe3069. doi: 10.1126/science.abe3069
16. Martin R, Qi T, Zhang H, Liu F, King M, Toth C, et al. Structure of the Activated ROQ1 Resistosome Directly Recognizing the Pathogen Effector XopQ. *Science* (2020) 370(6521):eabd9993. doi: 10.1126/science.abd9993
17. Spear AM, Loman NJ, Atkins HS, Pallen MJ. Microbial TIR Domains: Not Necessarily Agents of Subversion? *Trends Microbiol* (2009) 17:393–8. doi: 10.1016/j.tim.2009.06.005
18. Sengupta D, Koblansky A, Gaines J, Brown T, West AP, Zhang D, et al. Subversion of Innate Immune Responses by Brucella Through the Targeted Degradation of the TLR Signaling Adapter, MAL. *J Immunol* (2010) 184:956–64. doi: 10.4049/jimmunol.0902008
19. Snyder GA, Deredge D, Waldhuber A, Fresquez T, Wilkins DZ, Smith PT, et al. Crystal Structures of the Toll/Interleukin-1 Receptor (TIR) Domains From the Brucella Protein TcpB and Host Adaptor TIRAP Reveal Mechanisms of Molecular Mimicry. *J Biol Chem* (2014) 289:669–79. doi: 10.1074/jbc.M113.523407
20. Essuman K, Summers DW, Sasaki Y, Mao X, Yim AKY, DiAntonio A, et al. TIR Domain Proteins Are an Ancient Family of NAD+-Consuming Enzymes. *Curr Biol* (2018) 28:421–30.e424. doi: 10.1016/j.cub.2017.12.024
21. Coronas-Serna JM, Louche A, Rodríguez-Escudero M, Roussin M, Imbert PR, Rodríguez-Escudero I, et al. The TIR-Domain Containing Effectors BtpA and BtpB From Brucella Abortus Impact NAD Metabolism. *PloS Pathog* (2020) 16:e1007979. doi: 10.1371/journal.ppat.1007979
22. Ka D, Oh H, Park E, Kim J-H, Bae E. Structural and Functional Evidence of Bacterial Antiphage Protection by Thoeris Defense System *via* NAD+ Degradation. *Nat Comm* (2020) 11:1–8. doi: 10.1038/s41467-020-16703-w
23. Doron S, Melamed S, Ofir G, Leavitt A, Lopatina A, Keren M, et al. Systematic Discovery of Antiphage Defense Systems in the Microbial Pangenome. *Science* (2018) 359(6739):eaar4120. doi: 10.1126/science.aar4120
24. Ofir G, Herbst E, Baroz M, Cohen D, Millman A, Doron S, et al. Antiviral Activity of Bacterial TIR Domains *via* Signaling Molecules That Trigger Cell Death. *bioRxiv* (2021). doi: 10.1101/2021.01.06.425286. 2021.2001.2006.425286.
25. Rodriguez Gama A, Miller T, Halfmann R. Mechanics of a Molecular Mousetrap-Nucleation-Limited Innate Immune Signaling. *Biophys J* (2021) 120:1150–60. doi: 10.1016/j.bpj.2021.01.007

26. Ve T, Vajjhala PR, Hedger A, Croll T, DiMaio F, Horsefield S, et al. Structural Basis of TIR-Domain-Assembly Formation in MAL- and MyD88-Dependent TLR4 Signaling. *Nat Struct Mol Biol* (2017) 24:743–51. doi: 10.1038/nsmb.3444
27. Clabbers MT, Holmes S, Muusse TW, Vajjhala PR, Thygesen SJ, Malde AK, et al. MyD88 TIR Domain Higher-Order Assembly Interactions Revealed by Microcrystal Electron Diffraction and Serial Femtosecond Crystallography. *Nat Comm* (2021) 12:1–14. doi: 10.1038/s41467-021-22590-6
28. Ve T, Gay NJ, Mansell A, Kobe B, Kellie S. Adaptors in Toll-Like Receptor Signaling and Their Potential as Therapeutic Targets. *Curr Drug Targets* (2012) 13:1360–74. doi: 10.2174/138945012803530260
29. Behzadi P, Garcia-Perdomo HA, Karpinski TM. Toll-Like Receptors: General Molecular and Structural Biology. *J Immunol Res* (2021) 2021:9914854. doi: 10.1155/2021/9914854
30. Asami J, Shimizu T. Structural and Functional Understanding of the Toll-Like Receptors. *Protein Sci* (2021) 30:761–72. doi: 10.1002/pro.4043
31. Ishida H, Asami J, Zhang Z, Nishizawa T, Shigematsu H, Ohto U, et al. Cryo-EM Structures of Toll-Like Receptors in Complex With UNC93B1. *Nat Struct Mol Biol* (2021) 28:173–80. doi: 10.1038/s41594-020-00542-w
32. Tao X, Xu Y, Zheng Y, Beg AA, Tong L. An Extensively Associated Dimer in the Structure of the C713S Mutant of the TIR Domain of Human TLR2. *Biochem Biophys Res Commun* (2002) 299:216–21. doi: 10.1016/s0006-291x(02)02581-0
33. Jang T-h, Park HH. Crystal Structure of TIR Domain of TLR6 Reveals Novel Dimeric Interface of TIR–TIR Interaction for Toll-Like Receptor Signaling Pathway. *J Mol Biol* (2014) 426:3305–13. doi: 10.1016/j.jmb.2014.07.024
34. Nyman T, Stenmark P, Flodin S, Johansson I, Hammarström M, Nordlund P. The Crystal Structure of the Human Toll-Like Receptor 10 Cytoplasmic Domain Reveals a Putative Signaling Dimer. *J Biol Chem* (2008) 283:11861–5. doi: 10.1074/jbc.C800001200
35. Poltorak A, He X, Smirnova I, Liu M-Y, Van Huffel C, Du X, et al. Defective LPS Signaling in C3H/HeJ and C57BL/10ScCr Mice: Mutations in Tlr4 Gene. *Science* (1998) 282:2085–8. doi: 10.1126/science.282.5396.2085
36. Nunez Miguel R, Wong J, Westoll JF, Brooks HJ, O'Neill LA, Gay NJ, et al. A Dimer of the Toll-Like Receptor 4 Cytoplasmic Domain Provides a Specific Scaffold for the Recruitment of Signalling Adaptor Proteins. *PloS One* (2007) 2:e788. doi: 10.1371/journal.pone.0000788
37. Lech M, Skuginna V, Kulkarni OP, Gong J, Wei T, Stark RW, et al. Lack of SIGIRR/TIR8 Aggravates Hydrocarbon Oil-Induced Lupus Nephritis. *J Pathol* (2010) 220:596–607. doi: 10.1002/path.2678
38. Boraschi D, Tagliabue A. The Interleukin-1 Receptor Family. *Semin Immunol* (2013) 25:394–407. doi: 10.1016/j.smim.2013.10.023
39. Boraschi D, Italiani P, Weil S, Martin MU. The Family of the Interleukin-1 Receptors. *Immunol Rev* (2018) 281:197–232. doi: 10.1111/imr.12606
40. Fields JK, Günther S, Sundberg EJ. Structural Basis of IL-1 Family Cytokine Signaling. *Front Immunol* (2019) 10:1412. doi: 10.3389/fimmu.2019.01412
41. Khan JA, Brint EK, O'Neill LA, Tong L. Crystal Structure of the Toll/interleukin-1 Receptor Domain of Human IL-1rapl. *J Biol Chem* (2004) 279:31664–70. doi: 10.1074/jbc.M403434200
42. Montani C, Gritti L, Beretta S, Verpelli C, Sala C. The Synaptic and Neuronal Functions of the X-Linked Intellectual Disability Protein Interleukin-1 Receptor Accessory Protein Like 1 (Il1rapl1). *Dev Neurobiol* (2019) 79:85–95. doi: 10.1002/dneu.22657
43. Gay NJ, Symmons MF, Gangloff M, Bryant CE. Assembly and Localization of Toll-Like Receptor Signalling Complexes. *Nat Rev Immunol* (2014) 14:546–58. doi: 10.1038/nri3713
44. Carty M, Goodbody R, Schröder M, Stack J, Moynagh PN, Bowie AG. The Human Adaptor SARM Negatively Regulates Adaptor Protein TRIF–dependent Toll-Like Receptor Signaling. *Nat Immunol* (2006) 7:1074–81. doi: 10.1038/ni1382
45. Halabi S, Sekine E, Verstak B, Gay NJ, Moncrieffe MC. Structure of the Toll/interleukin-1 Receptor (TIR) Domain of the B-Cell Adaptor That Links Phosphoinositide Metabolism With the Negative Regulation of the Toll-Like Receptor (TLR) Signalosome. *J Biol Chem* (2017) 292:652–60. doi: 10.1074/jbc.M116.761528
46. Troutman TD, Hu W, Fulenchek S, Yamazaki T, Kurosaki T, Bazan JF, et al. Role for B-Cell Adapter for PI3K (BCAP) as a Signaling Adapter Linking Toll-Like Receptors (TLRs) to Serine/Threonine Kinases PI3K/Akt. *Proc Natl Acad Sci USA* (2012) 109:273–8. doi: 10.1073/pnas.1118579109
47. Lin SC, Lo YC, Wu H. Helical Assembly in the MyD88-IRAK4-IRAK2 Complex in TLR/IL-1R Signalling. *Nature* (2010) 465:885–90. doi: 10.1038/nature09121
48. Von Bernuth H, Picard C, Jin Z, Pankla R, Xiao H, Ku C-L, et al. Pyogenic Bacterial Infections in Humans With MyD88 Deficiency. *Science* (2008) 321:691–6. doi: 10.1126/science.1158298
49. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim K-H, et al. Oncogenically Active MYD88 Mutations in Human Lymphoma. *Nature* (2011) 470:115–9. doi: 10.1038/nature09671
50. O'Carroll A, Chauvin B, Brown JWP, Meagher A, Coyle J, Schill J, et al. Pathological Mutations Differentially Affect the Self-Assembly and Polymerisation of the Innate Immune System Signalling Adaptor Molecule Myd88. *BMC Biol* (2018) 16:149. doi: 10.1186/s12915-018-0611-7
51. Osterloh JM, Yang J, Rooney TM, Fox AN, Adalbert R, Powell EH, et al. Dsarm/Sarm1 is Required for Activation of an Injury-Induced Axon Death Pathway. *Science* (2012) 337:481–4. doi: 10.1126/science.1223899
52. Gerdts J, Summers DW, Sasaki Y, DiAntonio A, Milbrandt J. Sarm1-Mediated Axon Degeneration Requires Both SAM and TIR Interactions. *J Neurosci* (2013) 33:13569–80. doi: 10.1523/JNEUROSCI.1197-13.2013
53. Hopkins EL, Gu W, Kobe B, Coleman MP. A Novel NAD Signaling Mechanism in Axon Degeneration and Its Relationship to Innate Immunity. *Front Mol Biosci* (2021) 8:703532. doi: 10.3389/fmolb.2021.703532
54. Carty M, Bowie AG. SARM: From Immune Regulator to Cell Executioner. *Biochem Pharmacol* (2019) 161:52–62. doi: 10.1016/j.bcp.2019.01.005
55. Angeletti C, Amici A, Gilley J, Loreto A, Trapanotto AG, Antoniou C, et al. Programmed Axon Death Executor SARM1 is a Multi-Functional NAD (P) Ase With Prominent Base Exchange Activity, All Regulated by Physiological Levels of NMN, NAD, NADP and Other Metabolites. *bioRxiv* (2021). doi: 10.1101/2021.07.14.451805. 2021.07.14.451805.
56. Sporny M, Guez-Haddad J, Lebendiker M, Ulisse V, Volf A, Mim C, et al. Structural Evidence for an Octameric Ring Arrangement of SARM1. *J Mol Biol* (2019) 431:3591–605. doi: 10.1016/j.jmb.2019.06.030
57. Summers DW, Gibson DA, DiAntonio A, Milbrandt J. SARM1-Specific Motifs in the TIR Domain Enable NAD+ Loss and Regulate Injury-Induced SARM1 Activation. *Proc Natl Acad Sci USA* (2016) 113:E6271–80. doi: 10.1073/pnas.1601506113
58. Bratkowski M, Xie T, Thayer DA, Lad S, Mathur P, Yang Y-S, et al. Structural and Mechanistic Regulation of the Pro-Degenerative NAD Hydrolase SARM1. *Cell Rep* (2020) 32:107999. doi: 10.1016/j.celrep.2020.107999
59. Figley MD, Gu W, Nanson JD, Shi Y, Sasaki Y, Cunnea K, et al. SARM1 is a Metabolic Sensor Activated by an Increased NMN/NAD(+) Ratio to Trigger Axon Degeneration. *Neuron* (2021) 109:1118–36.e1111. doi: 10.1016/j.neuron.2021.02.009
60. Jiang Y, Liu T, Lee CH, Chang Q, Yang J, Zhang Z. The NAD(+)-Mediated Self-Inhibition Mechanism of Pro-Neurodegenerative SARM1. *Nature* (2020) 588:658–63. doi: 10.1038/s41586-020-2862-z
61. Sasaki Y, Zhu J, Shi Y, Gu W, Kobe B, Ve T, et al. Nicotinic Acid Mononucleotide is an Allosteric SARM1 Inhibitor Promoting Axonal Protection. *Exp Neurol* (2021) 345:113842. doi: 10.1016/j.expneurol.2021.113842
62. Di Stefano M, Nascimento-Ferreira I, Orsomando G, Mori V, Gilley J, Brown R, et al. A Rise in NAD Precursor Nicotinamide Mononucleotide (NMN) After Injury Promotes Axon Degeneration. *Cell Death Differ* (2015) 22:731–42. doi: 10.1038/cdd.2014.164
63. Di Stefano M, Loreto A, Orsomando G, Mori V, Zamporlini F, Hulse RP, et al. NMN Deamidase Delays Wallerian Degeneration and Rescues Axonal Defects Caused by NMNAT2 Deficiency *In Vivo*. *Curr Biol* (2017) 27:784–94. doi: 10.1016/j.cub.2017.01.070
64. Chan SL, Mukasa T, Santelli E, Low LY, Pascual J. The Crystal Structure of a TIR Domain From Arabidopsis Thaliana Reveals a Conserved Helical Region Unique to Plants. *Protein Sci* (2010) 19:155–61. doi: 10.1002/pro.275
65. Williams SJ, Sohn KH, Wan L, Bernoux M, Sarris PF, Segonzac C, et al. Structural Basis for Assembly and Function of a Heterodimeric Plant Immune Receptor. *Science* (2014) 344:299–303. doi: 10.1126/science.1247357
66. Williams SJ, Yin L, Foley G, Casey LW, Outram MA, Ericsson DJ, et al. Structure and Function of the TIR Domain From the Grape NLR Protein RPV1. *Front Plant Sci* (2016) 7:1850. doi: 10.3389/fpls.2016.01850

67. Zhang X, Bernoux M, Bentham AR, Newman TE, Ve T, Casey LW, et al. Multiple Functional Self-Association Interfaces in Plant TIR Domains. *Proc Natl Acad Sci USA* (2017) 114:E2046–52. doi: 10.1073/pnas.1621248114
68. Bernoux M, Ve T, Williams S, Warren C, Hatters D, Valkov E, et al. Structural and Functional Analysis of a Plant Resistance Protein TIR Domain Reveals Interfaces for Self-Association, Signaling, and Autoregulation. *Cell Host Microbe* (2011) 9:200–11. doi: 10.1016/j.chom.2011.02.009
69. Nishimura MT, Anderson RG, Cherkis KA, Law TF, Liu QL, Machius M, et al. TIR-Only Protein RBA1 Recognizes a Pathogen Effector to Regulate Cell Death in Arabidopsis. *Proc Natl Acad Sci USA* (2017) 114:E2053–62. doi: 10.1073/pnas.1620973114
70. Uno M, Watanabe-Nakayama T, Konno H, Akagi KI, Tsutsumi N, Fukao T, et al. Intramolecular Interaction Suggests an Autosuppression Mechanism for the Innate Immune Adaptor Protein Myd88. *Chem Commun (Camb)* (2018) 54:12318–21. doi: 10.1039/c8cc06480f
71. Moncrieffe MC, Bollschweiler D, Li B, Penczek PA, Hopkins L, Bryant CE, et al. MyD88 Death-Domain Oligomerization Determines Myddosome Architecture: Implications for Toll-Like Receptor Signaling. *Structure* (2020) 28:281–9.e283. doi: 10.1016/j.str.2020.01.003

Conflict of Interest: BK is a shareholder of Disarm Therapeutics, a wholly owned subsidiary of Eli Lilly & Company. BK is a consultant to Disarm Therapeutics. BK and WG receive research funding from Disarm Therapeutics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Immunology

ORIGINAL RESEARCH
published: 18 November 2021
doi: 10.3389/fimmu.2021.745160

Insights Into the Immune Response of the Black Soldier Fly Larvae to Bacteria

Daniele Bruno[1†], Aurora Montali[1†], Maristella Mastore[2], Maurizio Francesco Brivio[2], Amr Mohamed[3], Ling Tian[4], Annalisa Grimaldi[1], Morena Casartelli[5,6] and Gianluca Tettamanti[1,6]*

[1] Laboratory of Invertebrate Biology, Department of Biotechnology and Life Sciences, University of Insubria, Varese, Italy, [2] Laboratory of Comparative Immunology, Department of Theoretical and Applied Sciences, University of Insubria, Varese, Italy, [3] Laboratory of Insect Biochemistry and Molecular Sciences, Department of Entomology, Faculty of Science, Cairo University, Giza, Egypt, [4] Guangdong Provincial Key Laboratory of Agro-Animal Genomics and Molecular Breeding, Guangdong Provincial Sericulture and Mulberry Engineering Research Center, College of Animal Science, South China Agricultural University, Guangzhou, China, [5] Laboratory of Insect Physiology and Biotechnology, Department of Biosciences, University of Milano, Milan, Italy, [6] Interuniversity Center for Studies on Bioinspired Agro-Environmental Technology (BAT Center), University of Napoli Federico II, Naples, Italy

OPEN ACCESS

Edited by:
Ioannis Eleftherianos, George Washington University, United States

Reviewed by:
Javad Karimi, Ferdowsi University of Mashhad, Iran
Jorge Contreras-Garduño, National Autonomous University of Mexico, Mexico

***Correspondence:**
Gianluca Tettamanti
gianluca.tettamanti@uninsubria.it

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *21 July 2021*
Accepted: *01 November 2021*
Published: *18 November 2021*

Citation:
Bruno D, Montali A, Mastore M, Brivio MF, Mohamed A, Tian L, Grimaldi A, Casartelli M and Tettamanti G (2021) Insights Into the Immune Response of the Black Soldier Fly Larvae to Bacteria. Front. Immunol. 12:745160. doi: 10.3389/fimmu.2021.745160

In insects, a complex and effective immune system that can be rapidly activated by a plethora of stimuli has evolved. Although the main cellular and humoral mechanisms and their activation pathways are highly conserved across insects, the timing and the efficacy of triggered immune responses can differ among different species. In this scenario, an insect deserving particular attention is the black soldier fly (BSF), *Hermetia illucens* (Diptera: Stratiomyidae). Indeed, BSF larvae can be reared on a wide range of decaying organic substrates and, thanks to their high protein and lipid content, they represent a valuable source of macromolecules useful for different applications (e.g., production of feedstuff, bioplastics, and biodiesel), thus contributing to the development of circular economy supply chains for waste valorization. However, decaying substrates bring the larvae into contact with different potential pathogens that can challenge their health status and growth. Although these life strategies have presumably contributed to shape the evolution of a sophisticated and efficient immune system in this dipteran, knowledge about its functional features is still fragmentary. In the present study, we investigated the processes underpinning the immune response to bacteria in *H. illucens* larvae and characterized their reaction times. Our data demonstrate that the cellular and humoral responses in this insect show different kinetics: phagocytosis and encapsulation are rapidly triggered after the immune challenge, while the humoral components intervene later. Moreover, although both Gram-positive and Gram-negative bacteria are completely removed from the insect body within a few hours after injection, Gram-positive bacteria persist in the hemolymph longer than do Gram-negative bacteria. Finally, the activity of two key actors of the humoral response, i.e., lysozyme and phenoloxidase, show unusual dynamics as compared to other insects. This study represents the first detailed

characterization of the immune response to bacteria of *H. illucens* larvae, expanding knowledge on the defense mechanisms of this insect among Diptera. This information is a prerequisite to manipulating the larval immune response by nutritional and environmental factors to increase resistance to pathogens and optimize health status during mass rearing.

Keywords: cellular response, hemocytes, *Hermetia illucens*, humoral response, immune system

INTRODUCTION

In recent years, the possibility of rearing saprophagous insects on organic residues has been attracting increasing interest as insect-mediated bioconversion not only represents a strategy to valorize waste, but also to obtain bioproducts, thus generating circular economy value chains (1). In this scenario, the larvae of the black soldier fly (BSF), *Hermetia illucens* (Diptera: Stratiomyidae), play a key role since they have an outstanding ability to grow on a variety of by-products of different supply chains (2–4) and larval proteins can be used not only for formulating feedstuff for poultry, pigs, and fish (5), but also for manufacturing bioplastics (6). Moreover, BSF larvae represent a valuable source of lipids for producing biodiesel (7), chitin (8), and antimicrobial peptides (9).

Great effort has been invested to improve the quality of BSF larvae in terms of protein and lipid content (10–12), mainly by modulating the formulation of their feeding substrate, rearing temperature, humidity, and larval density. These parameters have, however, completely disregarded the fact that most of BSF-based applications involve the use of decaying waste substrates, thus potentially bringing the insect into contact with different pathogens that can challenge its health status and development. Only a few studies have evaluated the ability of *H. illucens* larvae to reduce the presence of pathogens in the rearing substrates (13, 14), although Huang and colleagues (15) have demonstrated that this insect can reduce pathogenic bacteria in the ingested substrate by means of gut immune mechanisms. This finding suggests that BSF larvae have a particularly efficient immune system, ensuring growth and development even in unhealthy organic substrates full of pathogens. However, information on the immune system of BSF larvae is incomplete (16), hindering further exploitation of this insect.

The immune system of insects possesses a sophisticated set of cellular and humoral innate mechanisms that can be rapidly activated in the presence of infections. The main actors of the cellular response are hemocytes, circulating cells involved in phagocytosis, encapsulation, and nodulation of non-self-antigens (17). In parallel, the humoral response is triggered by the interaction of Pattern Recognition Receptors (PRRs), located on the surface of insect cells, and Pathogen-associated Molecular Patterns (PAMPs) such as peptidoglycans, lipopolysaccharides, and β-1,3 glucans that are present on the pathogen surface (18). Humoral responses include the phenoloxidase (PO) system, an enzymatic cascade whose activation culminates with hemolymph clotting and melanin production, antimicrobial peptides (AMPs), lysozyme, and reactive oxygen species (19–21). Although the activation pathways are highly conserved in insects, the triggered cellular and humoral responses can differ between orders and species (18). In some Coleoptera, for example, the maximum number of hemocytes involved in the response to bacteria can be observed just a few hours after infection (22, 23), while in *Anopheles gambiae* this requires 24 hours (24). This finding indicates that, after an immune challenge, different amounts of time are required for circulating hemocytes to proliferate and thus counteract the threat. The pathways that regulate the humoral responses can greatly differ, too. For example, although the honey bee, *Apis millifera*, retains the major pathways typical of insect immune responses, the number of genes involved in the response to different non-self-antigens, including recognition, signaling, and effectors molecules, is reduced by one-third compared to *Drosophila melanogaster* (25), although this difference might be related to the existence of collective immunity that guarantees defenses against diseases in social insects (26). Another example comes from the pea aphid, *Acyrtosiphon pisum*. In this insect, the highly conserved peptidoglycan receptor proteins, as well as other IMD-associated genes involved in the defense against Gram-negative bacteria, including the AMPs Defensin and Cecropin, appear to be missing and Gram-positive bacteria are likely recognized by Gram-negative binding proteins or other receptors (27). Moreover, despite bacterial infections inducing the activation of lysozyme in that aphid, this enzyme does not seem to have a direct role in the host immune response against bacteria (28). Regarding the Diptera, although the immune mechanisms have been exhaustively and comprehensively characterized in the model *D. melanogaster*, it could be informative to investigate the peculiarities of the immune response in those species that have particular lifestyles. In the housefly *Musca domestica*, for example, many immune-related genes significantly expanded and diversified during evolution, as an adaptation to highly septic environments (29).

To our knowledge, and apart from the AMPs, only one study has investigated the larval immune system of BSF (16): here, the authors performed preliminary work on some elements of the humoral response, such as phenoloxidase, lysozyme, and antimicrobial peptides, completely overlooking the cellular mechanisms. More attention has been paid to the AMPs due to biotechnological applications and medical interests, some of which have been recently characterized and cloned from BSF larvae (30–33). Moreover, Vogel et al. (9) performed a RNASeq analysis on whole larvae grown on different rearing substrates (including those with high bacterial loads) and on selected tissues, demonstrating that the AMP expression profile

depends not only on the bacterial load of the substrate, but also on its nutrient composition.

In this paper we analyzed the cellular and humoral response of *H. illucens* larvae subjected to bacterial infection, focusing on the dynamics of the reactions. In particular, after injecting different immune elicitors (i.e., Gram-negative and Gram-positive bacteria, and chromatographic beads) into the hemocoel, we investigated the activity of hemocytes in phagocytosis and encapsulation processes, evaluated the antimicrobial activity of hemolymph as such and of the humoral components, analyzed the activity of the PO system and lysozyme, and quantified the expression of genes coding for AMPs.

This study is the first to characterize the immune response of *H. illucens* larvae to bacteria in detail. Resistance to diseases is a key issue for breeding BSF larvae in habitats with high bacterial load, which constantly expose the insect to infections that may reduce its survival and reproductive potential (34). In particular, we analyzed the innate immune response triggered by the Gram-positive saprotrophic bacterium *Micrococcus luteus* and the Gram-negative bacterium *Escherichia coli*, which are widely represented in environmental and soil samples and in organic wastes processed by BSF and other dipteran larvae (35, 36), upon their entrance in the insect hemocoel through lesions on the integument. Understanding the dialogue between cellular and humoral components of the immune system to counteract bacterial infections can be extremely important for maximining the productivity of BSF mass rearing activities and, ultimately, contributing to a wider utilization of the larvae as a viable feed source for livestock and as a sustainable waste management system.

MATERIALS AND METHODS

Insect Rearing

Larvae of *H. illucens* used in this study were obtained from a colony kept for six years at the University of Insubria (Varese, Italy).

Eggs were collected in Petri dishes and maintained at 27 ± 0.5°C until hatching, according to the procedures of Pimentel et al. (37). Neonate larvae were fed with a standard diet for Diptera (50% wheat bran, 30% corn meal, and 20% alfalfa meal mixed at a ratio of 1:1 dry matter:water) (38). Four days after hatching, batches of 300 larvae were transferred to plastic containers (16 x 16 x 9 cm), fed with standard diet, and reared in the dark at 27 ± 0.5°C and 70 ± 0.5% relative humidity. Once at the pupal stage, insects were removed from the substrate and transferred to a cage (70 x 70 x 120 cm) until adult eclosion. Flies were kept at 30 ± 0.5°C, 70 ± 5% relative humidity and with a 12:12 hours light:dark photoperiod, according to Bruno et al. (39).

Bacterial Strains

The bacterial strains *Escherichia coli* (Strain K12, Sigma-Aldrich, USA) and *Micrococcus luteus* ATCC No. 4698 (Sigma-Aldrich) were used for all the analyses except the encapsulation assay. We selected these bacteria because they are widely represented in organic wastes that can be used by *H. illucens* larvae to grow (36) and can infect the insect through superficial lesions of the integument.

Bacteria were grown in 10 ml of LB broth (Sigma-Aldrich) overnight at 37°C under shaking at 160 rpm. Then, 1-ml aliquots were centrifuged at 1620 x *g* for 15 minutes and cell pellets were washed three times with phosphate buffer (38 mM KH_2PO_4, 61.4 mM K_2HPO_4, pH 7.4). Finally, after measuring the optical density of the culture at 600 nm (OD_{600nm}; one unit of OD_{600nm} corresponds to 4.12 x 10^8 CFU/ml of *E. coli* and 1.83 x 10^7 CFU/ml of *M. luteus*), cells were resuspended and diluted in phosphate-buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4/KH_2PO_4, pH 7.4) to the final concentration for injection into larvae (40).

Injection of Larvae, Hemolymph Collection, and Control Groups

Last instar larvae were carefully washed with tap water to remove diet debris, then with 0.5% sodium hypochlorite (in tap water, v/v) and 70% ethanol (in distilled water, v/v). To evaluate the activation of cellular and humoral mechanisms by the co-administration of Gram-positive and Gram-negative bacteria, an *E. coli/M. luteus* mix was used to infect the larvae (5µl of the bacterial mix for each larva), unless for those markers where the response of the immune system to Gram-positive or Gram-negative bacteria needed to be specifically evaluated (i.e., antimicrobial activity and phagocytosis). For details on the injection of pHrodo-conjugated bacteria and chromatographic beads, please see the sections below.

All the injections were performed between the third last and penultimate metamere by using a Hamilton 700 10-µl syringe (Hamilton, USA). After injection, larvae were placed in sterile Petri dishes at 27 ± 0.5°C and 70 ± 0.5% relative humidity without food until bleeding.

Hemolymph was collected by piercing the larvae, previously anesthetized on ice, between the last and penultimate metamere with a sterile needle. Each experiment was performed by pooling samples of hemolymph collected from at least 15 larvae and 3 experiments were performed for each test.

As the aim of the present work was to analyze the immune response triggered by the entrance of bacteria through lesions on the body surface, uninjected larvae (naïve) were considered as control groups for all the experiments, as detailed below. To exclude potential side effects of the puncture or the buffer used to resuspend the bacteria, larvae punctured with a sterile needle only or injected with sterile PBS (5 µl) were analyzed at different time points (negative controls) (**Figure S1**). Additional controls were performed by examining starved naïve larvae at different time points (3, 6, 14, 24, and 48 hours) to exclude possible effects of food deprivation on different immune markers (**Figure S2**).

Determination of the Optimal Bacterial Dose for the Infection of Larvae

The larval immune response was triggered by injecting a mixture of Gram-negative and Gram-positive bacteria into the hemocoel. To determine the adequate bacterial dose for infecting insects,

larvae were injected with 5 µl of different concentrations of an *E. coli/M. luteus* mix (10^4, 10^5, 10^6, 10^7, 10^8, and 10^9 CFU/ml) and their mortality was monitored every 24 hours for 3 days. Twenty-five larvae for each condition were used and the experiment was conducted in triplicate. This preliminary experiment was carried out to define the bacterial dose at which the immune system of BSF larvae could be stimulated without causing significant insect mortality during the time considered for the analyses. According to the results obtained in this assay (**Figure S3**), larvae were injected with 5µl of *E. coli/M. luteus* mix at a concentration of 10^5 CFU/ml for all the experiments, unless otherwise specified.

Evaluation of Antimicrobial Activity of the Hemolymph

CFU Count by the Spread-Plating Method

To determine the bacterial load (CFU/ml) in the hemocoel, hemolymph was collected 6, 14, 24, and 48 hours after infection with *E. coli* or *M. luteus* at a concentration of 10^5 CFU/ml. Samples were diluted 1:100 with sterile PBS and then plated onto 20 ml of LB broth agar (Sigma-Aldrich). The plates were incubated at 37°C for 24 hours and then colonies were counted. The CFU/ml was calculated as CFU/ml=n° colonies x dilution factor.

CFU Count by the Track-Dilution Method

Larvae were injected as described in the section *"Injection of larvae, hemolymph collection, and control groups"* and placed in sterile Petri dishes for 6, 14, 24, or 48 hours. After collection, hemolymph was centrifuged at 250 x *g* to remove the hemocytes. Then, 10 µl of cell-free hemolymph were incubated with 90 µl of 10^6 CFU/ml *E. coli* or *M. luteus* at 37°C for 3 hours, according to Mastore and Brivio (41). After incubation, samples were serially diluted and then 10 µl of each dilution were dropped onto agar plates. Plates were tilted to allow drops to flow downward and then incubated overnight at 37°C. The colonies (CFU) in each plate were counted and the concentration was calculated as CFU/ml=n° colonies x 10 x dilution factor. *E. coli* or *M. luteus* (10^6 CFU/ml) without the addition of cell-free hemolymph, and sterile PBS were used as controls. Time point 0 (T_0) corresponded to cell-free hemolymph derived from naïve larvae incubated for 3 hours with 10^6 CFU/ml *E. coli* or *M. luteus*.

Hemocyte Counts

To quantify the hemocytes, hemolymph samples were extracted from the larvae 6, 14, 24, or 48 hours after the infection. Hemocytes were counted by loading diluted hemolymph (10 µl hemolymph added to 90 µl of 0.4% Trypan blue, ThermoFisher, USA) into FAST READ 102 counting chambers (Biosigma S.R.L., Italy) according to the manufacturer's instructions.

Phagocytosis Assay

To investigate the phagocytic activity of hemocytes, fluorogenic probes that dramatically increase their fluorescence as pH decreases from neutral to acidic were used. For this purpose, 5 µl of pHrodo™ Red *Staphylococcus aureus* BioParticles Conjugate (0.2 mg/ml) or pHrodo™ Green *E. coli* BioParticles Conjugate (0.2 mg/ml) (Molecular Probes, USA) were injected into the larval hemocoel. Five minutes and 1, 14, and 24 hours later, the hemolymph was extracted from the larvae and 200 µl were loaded on round glass coverslips for 15 minutes in the dark to allow cell adhesion. After discarding the liquid fraction, the hemocytes were fixed with 4% paraformaldehyde in PBS for 5 minutes and then washed twice with sterile PBS. Cells were incubated with DAPI (100 ng/ml in PBS) for nuclear detection and then coverslips were mounted on slides with Citifluor (Citifluor Ltd, United Kingdom) (42). The hemocytes were analyzed by using an Eclipse Ni-U fluorescence microscope (Nikon, Japan) equipped with a DS-SM-L1 digital camera (Nikon, Japan). Phagocytic cells were quantified by analyzing five different images for each independent experiment. The experiments were conducted in triplicate for each time analyzed.

Encapsulation Assay

B-Agarose, B-Sephadex G-100, CM-Sephadex C-25, and DEAE-Sephadex A-25 beads (Sigma-Aldrich) were used to analyze the encapsulation process. The method was adapted from Mastore and Brivio (43). Briefly, beads were first washed three times with sterile PBS and resuspended in the same buffer, centrifuging at 1620 x *g* for 2 minutes after each wash. Of the final preparation (approximately 5 beads/µl), 5 µl were injected into the larvae. The hemolymph was collected at different time points (2, 4, 14, and 24 hours) and diluted 2:1 (hemolymph/medium) (v/v) with Schneider's Insect Medium. The encapsulation process was evaluated by using an Olympus IX51 microscope (Olympus, Japan) equipped with an Optika C-P20M camera (Nikon). Encapsulation is an extremely dynamic process due to the different number of hemocytes that adhere to the bead over time, therefore, a qualitative approach was used for this assay. The time at which at least 30% of the beads were at least partially covered by hemocytes was selected as starting point of the process. The experiments were conducted in triplicate for each time point analyzed.

PO System Activity

To monitor the activity of the PO system after infection, larvae were injected with increasing concentrations of the *E. coli/M. luteus* mix (i.e., 10^5, 10^7, 10^8, or 10^9 CFU/ml). After 7 minutes, (according to preliminary tests this is the time needed to activate the PO system in *H. illucens* larvae), the hemolymph was collected and centrifuged at 250 x *g* for 5 minutes at 4°C to obtain the humoral (cell-free) fraction. The method used to analyze the PO activity was adapted from Garriga et al. (44). Briefly, 10 µl of cell-free hemolymph were added to 990 µL of 8 mM L-Dopa (L-3-4 dihydroxyphenylalanine) (Sigma-Aldrich) in 10 mM Tris-HCl, pH 7.4. Activation of the PO system was also evaluated by incubating 10 µl of cell-free hemolymph collected from larvae infected with 10^5 CFU/ml of the *E. coli/M. luteus* mix, with 10 µl of β-glucans from *Saccharomyces cerevisiae* (Zymosan®, Sigma-Aldrich) in 980 µl of 8 mM L-Dopa in 10 mM Tris-HCl, pH 7.4.

The formation of dopachrome was evaluated with a V-560 double-beam spectrophotometer (Jasco, USA). Before measurement, samples were incubated at room temperature for 10 minutes and the absorbance was measured at 490 nm every 10

minutes. The increase in absorbance over time (ΔAbs) was determined in the linear range of the curve within 60 minutes. The range in which the absorbance values were registered was 0.1-1 according to Harvey (45).

Lysozyme Activity

Larvae were infected and placed in sterile Petri dishes for 6, 14, 24, and 48 hours. After collecting the hemolymph, the activity of lysozyme was evaluated according to Garriga et al. (44) with slight modifications. Briefly, N-Phenylthiourea (PTU, Sigma-Aldrich) crystals were added to the samples to prevent activation of the PO system. Samples were centrifuged twice at 250 x *g* for 5 minutes at 4°C, and then at 1600 x *g* for 10 minutes at 4°C, collecting the supernatant. Cell-free hemolymph was diluted 1:10 with sterile PBS. Thereafter, 100 µl of the samples were added to 150 µl of *Micrococcus lysodeikticus* (0.45 mg/ml in 30 mM phosphate buffer, pH 7.2; Sigma-Aldrich), with an optical density of 0.6-0.7, which was used as substrate to measure lysozyme relative activity. *M. lysodeikticus* alone and cell-free hemolymph without addition of the bacterium were used as controls. Absorbance at 450 nm was measured every 30 s for 10 minutes in 96-well plates using a Bio-Rad iMark™ Microplate Absorbance reader (Bio-Rad, USA).

qRT-PCR Analysis

For qRT-PCR analysis, larvae at 3, 6, 14, 24, and 48 hours after the infection were anesthetized on ice and then dissected. The fat body was isolated, frozen in liquid nitrogen, and stored at -80°C until use. In parallel, the hemolymph collected from larvae infected for 3, 6, 14, and 24 hours was isolated as described in the section *"Injection of larvae and hemolymph collection"* and placed in sterile tissue culture Petri dishes (5.5 x 1.5 cm) for 20 minutes in the dark to allow cell adhesion. Afterwards, the hemocytes were washed twice with Schneider's Insect Medium to remove fat body debris, then scraped from the Petri dish with Trizol Reagent (Life Technologies, USA), and stored at -80°C until use.

RNA was extracted from 30-40 mg of fat body and 10^6 hemocytes/ml with Trizol Reagent. Samples were treated with TURBO DNA-free Kit (Life Technologies) to avoid genomic DNA contamination and RNA quality was verified through gel electrophoresis.

Retrotranscription was performed with M-MLV reverse transcriptase (Life Technologies). Primers used for qRT-PCR (**Table 1**) were designed based on the sequences of Defensin, Diptericin, Lysozyme, and Ribosomal Protein L5 (RPL5) obtained from *de novo* transcriptome of *H. illucens* larvae (Transcriptome accession number: ERP122672) (46). RPL5 was used as housekeeping gene (34, 39). Real-Time PCR was performed using the iTaq Universal SYBR Green Supermix (Bio-Rad) and the CFX Connect Real-Time PCR Detection System (Bio-Rad), according to (47). Relative expression of the genes was calculated with the $2^{-\Delta\Delta Ct}$ method. Each value was the result of the analysis of five series of samples.

Statistical Analysis

Statistical analyses were performed using GraphPad Prism version 7.00 (GraphPad software, La Jolla, USA). The antimicrobial and lysozyme activity, AMPs and lysozyme expression, and the activation of PO system were analyzed using One-Way ANOVA followed by Tukey's multiple-comparison *post hoc* test. Hemocyte counts were analyzed using the Unpaired Student's *t*-test while survival curves of the larvae were compared using Kaplan-Meier followed by log-rank analysis. All the data were subjected to logarithmic ($\log_{10}$) transformation for normality before statistical analysis. Statistical differences between groups were considered significant at a p-value < 0.05.

RESULTS

Determination of the Optimal Bacterial Dose for the Infection of Larvae

The bacterial concentration used in this study was selected by monitoring larvae injected with different concentrations of the *E. coli*/*M. luteus* mix (from 10^4 to 10^9 CFU/ml) for up to 72 hours. As shown in **Figure S3**, survival of the larvae was inversely proportional to the bacterial dose administered. In particular, while 100% of the larvae injected with 10^4 CFU/ml bacteria were alive 72 hours after the infection, higher bacterial doses (10^5, 10^6, 10^7, and 10^8 CFU/ml) progressively decreased the number of live larvae over time (72%, 64%, 60%, and 16%, respectively, at 72 hours). All the insects died within 24 hours at the highest concentration of bacteria (10^9 CFU/ml). According to these results, for all the subsequent experiments, larvae were infected by injecting 10^5 CFU/ml of the bacterial mix as this was the lowest concentration able to reduce the welfare of the larvae, without causing high mortality, thus allowing us to monitor the immune response over time.

TABLE 1 | Sequence of primers used in this study.

Gene name	Transcriptome Accession number	Contig number	Primer sequences
HiDefensin	ERP122672	TRINITY_DN10226_C0_G1_I2	F: GCGTTCTATTCTCGTCTTGG R: TGCTGTTCCACTACCTGACT
HiDiptericin	ERP122672	TRINITY_DN6246_C0_GI_I1	F: CCCAGTGAGCGATGAGGAA R: GTGAAGGGTATTGCGTCCAT
HiLysozyme	ERP122672	TRINITY_DN12175_C0_G2_I1	F: GCCCAAGGCAAGGTTTACA R: TGGCGAGGGTGGTTAGATTC
HiRPL5	ERP122672	TRINITY_DN8551_C0_G1_I2	F: AGTCAGTCTTTCCCTCACGA R: GCGTCAACTCGGATGCTA

Analysis of the Antimicrobial Activity in the Hemolymph

The bacterial load in the hemolymph was evaluated at different time points after the bacterial infection. While the *E. coli* concentration dropped to zero 14 hours after the infection (**Figure 1A**), *M. luteus* proliferation at 6 hours (4.34 ± 4.2 x 10^5 CFU/ml) was higher than that observed for *E. coli* at the same time point after infection (5.73 ± 1.4 x 10^2 CFU/ml) and even increased at 14 hours (2.62 ± 0.8 x 10^6 CFU/ml). However, no CFUs were detected 24 hours from the infection onwards for the two bacteria (**Figures 1A, B**).

The results of a track dilution method showed a time-dependent trend in the antimicrobial activity of the cell-free hemolymph against *E. coli*. In fact, the initial bacterial concentration (2.26 ± 0.4 x 10^8 CFU/ml) dropped to 1.16 ± 0.5 x 10^6 CFU/ml and to 3.08 ± 2.9 x 10^4 CFU/ml by adding humoral components of cell-free hemolymph collected from larvae at 6 and 14 hours after the infection, respectively, and no colonies were found by testing hemolymph from larvae at 48 hours after the infection (**Figure 1C**). The antimicrobial activity against *M. luteus* showed a different trend. Although the initial concentration of this bacterium (4.9 ± 2.3 x 10^9 CFU/ml) constantly fell at different time points, a significant number of colonies were still observed in hemolymph collected at 48 hours (**Figure 1D**). These data demonstrate that bacterial infection is rapidly reduced due to the cooperation of cellular and humoral responses (**Figures 1A, B**) and indicate that the cellular response plays a key role in counteracting the infection of *M. luteus* since, unlike the response to *E. coli*, cell-free hemolymph collected from larvae at 48 hours after infection is unable to completely inhibit the bacterial growth (**Figures 1C, D**).

Cellular Response

Hemocyte Number

To evaluate possible variations in the number of circulating hemocytes after infection with the *E. coli*/*M. luteus* mix, hemocytes were counted. A significant difference between the hemocyte number in control (8.45 ± 2.2 x 10^5 cells/ml)

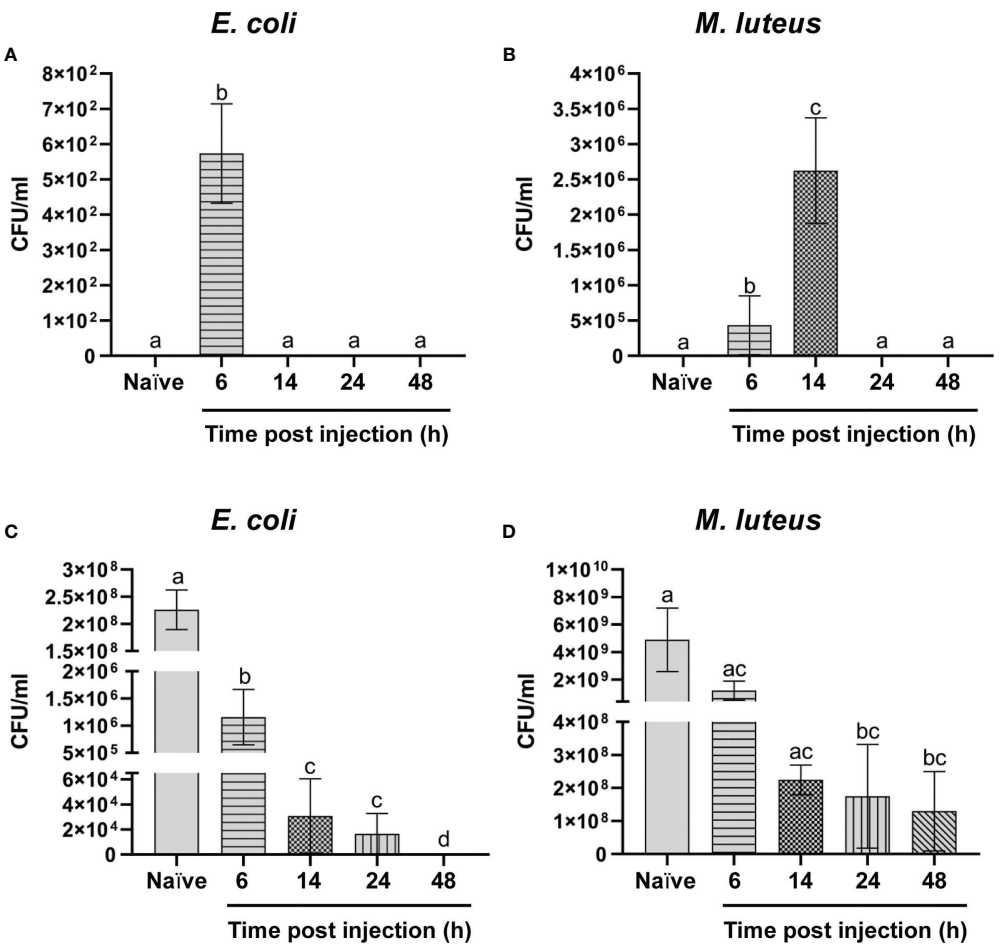


FIGURE 1 | Analysis of the antimicrobial activity. **(A, B)** Humoral and cellular activity against *Escherichia coli* **(A)** and *Micrococcus luteus* **(B)**. **(C, D)** Humoral antimicrobial activity against *E. coli* **(C)** and *M. luteus* **(D)**. Values represent mean ± s.e.m. Different letters indicate statistically significant differences among treatments (One-Way ANOVA: **A**) $F_{4\text{-}10} = 91.65$, $p < 0.0001$; **B**) $F_{4\text{-}10} = 74.62$, $p < 0.0001$; **C**) $F_{4\text{-}10} = 52.36$, $p < 0.0001$; **D**) $F_{4\text{-}10} = 4.694$, $p < 0.0216$).

and infected (2.19 ± 0.2 x 10^6 cells/ml) larvae 6 hours after injecting the bacteria was observed (**Figure 2**). This difference remained stable up to 24 hours (4.59 ± 1.4 x 10^5 cells/ml vs 1.82 ± 0.2 x 10^6 cells/ml) and disappeared 48 hours after infection (6.43 ± 2.3 x 10^5 cells/ml vs 1.55 ± 0.5 x 10^6 cells/ml) (**Figure 2**).

Phagocytosis

To analyze the trend of the phagocytosis process over time, Gram-positive and Gram-negative bacteria conjugated with pHrodo fluorophores were injected in the larvae and the hemocytes examined at different time points (**Table 2**). For both bacterial species, a time-dependent increase in phagocytic activity was observed up to 1 hour after injection in the larvae. In particular, while 5 minutes after infection the counts for hemocytes that had engulfed fluorescent bacteria were 22 ± 4% and 45 ± 3% for *E. coli* and *S. aureus*, respectively (**Figures 3A, B**), an increasing number of active phagocytes was observed after 1 hour (41 ± 2% and 86 ± 6% for *E. coli* and *S. aureus*, respectively) (**Figures 3C, D**). Moreover, a few, engulfed Gram-positive bacteria were detected after 14 hours (6 ± 1%) (**Figures 3E, F**), while no signal was found at 24 hours for either bacterial species (**Figures 3G, H**).

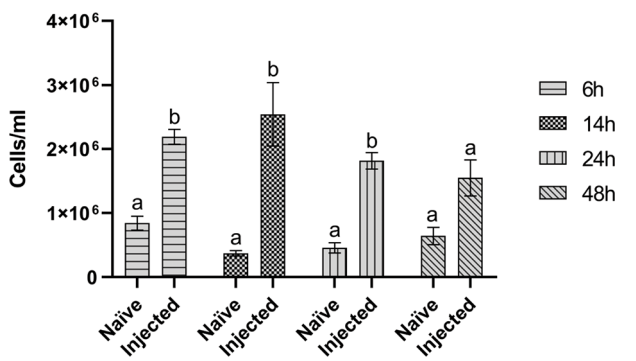


FIGURE 2 | Hemocyte counts. Hemocyte counts of naïve larvae (Naïve) and larvae infected with 10^5 CFU/ml *E. coli/M. luteus* mix (Injected) at different time points (6, 14, 24, and 48 hours after infection). Values represent mean ± s.e.m. Different letters indicate statistically significant differences among naïve and injected larvae (Unpaired Student's *t*-test: for 6h $t = 3.289$, $df = 5$, $p = 0.0217$; for 14h $t = 7.864$, $df = 5$, $p = 0.0005$; for 24h $t = 4.661$, $df = 4$, $p = 0.0096$; for 48h $t = 1.595$, $df = 4$, $p = 0.1859$).

TABLE 2 | Quantification of cells undergoing phagocytosis.

Time post infection	Percentage of phagocytic cells	
	E. coli	*S. aureus*
5 minutes	22 ± 4[a]	45 ± 3[a]
60 minutes	41 ± 2[b]	86 ± 6[b]
14 hours	0 ± 0[c]	6 ± 1[c]
24 hours	0 ± 0[c]	0 ± 0[d]

Values represent mean ± s.e.m. Different letters indicate statistically significant differences among cells undergoing phagocytosis (E. coli or S. aureus) and non-phagocytic cells at different time points (One-Way ANOVA: for E. coli $F_{3-11} = 876.8$, $p < 0.0001$; for S. aureus $F_{3-11} = 910.8$, $p < 0.0001$).

Encapsulation

We analyzed the qualitative trend of the encapsulation process and melanin deposition over time by injecting different beads into the larval hemocoel. First, we compared the effects of two types of neutrally charged beads, i.e., agarose and dextran. B-Agarose beads partially surrounded by hemocytes were visible 2 hours after being injected in the larva (**Figure 4A**) and, despite the extensive recruitment of cells adhering to the beads over time (**Figures 4E, I, M**), production and deposition of melanin were never observed. Conversely, beads surrounded by hemocytes and melanin deposition could be observed from 14 hours onwards in larvae injected with B-Sephadex beads (**Figures 4B, F, J, N**).

We then investigated whether the charge of the beads could affect the encapsulation process and melanin deposition. For this, negatively and positively charged dextran beads were tested. In larvae treated with CM-Sephadex beads (negatively charged), hemocytes began to adhere to the beads very quickly (2-4 hours after the injection) (**Figures 4C, G**). Although the beads were not completely surrounded by hemocytes at longer times after injection (**Figure 4K**), full melanization could be seen 24 hours after being injected (**Figure 4O**). Conversely, for DEAE-Sephadex beads (positively charged) (**Figure 4D**), close adhesion of the cells to the foreign body and its melanization could already be observed 4 hours after their injection (**Figure 4H**), and a typical melanotic capsule could be detected after 14 hours (**Figures 4L, P**).

Humoral Response

PO System

Activation of the PO system was first evaluated in larvae injected with 10^5 CFU/ml *E. coli/M. luteus* mix. The ΔAbs value in the cell-free hemolymph from control larvae was significantly higher than that of larvae infected for 7 minutes (**Figure 5A**). The pattern was unchanged in hemolymph collected from larvae infected with the bacterial mix for longer times (30 and 60 minutes) (**Figure 5A**), suggesting the possibility of a potential inhibition/inactivation of the PO system.

Zymosan, a specific activator of the PO system was added to the cell-free fraction of hemolymph from infected larvae to verify enzymatic system functioning (**Figure 5B**). The results showed a consistent activation of phenoloxidase under this condition and values were comparable to those of hemolymph from naïve larvae treated with Zymosan (**Figure 5B**). Moreover, although the addition of Zymosan to the hemolymph collected from larvae infected for longer times (30 and 60 minutes) reduced PO system activity compared to larvae infected for 7 minutes (**Figure 5B**), Zymosan always activated the enzyme system.

As these experiments with Zymosan demonstrated PO system functioning, we evaluated whether activation might possibly depend on the bacterial concentration. Therefore, we compared phenoloxidase activation in the cell-free fraction of hemolymph from larvae injected with increasing concentrations of the bacterial mix (10^5, 10^7, 10^8, and 10^9 CFU/ml) (**Figure 5C**). The data showed that the PO system is activated at a concentration of bacteria higher than or equal to 10^9 CFU/ml, indicating that this concentration is the threshold for full activation of the enzyme complex.

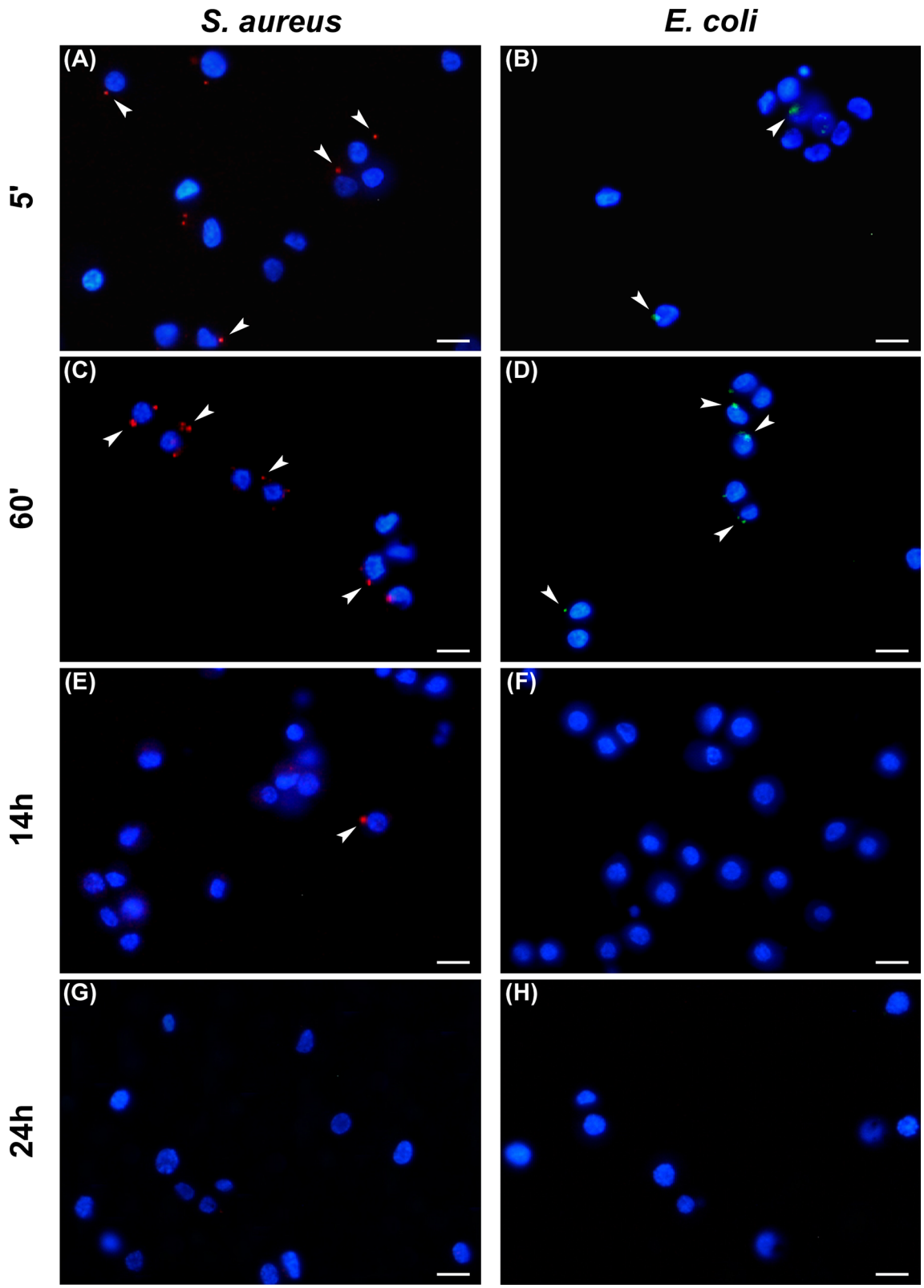


FIGURE 3 | Analysis of phagocytosis. Hemocytes undergoing phagocytosis 5 minutes **(A, B)**, 60 minutes **(C, D)**, 14 hours **(E, F)**, and 24 hours **(G, H)** after injection of *Staphylococcus aureus* **(A, C, E, G)** and *Escherichia coli* **(B, D, F, H)** conjugated with pHrodo. Arrows indicate phagocytosed bacteria. Bars: 10 µm.

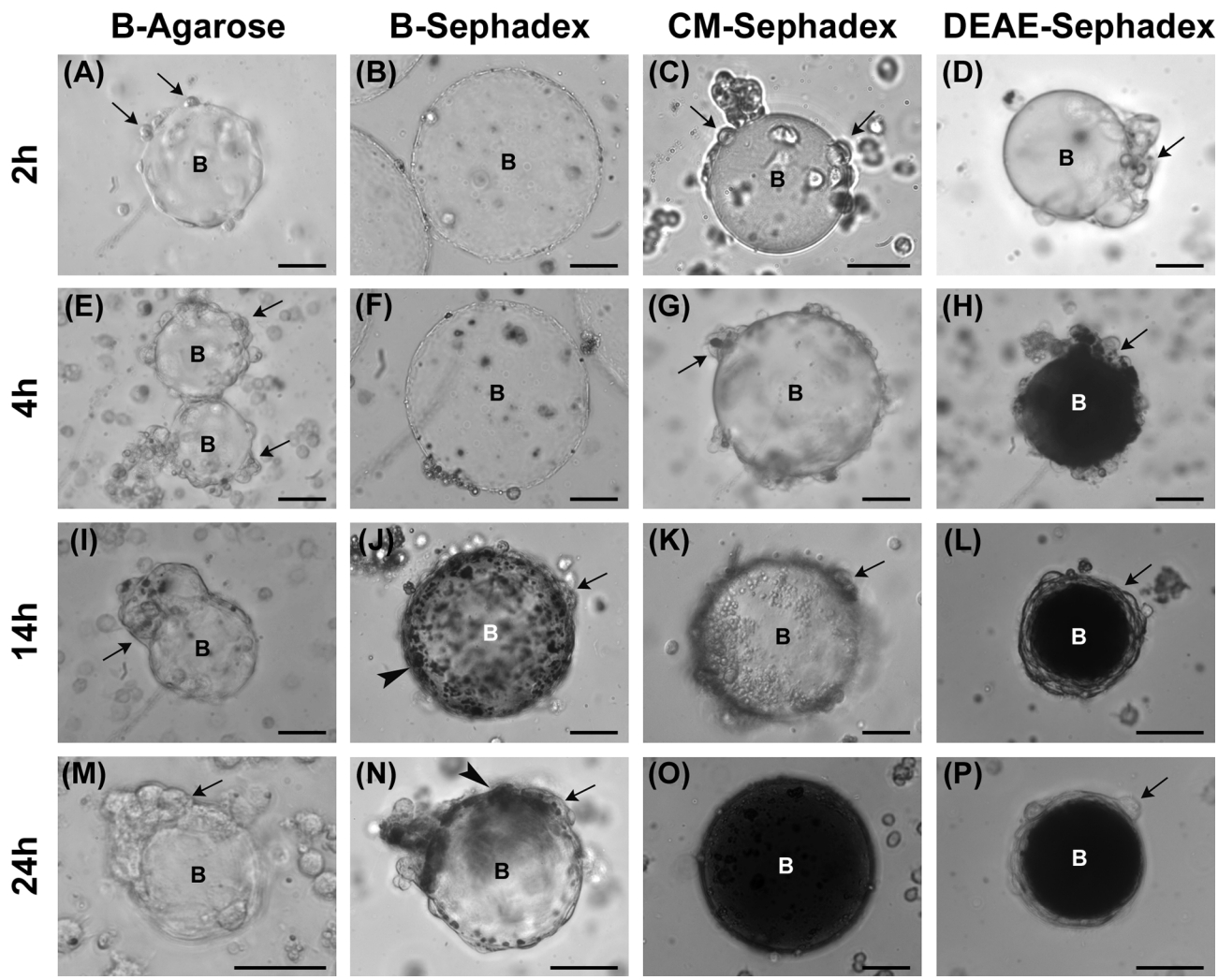


FIGURE 4 | Analysis of encapsulation. Effect of B-Agarose **(A, E, I, M)**, B-Sephadex **(B, F, J, N)**, CM-Sephadex **(C, G, K, O)**, and DEAE-Sephadex **(D, H, L, P)** beads on encapsulation and melanization. Hemolymph was analyzed 2 **(A–D)**, 4 **(E–H)**, 14 **(I–L)**, and 24 (M-P) hours after injecting the beads in the larvae. Arrows, adherent hemocytes; arrowheads, melanin deposition; B, beads. Bars: 40 µm.

Expression and Activity of Lysozyme

In cell-free hemolymph from naïve larvae, we measured a basal activity of lysozyme (2288 ± 211 U/ml), which progressively increased after infection with the bacterial mix (**Figure 6A**). Specifically, the activity of the enzyme increased from 6633 ± 133 U/ml after 6 hours up to 9100 ± 900 U/ml after 14 hours and 11316 ± 687 U/ml at 24 hours (**Figure 6A**).

We also analyzed the transcription of *HiLysozyme* both in the fat body and in hemocytes. While the mRNA levels of this gene in the fat body of infected larvae were not significantly higher than in controls (data not shown), levels in hemocytes varied over time. In fact, a rapid and marked increase (15-fold) in *HiLysozyme* mRNA levels was observed within 3 hours after infection. Then, mRNA levels decreased at 6 hours and returned to basal values after 14 hours (**Figure 6B**).

Expression of Antimicrobial Peptides

The expression of *HiDiptericin* and *HiDefensin* was investigated both in the fat body and in hemocytes. These two representative AMPs were selected to obtain information about AMP response to Gram-negative and Gram-positive bacteria, respectively. In the fat body the slight increase in mRNA levels of both AMPs 3 hours after the infection changed at later times. In particular, a >250-fold increase in the expression of Diptericin was observed at 6-14 hours; then, mRNA levels markedly decreased within 24 hours (**Figure 7A**). Conversely, Defensin expression increased later compared to Diptericin (110-fold change at 14 hours) and levels remained stable up to 48 hours (**Figure 7B**).

The expression levels of the two AMPs in hemocytes displayed a peak within 3 hours from the infection [1400-fold for Diptericin and 90-fold for Defensin (**Figures 7C, D**)]; then,

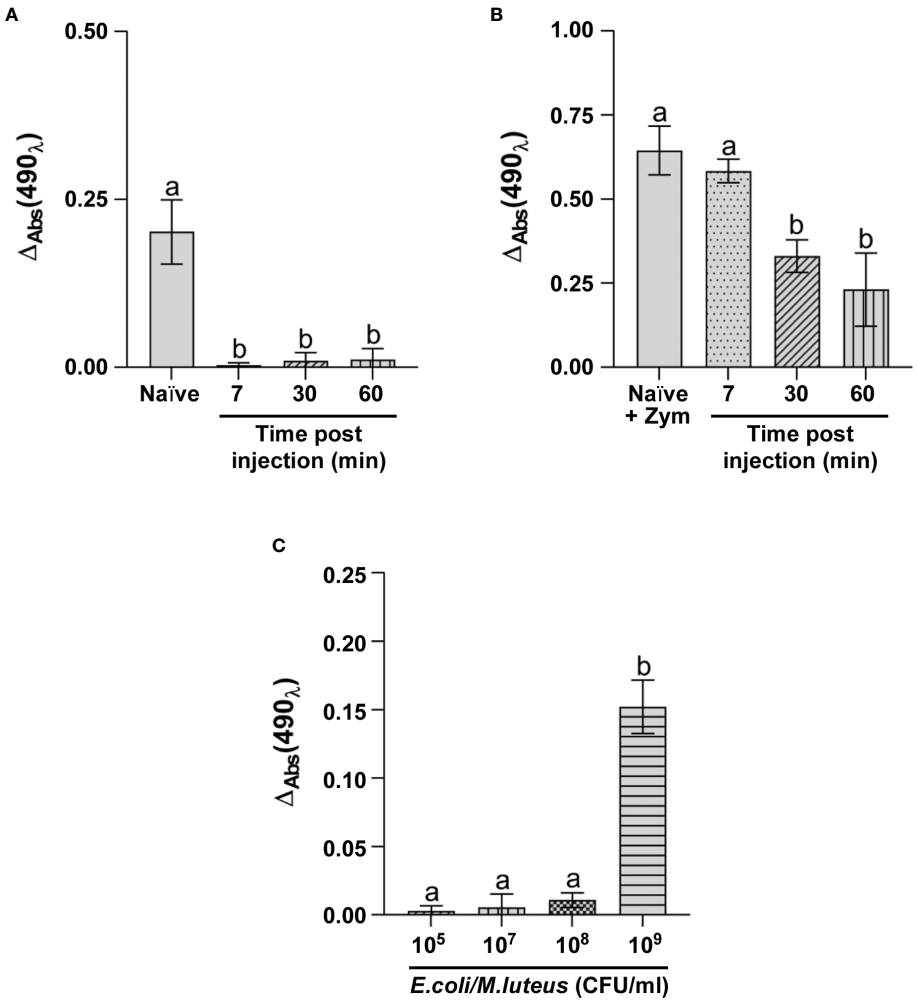


FIGURE 5 | Analysis of PO system activation. ΔAbs at 490 nm of phenoloxidase in **(A)** hemolymph of naïve larvae (Naïve) and larvae infected for 7, 30 and 60 minutes with 10^5 CFU/ml of *E. coli/M. luteus* mix; **(B)** hemolymph isolated from naïve larvae (Naïve + Zym) and larvae infected for 7, 30, and 60 minutes with 10^5 CFU/ml of *E. coli/M. luteus*, treated with Zymosan; **(C)** hemolymph isolated from larvae infected for 7 minutes with different concentrations of *E. coli/M. luteus* mix. Values represent mean ± s.e.m. Different letters indicate statistically significant differences among treatments (One Way ANOVA: **A**) $F_{3\text{-}8} = 41.54$, $p < 0.0001$; **B**) $F_{3\text{-}8} = 22.95$, $p = 0.0003$; **C**) $F_{3\text{-}8} = 122.7$, $p < 0.0001$).

mRNA levels gradually decreased at the later times (**Figures 7C, D**).

DISCUSSION

Over the last few years, considerable effort has been devoted to investigating BSF biology. In particular, the literature provides information regarding the functional properties of its digestive (39, 46, 48–51) and reproductive (52, 53) system, the fat body (37), and other organs and structures (54–58). Although interest in these aspects is increasing, information about the immune system of BSF is still fragmentary. In fact, different studies have focused attention on the antimicrobial activity of BSF larval extracts (9, 31, 32, 59) while, to the best of our knowledge, only one considered the response of the immune components after an immune challenge (16). However, it must be highlighted that, in contrast to this latter study in which larvae were infected by pricking with an unknown bacterial load, the present study applied a controlled immune challenge by injecting a defined amount of bacteria into the hemocoel. The bacterial load for the infections was accurately selected to reduce insect welfare, stimulating the immune system, but ensuring a high survival rate of the larvae to analyze the cellular and humoral responses in detail over time (**Figure 8**).

In insects, a continuous cross-talk between hemocytes and humoral molecules is fundamental for maintaining the hemolymph devoid of pathogens and parasites (20, 60). During infection, the cellular response is triggered when transmembrane or soluble PRPs recognize PAMPs and usually ensues more

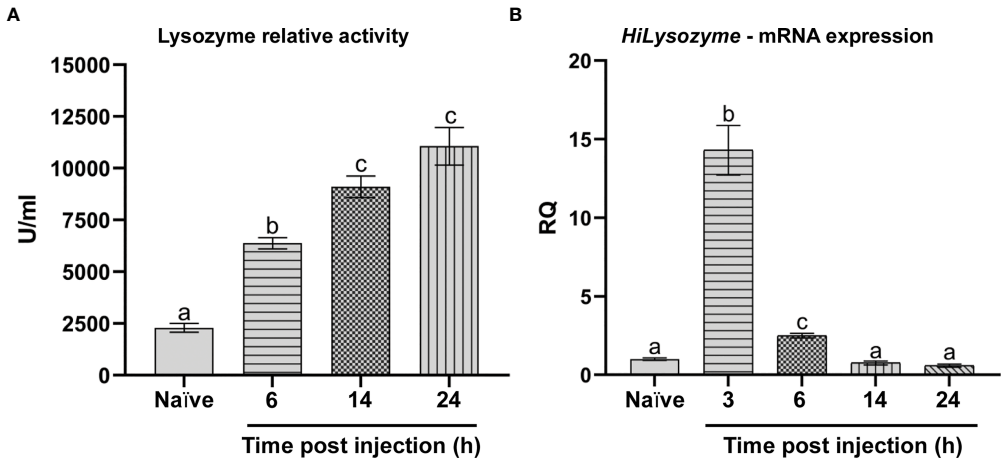


FIGURE 6 | Lysozyme activity and *HiLysozyme* expression. **(A)** Lysozyme relative activity in the hemolymph of naïve and infected larvae at different time points. **(B)** *HiLysozyme* mRNA levels in hemocytes of naïve and infected larvae at different time points. Naïve larvae starved for 6 and 3 hours were used as controls for the activity and the expression of lysozyme, respectively. Values represent mean ± s.e.m. Different letters indicate statistically significant differences among treatments (One-Way ANOVA: **A**) $F_{4\text{-}10} = 128.1$, $p < 0.0001$; **B**) $F_{4\text{-}10} = 107.2$, $p < 0.0001$).

rapidly than the humoral response to quickly reduce the pathogen load in the hemolymph. The few remaining microorganisms not eliminated by hemocytes are killed by humoral molecules (61). Our results confirm this sequence of events and demonstrate the need for synergic action of the two components of the immune system to efficiently remove invading bacteria in *H. illucens* larvae. In particular, the antimicrobial activity against *E. coli* and *M. luteus* elicited by the whole hemolymph and that of the cell-free fraction were significantly different, with a quite rapid clearance of bacteria displayed by the former versus the inability of the latter to kill both bacteria in a short time. Moreover, unlike *E. coli*, which is totally removed within 48 hours by both the hemolymph and cell-free fraction, the latter was unable to eliminate *M. luteus* even at longer times from the infection. The longer time required for humoral molecules to remove *M. luteus* could be due to the potential capability of this bacterium to elude the action of AMPs, as previously demonstrated for some Gram-positive bacteria (62–64). In addition, the amount of time needed to fully activate the Toll and IMD pathways, which are involved in the defense against Gram-positive and Gram-negative bacteria, respectively, could be different, thus causing a quicker elimination of *E. coli* than of *M. luteus* (65). These two hypotheses are part of a more general strategy that leads the immune system of insects to selectively activate specific responses to minimize the cost of the immune defense (66). Different reaction times against specific foreign agents can thus be the result of this cost minimization criterion, which, in turn, can sometimes reduce the efficacy of the immune response. The lack of a bactericidal action of cell-free hemolymph against *M. luteus* and the need for hemocytes to support the immune response and counteract this pathogen is confirmed by the marked increase in circulating cells, starting 6 hours after the infection, which is maintained for longer times.

Although hemocytes are responsible for producing humoral molecules (67, 68), they are mainly involved in phagocytosis, encapsulation, and nodulation (17). In Diptera, phagocytosis is triggered very rapidly after an immune challenge [from 5 to 15 minutes from the infection depending on the type of bacteria (69–72)] and can last up to 24 hours (70, 71). Here, we showed that phagocytosis is rapidly activated by both Gram-positive and Gram-negative bacteria, also in *H. illucens* larvae, as confirmed by pHrodo-conjugated bacteria, which are phagocytosed by circulating hemocytes only 5 minutes after the infection. In contrast to other Diptera such as *Armigeres subalbatus* and *Aedes aegypti* (70, 71), the engulfment activity in *H. illucens* larvae declines rapidly and is absent at the longer time points (i.e., 24 hours). This difference could be attributed to the infection protocol as, here, we injected a defined and limited concentration of Gram-positive and Gram-negative bacteria in the larvae instead of using an "uncontrolled" pricking infection (70, 71). However, a more intriguing hypothesis could be the existence in BSF of phagocytes that are quicker and more efficient in removing bacteria than in other Diptera as a possible adaptation of these larvae to substrates with high bacterial loads, which require a very rapid and robust cellular immune response to counteract potential, dangerous infections.

We showed that the kinetics of encapsulation and melanin deposition in BSF are affected by the non-self-antigen. In fact, as in other Diptera and Lepidoptera (73, 74), melanotic capsule formation strongly depends on the matrix, with agarose beads that are never melanized although recognized and surrounded by hemocytes, in contrast to dextran beads, which are rapidly encapsulated and melanized. Moreover, a correlation between the bead charge and the degree of melanin deposition has been observed. In particular, positive-charged beads elicited the strongest response in terms of melanin deposition, which was also observed in other Diptera (75) and Lepidoptera (74, 76). As

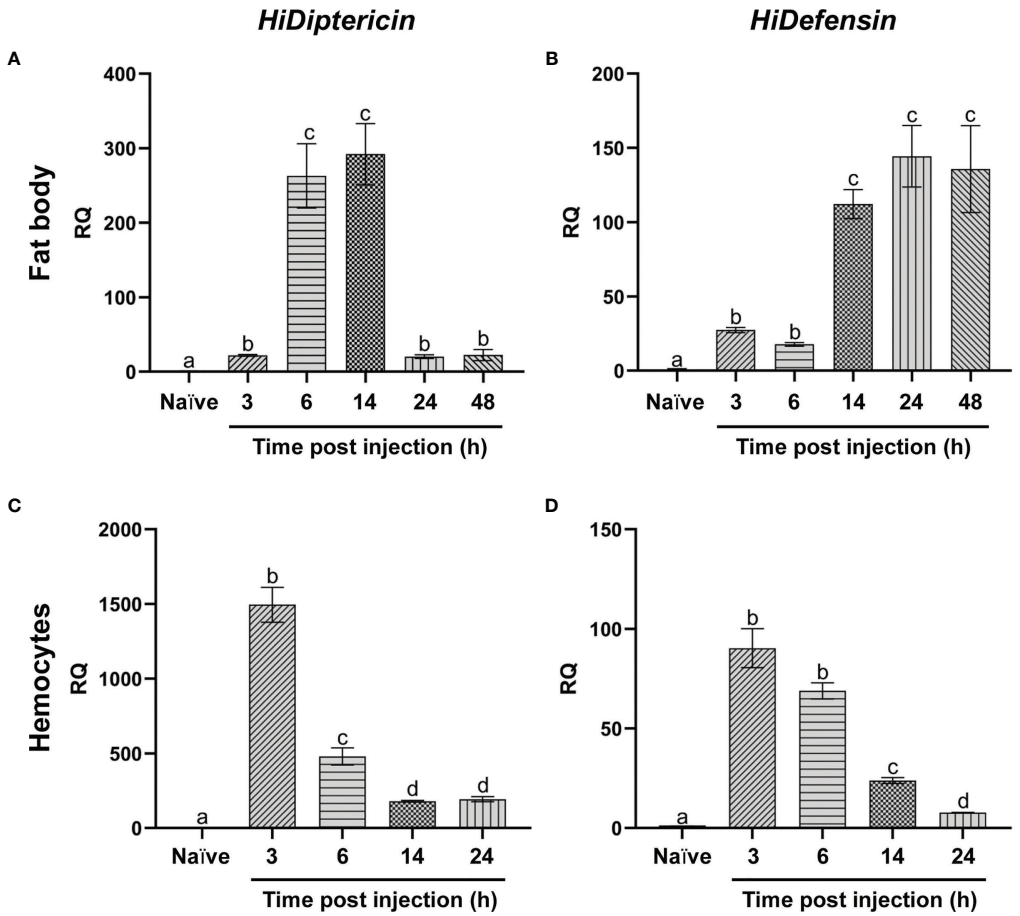


FIGURE 7 | qRT-PCR analysis of antimicrobial peptides. **(A, C)** mRNA levels of *HiDiptericin* in the fat body **(A)** and hemocytes **(C)** of naïve and infected larvae at different time points; **(B, D)** mRNA levels of *HiDefensin* in the fat body **(B)** and hemocytes **(D)** of naïve and infected larvae at different time points. Naïve larvae starved for 3 hours were used as controls. Values represent mean ± s.e.m. Different letters indicate statistically significant differences among treatments (One-Way ANOVA: **A**) $F_{5\text{-}12} = 178.8$, $p < 0.0001$; **B**) $F_{5\text{-}12} = 127$, $p < 0.0001$; **C**) $F_{4\text{-}10} = 532.6$, $p < 0.0001$; **D**) $F_{4\text{-}10} = 658.3$, $p < 0.0001$).

encapsulation in insects usually involves the joint action of different hemocyte populations (i.e., granulocytes, involved in the release of lectins for PO activation, actively recruit plasmatocytes, which leads to capsule formation) (77, 78), ultrastructural analyses of the encapsulated beads are mandatory to address how the three hemocyte populations reported in *H. illucens* larvae (16) take part in this process. Moreover, our results may indicate how BSF larvae react to larger parasites characterized by different surface charge (79).

The lack of activation of the PO system at low bacterial concentrations indicates that melanization depends on the bacterial load in the hemocoel and here we have demonstrated the existence of a threshold level over which melanin is produced. The inhibition of melanin formation has been observed in immunized *Galleria mellonella* larvae (80), too, and it could be mediated by the presence in the hemolymph of active inhibitors of the enzymatic cascade, such as serpins and CLSP2 protein (81, 82). These inhibitors, which are only triggered when the PO cascade is activated, keep melanin formation under strict control, preventing an unnecessary production/overproduction of this pigment, which can be highly toxic for the insect itself (83–85). Experiments with hemolymph from naïve larvae indicate that such regulatory mechanisms exist in *H. illucens*. In fact, these inhibitors are present in their inactive form since the PO system has not been activated yet: once isolated from the larval body, simple exposure to the air activates the PO system, which is not subjected to such negative regulation, as previously seen in *Calliphora* larvae (86). Therefore, when *H. illucens* larvae experience a limited bacterial infection, phagocytosis, antimicrobial peptide production, and lysozyme activity are sufficient to counteract the threat, inhibiting at the same time the activation of the PO system. Conversely, in the presence of a significant bacterial load in the hemocoel, these cellular and humoral systems are unable to efficiently contain the bacterial spread and the PO system is activated, leading to melanin production and release in the larval body (**Figure 8**). In BSF an additional inhibitory effect on this enzyme system could be

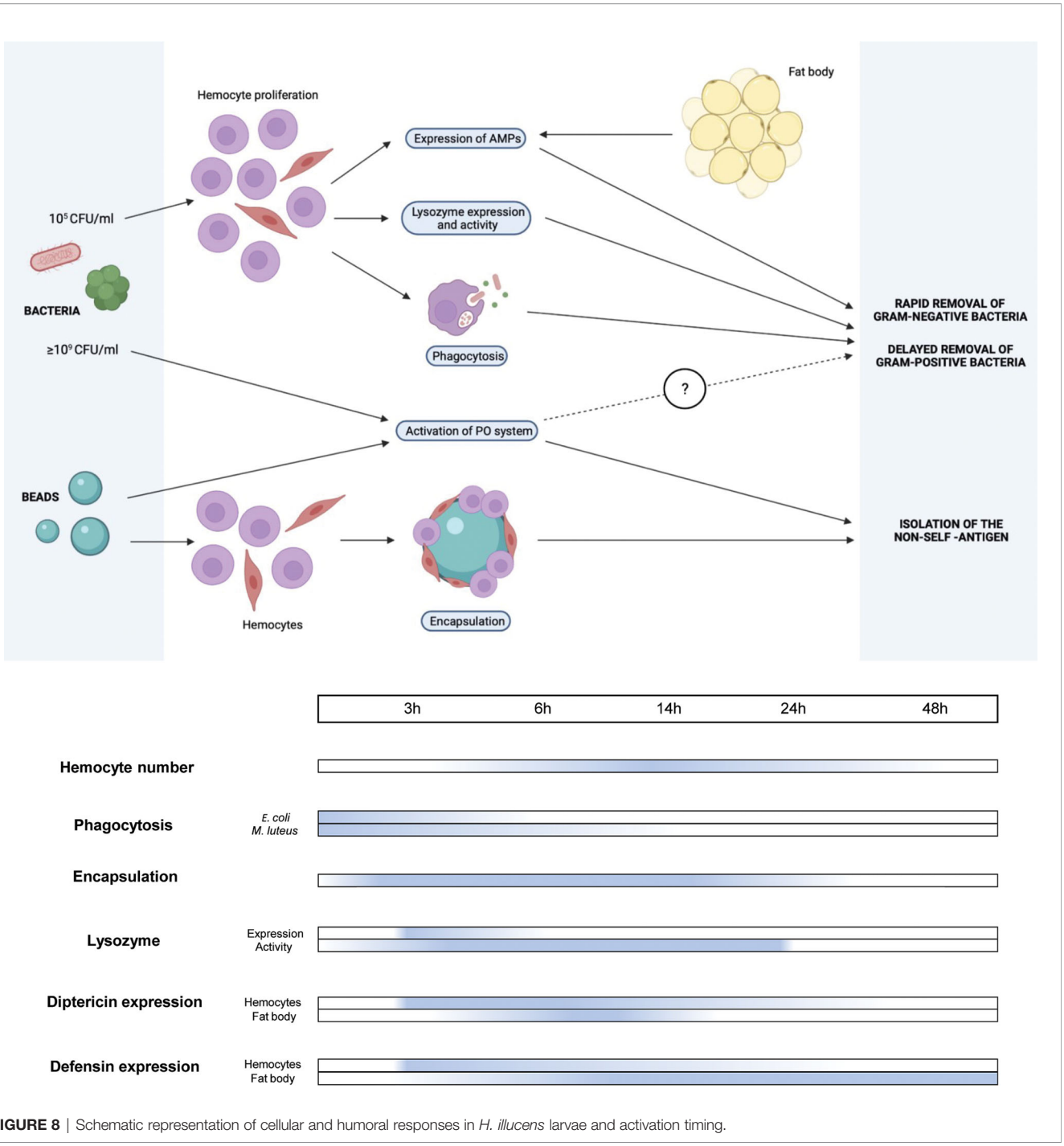


FIGURE 8 | Schematic representation of cellular and humoral responses in *H. illucens* larvae and activation timing.

exerted by lysozyme and AMPs, two known negative regulators of PO (87, 88), as they increase in parallel with inhibition of the PO system.

Interestingly, expression levels of *HiLysozyme* in the fat body did not increase after injecting bacteria, suggesting that the constitutive activity of this enzyme can counteract the bacterial threat, at least at the beginning of the infection (89). Moreover, the initial degradation of bacteria mediated by the constitutive activity of lysozyme could be useful for generating PAMPs, which stimulate the host immune system to obtain a stronger response against non-self-antigens (90). A temporally limited increase in mRNA levels was registered in hemocytes, indicating that circulating cells are the main source of lysozyme production during infection (91, 92) and their synthesizing activity is needed to maintain high levels of lysozyme in the hemolymph. Accordingly, the activity of this enzyme also remained elevated at longer times after the infection (24 hours), although no more bacterial colonies were detected. This pattern could be useful to

the larva to promptly respond to a second infection, suggesting the existence of immune priming in BSF (93).

In contrast to lysozyme, the transcription of both AMPs was highly stimulated by infection both in the fat body and in hemocytes. In particular, the early increase in AMP mRNA levels in circulating cells (3-6 hours), probably required to start challenging the infection, is rapidly supported by a later, but prolonged (6-48 hours) production of AMPs in the fat body, which represents the main organ involved in their synthesis (94). Interestingly, while the transcription of Diptericin, which is active only against Gram-negative bacteria (95), returns to low levels 24 hours after the infection, the trend observed for Defensin, which is active against Gram-positive bacteria (96), is different. In fact, although no colonies of *M. luteus* are found 24 hours after the infection, the expression levels of Defensin remain high up to 48 hours. This pattern may be due to the enhanced ability of Gram-positive bacteria to counteract the immune response of *H. illucens*; thus, high levels of this AMP are maintained to better counteract a potential, second infection.

Insect immunity with its cellular and humoral arms has evolved under substantial selective pressure to protect the host against foreign invaders (17, 97). The biological significance of the present study lies in its more detailed appreciation of how the speed and kinetics of immune activation influence the ability of the host (i.e., BSF larvae, whose immune system is poorly studied despite the great interest in this insect for its bioconversion ability) to counteract a specific category of invaders (i.e., Gram-positive and Gram-negative bacteria). The data reported herein highlight that the two branches of the immune system have different activation dynamics in this insect as phagocytosis and encapsulation are rapidly triggered once the larva is exposed to the foreign antigen, while the humoral components intervene later. This evidence - although important because it fills a gap in our knowledge of BSF larvae - is common among insects, that are able to activate different defense mechanisms against infections *via* different speeds, specificities, and routes (17, 97). The possible variation in numbers and populations of circulating hemocytes, and the rate of their recruitment to overcome the intruders likely affect the extent and time course of phagocytosis and encapsulation among different insects and within the same species depending on the foreign invader (17). Similarly, the variation in the evoking potency of the bacterial cell wall PAMPs, e.g., peptidoglycan or lipopolysaccharides of Gram-positive and Gram-negative bacteria, respectively, and the efficiency of the subsequent response mechanisms may contribute, in part, to the speed and specificity of immune phenomena in different challenged insects, e.g., Orthoptera (98), Lepidoptera (99), Hymenoptera (100), Coleoptera (101) and other Diptera (102). Moreover, the infection route itself can also affect the timing and expression of AMPs. Here, we employed a method to directly deliver pathogens to the larva *via* injection, but it will be necessary to investigate the effects of oral administration of bacteria as well. In fact, injection of the sand fly *Lutzomyia longipalpis* with *M. luteus* or *Serratia marcescens* did not result in increased Defensin expression until 72 hours later (103). Conversely, oral infection strongly induced an acute antimicrobial response by upregulating Defensin expression in both larval and adult stages. Thus, viewed through the lens of evolution, the immune system of modern insects may be so exquisite and fine-tuned to the point of providing apparently paradoxical and contradictory observations.

Summarizing, this study: i) provides new information on the BSF immune system and expands knowledge on the defense mechanisms of Diptera, ii) represents a platform of knowledge for future studies on the immune response of BSF larvae challenged with other infectious agents, such as fungi or viruses, and iii) is a prerequisite to manipulating the larval immune response by nutritional (i.e., the diet) (9) or environmental (i.e., rearing temperature or larval density) (104, 105) factors, to increase resistance to pathogens and optimize health status during mass rearing.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

DB, MC, and GT designed the study. DB, AuM, MM, and AG performed the research. DB, AuM, MB, MC, and GT analyzed the data. DB, AuM, MFB, AmM, LT, MC, and GT wrote the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Fondazione Cariplo (grant number 2020-0900).

ACKNOWLEDGMENTS

The authors are indebted to Sir Norman Ratcliffe, Prof. Bernard Duvic and Prof. David Stanley for their valuable suggestions on the manuscript. We thank Sherryl Sundell for English editing. AuM is a Ph.D. student of the 'Life Science and Biotechnology' course at University of Insubria. The authors appreciated the support from Consorzio Interuniversitario per le Biotecnologie (CIB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.745160/full#supplementary-material

REFERENCES

1. Cappellozza S, Leonardi MG, Savoldelli S, Carminati D, Rizzolo A, Cortellino G, et al. A First Attempt to Produce Proteins From Insects by Means of a Circular Economy. *Animals* (2019) 9:278. doi: 10.3390/ani9050278
2. Lopes IG, Lalander C, Vidotti RM, Vinnerås B. Using *Hermetia illucens* Larvae to Process Biowaste From Aquaculture Production. *J Clean Prod* (2020) 251:119753. doi: 10.1016/j.jclepro.2019.119753
3. Saadoun JH, Montevecchi G, Zanasi L, Bortolini S, Macavei LI, Masino F, et al. Lipid Profile and Growth of Black Soldier Flies (*Hermetia illucens*, Stratiomyidae) Reared On By-Products From Different Food Chains. *J Sci Food Agric* (2020) 100:3648–557. doi: 10.1002/jsfa.10397
4. Truzzi C, Giorgini E, Annibaldi A, Antonucci M, Illuminati S, Scarponi G, et al. Fatty Acids Profile of Black Soldier Fly (*Hermetia illucens*): Influence of Feeding Substrate Based on Coffee-Waste Silverskin Enriched With Microalgae. *Anim Feed Sci Technol* (2020) 259:114309. doi: 10.1016/j.anifeedsci.2019.114309
5. Hawkey KJ, Lopez-Viso C, Brameld JM, Parr T, Salter AM. Insects: A Potential Source of Protein and Other Nutrients for Feed and Food. *Annu Rev Anim Biosci* (2021) 9:333–54. doi: 10.1146/annurev-animal-021419-083930
6. Barbi S, Messori M, Manfredini T, Pini M, Montorsi M. Rational Design and Characterization of Bioplastics From *Hermetia illucens* Prepupae Proteins. *Biopolymers* (2019) 110:e23250. doi: 10.1002/bip.23250
7. Wang H, Rehman KU, Liu X, Yang Q, Zheng L, Li W, et al. Insect Biorefinery: A Green Approach for Conversion of Crop Residues Into Biodiesel and Protein. *Biotechnol Biofuels* (2017) 10:304. doi: 10.1186/s13068-017-0986-7
8. Jayanegara A, Haryati RP, Nafisah A, Suptijah P, Ridla M, Laconi EB. Derivatization of Chitin and Chitosan From Black Soldier Fly (*Hermetia illucens*) and Their Use as Feed Additives: An *In Vitro* Study. *Adv Anim Vet Sci* (2020) 8:472–7. doi: 10.17582/journal.aavs/2020/8.5.472.477
9. Vogel H, Müller A, Heckel DG, Gutzeit H, Vilcinskas A. Nutritional Immunology: Diversification and Diet-Dependent Expression of Antimicrobial Peptides in the Black Soldier Fly *Hermetia illucens*. *Dev Comp Immunol* (2018) 78:141–8. doi: 10.1016/j.dci.2017.09.008
10. Spranghers T, Ottoboni M, Klootwijk C, Ovyn A, Deboosere S, De Meulenaer B, et al. Nutritional Composition of Black Soldier Fly (*Hermetia illucens*) Prepupae Reared on Different Organic Waste Substrates. *J Sci Food Agric* (2017) 97:2594–600. doi: 10.1002/jsfa.8081
11. Barragan-Fonseca KB, Dicke M, van Loon JJA. Influence of Larval Density and Dietary Nutrient Concentration on Performance, Body Protein, and Fat Contents of Black Soldier Fly Larvae (*Hermetia illucens*). *Entomol Exp Appl* (2018) 166:761–70. doi: 10.1111/eea.12716
12. Fabrikov D, Morote E, Montes J, Sánchez-Muros MJ, Barroso FG, Rodríguez-Rodríguez M, et al. Facing the Challenge of Discarded Fish: Improving Nutritional Quality of Two Insect Species Larvae for Use as Feed and Food. *J Insects Food Feed* (2021) 7:1–12. doi: 10.3920/JIFF2020.0019
13. Lalander CH, Diener S, Magri ME, Zurbrügg C, Lindström A, Vinnerås B. Faecal Sludge Management With the Larvae of the Black Soldier Fly (*Hermetia illucens*) - From a Hygiene Aspect. *Sci Total Environ* (2013) 458-460:312–8. doi: 10.1016/j.scitotenv.2013.04.033
14. Lalander CH, Fidjeland J, Diener S, Eriksson S, Vinnerås B. High Waste-To-Biomass Conversion and Efficient *Salmonella* Spp. Reduction Using Black Soldier Fly for Waste Recycling. *Agron Sustain Dev* (2015) 35:261–71. doi: 10.1007/s13593-014-0235-4
15. Huang Y, Yu Y, Zhan S, Tomberlin JK, Huang D, Cai M, et al. Dual Oxidase Duox and Toll-Like Receptor 3 TLR3 in the Toll Pathway Suppress Zoonotic Pathogens Through Regulating the Intestinal Bacterial Community Homeostasis in *Hermetia illucens* L. *PLoS One* (2020) 15:e0225873. doi: 10.1371/journal.pone.0225873
16. Zdybicka-Barabas A, Bulak P, Polakowski C, Bieganowski A, Waśko A, Cytryńska M. Immune Response in the Larvae of The Black Soldier Fly *Hermetia illucens*. *ISJ Invertebr Surviv J* (2017) 14:9–17. doi: 10.25431/1824-307X/isj.v14i1.9-17
17. Eleftherianos I, Heryanto C, Bassal T, Zhang W, Tettamanti G, Mohamed A. Hemocyte-Mediated Immunity in Insects: Cells, Processes, and Associated Components in the Fight Against Pathogens and Parasites. *Immunology* (2021) 164:401–32. doi: 10.1111/imm.13390
18. Zhang W, Tettamanti G, Bassal T, Heryanto C, Eleftherianos I, Mohamed A. Regulators and Signalling in Insect Antimicrobial Innate Immunity: Functional Molecules and Cellular Pathways. *Cell Signal* (2021) 83:110003. doi: 10.1016/j.cellsig.2021.110003
19. Binda-Rossetti S, Mastore M, Protasoni M, Brivio MF. Effects of an Entomopathogen Nematode on the Immune Response of the Insect Pest Red Palm Weevil: Focus on the Host Antimicrobial Response. *J Invertebr Pathol* (2016) 133:110–9. doi: 10.1016/j.jip.2015.11.001
20. Tsakas S, Marmaras VJ. Insect Immunity and Its Signalling: An Overview. *ISJ Invertebr Surviv J* (2010) 7:228–38.
21. Rosales C. Cellular and Molecular Mechanisms of Insect Immunity. In: Shields VDC, editor. *Insect Physiology and Ecology*, London, UK: InTech Publication (2017). p. 179–212.
22. Hwang S, Bang K, Lee J, Cho S. Circulating Hemocytes From Larvae of the Japanese Rhinoceros Beetle *Allomyrina dichotoma* (iinnaeus) (Coleoptera: Scarabaeidae) and the Cellular Immune Response to Microorganisms. *PLoS One* (2015) 10:e0128519. doi: 10.1371/journal.pone.0128519
23. Al Mutawa MY, Ayaad TH, Shaurub EH. Hemocyte Profile, Phagocytosis, and Antibacterial Activity in Response to Immune Challenge of the Date Fruit Stalkborer, *Oryctes elegans*. *Invertebr Surviv J* (2020) 17:147–62. doi: 10.25431/1824-307X/isj.v0i0
24. Sigle LT, Hillyer JF. Mosquito Hemocytes Preferentially Aggregate and Phagocytose Pathogens in the Periostial Regions of the Heart That Experience the Most Hemolymph Flow. *Dev Comp Immunol* (2016) 55:90–101. doi: 10.1016/j.dci.2015.10.018
25. Evans JD, Aronstein K, Chen YP, Hetru C, Imler JL, Jiang H, et al. Immune Pathways and Defence Mechanisms in Honey Bees *Apis mellifera*. *Insect Mol Biol* (2006) 15:645–56. doi: 10.1111/j.1365-2583.2006.00682.x
26. Spivak M, Danka RG. Perspectives on Hygienic Behavior in *Apis mellifera* and Other Social Insects. *Apidologie* (2021) 52:1–16. doi: 10.1007/s13592-020-00784-z
27. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, et al. Immunity and Other Defenses in Pea Aphids, *Acyrthosiphon pisum*. *Genome Biol* (2010) 11:R21. doi: 10.1186/gb-2010-11-2-r21
28. Altincicek B, Gross J, Vilcinskas A. Wounding-Mediated Gene Expression and Accelerated Viviparous Reproduction of the Pea Aphid *Acyrthosiphon pisum*. *Insect Mol Biol* (2008) 17:711–6. doi: 10.1111/j.1365-2583.2008.00835.x
29. Sackton TB, Lazzaro BP, Clark AG. Rapid Expansion of Immune-Related Gene Families in the House Fly, *Musca domestica*. *Mol Biol Evol* (2017) 34:857–72. doi: 10.1093/molbev/msw285
30. Choi WH, Choi H-J, Goo TW, Quan F-S. Novel Antibacterial Peptides Induced by Probiotics in *Hermetia illucens* (Diptera: Stratiomyidae) Larvae. *Entomol Res* (2018) 48:237–47. doi: 10.1111/1748-5967.12259
31. Alvarez D, Wilkinson KA, Treilhou M, Téné N, Castillo D, Sauvain M. Prospecting Peptides Isolated From Black Soldier Fly (Diptera: Stratiomyidae) With Antimicrobial Activity Against *Helicobacter pylori* (Campylobacterales: Helicobacteraceae). *J Insect Sci* (2019) 19:17. doi: 10.1093/jisesa/iez120
32. Lee K-S, Yun E-Y, Goo T-W. Antimicrobial Activity of an Extract of *Hermetia illucens* Larvae Immunized With *Lactobacillus casei* Against *Salmonella* Species. *Insects* (2020) 11:704. doi: 10.3390/insects11100704
33. Xu J, Luo X, Fang G, Zhan S, Wu J, Wang D, et al. Transgenic Expression of Antimicrobial Peptides From Black Soldier Fly Enhance Resistance Against Entomopathogenic Bacteria in the Silkworm, *Bombyx mori*. *Insect Biochem Mol Biol* (2020) 127:103487. doi: 10.1016/j.ibmb.2020.103487
34. De Smet J, Wynants E, Cos P, Van Campenhout L. Microbial Community Dynamics During Rearing of Black Soldier Fly Larvae (*Hermetia illucens*) and Impact on Exploitation Potential. *Appl Environ Microbiol* (2018) 84:e02722–17. doi: 10.1128/AEM.02722-17
35. Perotti MA, Lysyk TJ, Kalischuk-Tymensen LD, Yanke LJ, Selinger LB. Growth and Survival of Immature *Heamatobia irritans* (Diptera: Muscidae) Is Influenced by Bacterial Isolated From Cattle Manure and Conspecific Larvae. *J Med Entomol* (2001) 38:180–7. doi: 10.1603/0022-2585-38.2.180
36. Erickson MC, Islam M, Sheppard C, Liao J, Doyle MP. Reduction of *Escherichia coli* O157:H7 and *Salmonella enterica* Serovar Enteritidis in

Chicken Manure by Larvae of the Black Soldier Fly. *J Food Prot* (2004) 67:685–90. doi: 10.4315/0362-028x-67.4.685
37. Pimentel AC, Montali A, Bruno D, Tettamanti G. Metabolic Adjustment of the Larval Fat Body in *Hermetia illucens* to Dietary Conditions. *J Asia Pac Entomol* (2017) 20:1307–13. doi: 10.1016/j.aspen.2017.09.017
38. Hogsette JA. New Diets for Production of House Flies and Stable Flies. *J Econ Entomol* (1992) 85:2291–304. doi: 10.1093/jee/85.6.2291
39. Bruno D, Bonelli M, De Filippis F, Di Lelio I, Tettamanti G, Casartelli M, et al. The Intestinal Microbiota of *Hermetia illucens* Larvae Is Affected by Diet and Shows a Diverse Composition in the Different Midgut Regions. *Appl Environ Microbiol* (2019) 85:e01864–18. doi: 10.1128/AEM.01864-18
40. Mastore M, Binda Rossetti S, Giovannardi S, Scarì G, Brivio MF. Inducible Factors With Antimicrobial Activity After Immune Challenge in the Haemolymph of Red Palm Weevil (Insecta). *Innate Immun* (2015) 21:392–405. doi: 10.1177/1753425914542446
41. Mastore M, Brivio MF. Basic Methods to Evaluate Humoral Immunity Processes in Lepidoptera Larvae. In: Sandrelli F, Tettamanti G, editors. *Immunity in Insects*. New York: Springer Protocols Handbooks (2020). p. 127–44.
42. Liégeois LS, Wang W, Ferrandon D. Methods to Quantify *In Vivo* Phagocytic Uptake and Opsonization of Live or Killed Microbes in *Drosophila melanogaster*. In: Sandrelli F, Tettamanti G, editors. *Immunity in Insects*. New York: Springer Protocols Handbooks (2020). p. 79–95.
43. Mastore M, Brivio MF. Cuticular Surface Lipids Are Responsible for Disguise Properties of an Entomoparasite Against Host Cellular Responses. *Dev Comp Immunol* (2008) 32:1050–62. doi: 10.1016/j.dci.2008.02.003
44. Garriga A, Mastore M, Morton A, Del Pino FG, Brivio MF. Immune Response of *Drosophila suzukii* Larvae to Infection With the Nematobacterial Complex *Steinernema carpocapsae-Xenorhabdus nematophila*. *Insects* (2020) 11:210. doi: 10.3390/insects11040210
45. Harvey D. Spectroscopic Methods. In: Harvey D, editor. *Analytical Chemistry 2.1*. LibreTexts, DePauw University (2016). p. 517–616.
46. Bonelli M, Bruno D, Brilli M, Gianfranceschi N, Tian L, Tettamanti G, et al. Black Soldier Fly Larvae Adapt to Different Food Substrates Through Morphological and Functional Responses of the Midgut. *Int J Mol Sci* (2020) 21:4955. doi: 10.3390/ijms21144955
47. Romanelli D, Casartelli M, Cappellozza S, de Eguileor M, Tettamanti G. Roles and Regulation of Autophagy and Apoptosis in the Remodeling of the Lepidopteran Midgut Epithelium During Metamorphosis. *Sci Rep* (2016) 6:32939. doi: 10.1038/srep32939
48. Bonelli M, Bruno D, Caccia S, Sgambetterra G, Cappellozza S, Jucker C, et al. Structural and Functional Characterization of *Hermetia illucens* Larval Midgut. *Front Physiol* (2019) 10:204. doi: 10.3389/fphys.2019.00204
49. Bruno D, Bonelli M, Cadamuro AG, Reguzzoni M, Grimaldi A, Casartelli M, et al. The Digestive System of the Adult *Hermetia illucens* (Diptera: Stratiomyidae): Morphological Features and Functional Properties. *Cell Tissue Res* (2019) 378:221–38. doi: 10.1007/s00441-019-03025-7
50. Bruno D, Bonacci T, Reguzzoni M, Casartelli M, Grimaldi A, Tettamanti G, et al. An In-Depth Description of Head Morphology and Mouthparts in Larvae of the Black Soldier Fly *Hermetia illucens*. *Arthropod Struct Dev* (2020) 58:100969. doi: 10.1016/j.asd.2020.100969
51. Gold M, Egger J, Scheidegger A, Zurbrügg C, Bruno D, Bonelli M, et al. Estimating Black Soldier Fly Larvae Biowaste Conversion Performance by Simulation of Midgut Digestion. *Waste Manage* (2020) 112:40–51. doi: 10.1016/j.wasman.2020.05.026
52. Kotzé RCM, Muller N, du Plessis L, van der Horst G. The Importance of Insect Sperm: Sperm Ultrastructure of *Hermetia illucens* (Black Soldier Fly). *Tissue Cell* (2019) 59:44–50. doi: 10.1016/j.tice.2019.06.002
53. Malawey AS, Mercati D, Love CC, Tomberlin J. Adult Reproductive Tract Morphology and Spermatogenesis in the Black Soldier Fly (Diptera: Stratiomyidae). *Ann Entomol Soc Am* (2019) 112:1–11. doi: 10.1093/aesa/saz045
54. Barros-Cordeiro KB, Báo SN, Pujol-Luz JR. Intra-Puparial Development of the Black Soldier-Fly, *Hermetia illucens*. *J Insect Sci* (2014) 14:83. doi: 10.1093/jis/14.1.83
55. Birrel NW. Sensory Morphology, Investment and Plasticity in the Black Soldier Fly (*Hermetia illucens*). [PhD Thesis]. [Auckland]: University of Auckland (2018). doi: 10.31237/osf.io/v8zs2
56. Paulk A, Gilbert C. Proprioceptive Encoding of Head Position in the Black Soldier Fly, *Hermetia illucens* (L.) (Stratiomyidae). *J Exp Biol* (2016) 209:3913–24. doi: 10.1242/jeb.02438
57. Waśko A, Bulak P, Polak-Berecka M, Nowak K, Polakowski C, Bieganowski A. The First Report of the Physicochemical Structure of Chitin Isolated From *Hermetia illucens*. *Int J Biol Macromol* (2016) 92:316–20. doi: 10.1016/j.ijbiomac.2016.07.038
58. Pezzi M, Scapoli C, Bharti M, Faucheux MJ, Chicca M, Leis M, et al. Fine Structure of Maxillary Palps in Adults of *Hermetia illucens* (Diptera: Stratiomyidae). *J Med Entomol* (2020) 58:658–65. doi: 10.1093/jme/tjaa251
59. Park S-I, Chang BS, Yoe SM. Detection of Antimicrobial Substances From Larvae of the Black Soldier Fly, *Hermetia illucens* (Diptera: Stratiomyidae). *Entomol Res* (2014) 44:58–64. doi: 10.1111/1748-5967.12050
60. Elrod-Erickson S, Schneider D, Mishr M. Interactions Between the Cellular and Humoral Immune Responses in *Drosophila*. *Curr Biol* (2000) 10:781–4. doi: 10.1016/s0960-9822(00)00569-8
61. Haine ER, Moret Y, Siva-Jothy MT, Rolff J. Antimicrobial Defense and Persistent Infection in Insects. *Science* (2008) 322:1257–9. doi: 10.1126/science.1165265
62. Abi Khattar Z, Rejasse A, Destoumieux-Garzón D, Escoubas JM, Sanchis V, Lereclus D, et al. The Dlt Operon of *Bacillus cereus* Is Required for Resistance to Cationic Antimicrobial Peptides and for Virulence in Insects. *J Bacteriol* (2009) 191:7063–73. doi: 10.1128/JB.00892-09
63. Assoni L, Milani B, Carvalho MR, Nepomuceno LN, Waz NT, Souza Guerra ME, et al. Resistance Mechanisms to Antimicrobial Peptides in Gram-Positive Bacteria. *Front Microbiol* (2020) 11:593215. doi: 10.3389/fmicb.2020.593215
64. Nawrocki KL, Crispell EK, McBride SM. Antimicrobial Peptide Resistance Mechanisms of Gram-Positive Bacteria. *Antibiotics-Basel* (2014) 3:461–92. doi: 10.3390/antibiotics3040461
65. Tanji T, Hu X, Weber ANR, Ip YT. Toll and IMD Pathways Synergistically Activate an Innate Immune Response in *Drosophila melanogaster*. *Mol Cell Biol* (2007) 27:4578–88. doi: 10.1128/MCB.01814-06
66. Moret Y. Explaining Variable Costs of the Immune Response: Selection for Specific Versus Non-Specific Immunity and Facultative Life History Change. *Oikos* (2003) 102:213–6. doi: 10.1034/j.1600-0706.2003.12496.x
67. Lavine MD, Chen G, Strand MR. Immune Challenge Differentially Affects Transcript Abundance of Three Antimicrobial Peptides in Hemocytes From the Moth *Pseudoplusia includens*. *Insect Biochem Mol Biol* (2005) 35:1335–46. doi: 10.1016/j.ibmb.2005.08.005
68. Ma H, Abbas MN, Zhang K, Hu X, Xu M, Liang H, et al. 20-Hydroxyecdysone Regulates the Transcription of the Lysozyme *Via* Broad-Complex Z2 Gene in Silkworm, *Bombyx mori*. *Dev Comp Immunol* (2019) 94:66–72. doi: 10.1016/j.dci.2019.01.014
69. Hernández-Martínez S, Lanz H, Rodríguez MH, González-Ceron L, Tsutsumi V. Cellular-Mediated Reactions to Foreign Organisms Inoculated Into the Hemocoel of *Anopheles albimanus* (Diptera: Culicidae). *J Med Entomol* (2002) 39:61–9. doi: 10.1603/0022-2585-39.1.61
70. Hillyer JF, Schmidt SL, Christensen BM. Hemocyte-Mediated Phagocytosis and Melanization in the Mosquito *Armigeres subalbatus* Following Immune Challenge by Bacteria. *Cell Tissue Res* (2003) 313:117–27. doi: 10.1007/s00441-003-0744-y
71. Hillyer JF, Schmidt SL, Christensen BM. Rapid Phagocytosis and Melanization of Bacteria and Plasmodium Sporozoites by Hemocytes of the Mosquito *Aedes aegypti*. *J Parasitol* (2003) 89:62–9. doi: 10.1645/0022-3395(2003)089[0062:RPAMOB]2.0.CO;2
72. Rowley AF, Ratcliffe NA. An Ultrastructural Study of the *In Vitro* Phagocytosis of *Escherichia coli* by the Hemocytes of *Calliphora erythrocephala*. *J Ultrastruct Res* (1976) 55:193–202. doi: 10.1016/s0022-5320(76)80066-4
73. Gorman MJ, Schwartz AM, Paskewitz SM. The Role of Surface Characteristics in Eliciting Humoral Encapsulation of Foreign Bodies in plasmodium-Refractory and -Susceptible Strains of *Anopheles gambiae*. *J Insect Physiol* (1998) 44:947–54. doi: 10.1016/s0022-1910(98)00056-0
74. Lavine MD, Strand MR. Surface Characteristics of Foreign Targets That Elicit an Encapsulation Response by the Moth *Pseudoplusia includens*. *J Insect Physiol* (2001) 47:965–74. doi: 10.1016/s0022-1910(01)00071-3

75. Paskewitz S, Riehle MA. Response of *Plasmodium* Refractory and Susceptible Strains of *Anopheles gambiae* to Inoculated Sephadex Beads. *Dev Comp Immunol* (1994) 18:369–75. doi: 10.1016/0145-305x(94)90002-7
76. Vinson SB. The Role of the Foreign Surface and Female Parasitoid Secretions on the Immune Response of an Insect. *Parasitology* (1974) 68:27–33. doi: 10.1017/S0031182000045340
77. Pech LL, Strand MR. Granular Cells Are Required for Encapsulation of Foreign Targets by Insect Haemocytes. *J Cell Sci* (1996) 109:2053–60. doi: 10.1242/jcs.109.8.2053
78. Yu X-Q, Kanost MR. Immulectin-2, a Pattern Recognition Receptor That Stimulates Hemocyte Encapsulation and Melanization in the Tobacco Hornworm, *Manduca sexta*. *Dev Comp Immunol* (2004) 28:891–900. doi: 10.1016/j.dci.2004.02.005
79. Wootton EC, Dyrynda EA, Ratcliffe NA. Interaction Between Non-Specific Electrostatic Forces and Humoral Factors in Haemocyte Attachment and Encapsulation in the Edible Cockle, *Cerastoderma edule*. *J Exp Biol* (2006) 209:1326–35. doi: 10.1242/jeb.02118
80. Pye AE. Activation of Prophenoloxidase and Inhibition of Melanization in the Haemolymph of Immune *Galleria mellonella* Larvae. *Insect Biochem* (1978) 8:117–23. doi: 10.1016/0020-1790(78)90048-3
81. Shin SW, Zou Z, Raikhel AS. A New Factor in the *Aedes aegypti* Immune Response: CLSP2 Modulates Melanization. *EMBO Rep* (2011) 12:938–43. doi: 10.1038/embor.2011.130
82. Kanost MR, Jiang H, Yu X-Q. Innate Immune Responses of a Lepidopteran Insect, *Manduca sexta*. *Immunol Rev* (2004) 198:97–105. doi: 10.1111/j.0105-2896.2004.0121.x
83. Zlotkin E, Gurevitz M, Shulov A. The Toxic Effects of Phenoloxidases From the Haemolymph of Tenebrionid Beetles. *J Insect Physiol* (1973) 19:1057–65. doi: 10.1016/0022-1910(73)90031-0
84. An C, Budd A, Kanost MR, Michel K. Characterization of a Regulatory Unit That Controls Melanization and Affects Longevity of Mosquitoes. *Cell Mol Life Sci* (2011) 68:1929–39. doi: 10.1007/s00018-010-0543-z
85. Lu A, Zhang Q, Zhang J, Yang B, Wu K, Xie W, et al. Insect Prophenoloxidase: The View Beyond Immunity. *Front Physiol* (2014) 11:252. doi: 10.3389/fphys.2014.00252
86. Thomson JA, Sin YT. The Control of Prophenoloxidase Activation in Larval Haemolymph of *Calliphora*. *J Insect Physiol* (1970) 16:2063–74. doi: 10.1016/0022-1910(70)90079-x
87. Zdybicka-Barabas A, Mak P, Jakubowicz T, Cytryńska M. Lysozyme and Defense Peptides as Suppressors of Phenoloxidase Activity in *Galleria mellonella*. *Arch Insect Biochem Physiol* (2014) 87:1–12. doi: 10.1002/arch.21175
88. Rao X-J, Ling E, Yu X-Q. The Role of Lysozyme in the Prophenoloxidase Activation System of *Manduca sexta*: An *In Vitro* Approach. *Dev Comp Immunol* (2010) 34:264–71. doi: 10.1016/j.dci.2009.10.004
89. Dunn PE, Dai W, Kanost MR, Geng CX. Soluble Peptidoglycan Fragments Stimulate Antibacterial Protein Synthesis by Fat Body From Larvae of *Manduca sexta*. *Dev Comp Immunol* (1985) 9:559–68. doi: 10.1016/0145-305x(85)90019-9
90. Ragland SA, Criss AK. From Bacterial Killing to Immune Modulation: Recent Insights Into the Functions of Lysozyme. *PLoS Pathog* (2017) 13: e1006512. doi: 10.1371/journal.ppat.1006512
91. Wilson R, Ratcliffe NA. Effect of Lysozyme on the Lectin-Mediated Phagocytosis of *Bacillus cereus* by Haemocytes of the Cockroach, *Blaberus discoidalis*. *J Insect Physiol* (2000) 46:663–70. doi: 10.1016/S0022-1910(99)00154-7
92. Dorrah MA, Mohamed AA, Shaurub E-SH. Immunosuppressive Effects of the Limonoid Azadirachtin, Insights on a Nongenotoxic Stress Botanical, in Flesh Flies. *Pest Biochem Physiol* (2019) 153:55–66. doi: 10.1016/j.pestbp.2018.11.004
93. Sheehan G, Farrell G, Kavanagh K. Immune Priming: The Secret Weapon of the Insect World. *Virulence* (2020) 11:238–46. doi: 10.1080/21505594.2020.1731137
94. Yakovlev AY, Nesin AP, Simonenko NP, Gordya NA, Tulin DV, Kruglikova AA, et al. Fat Body and Hemocyte Contribution to the Antimicrobial Peptide Synthesis in *Calliphora vicina* R.-D. (Diptera: Calliphoridae) Larvae. *In Vitro Cell Dev Biol Anim* (2017) 53:33–42. doi: 10.1007/s11626-016-0078-1
95. Wu Q, Patočka J, Kuča K. Insect Antimicrobial Peptides, a Mini Review. *Toxins* (2018) 10:461. doi: 10.3390/toxins10110461
96. Koehbach J. Structure-Activity Relationships of Insect Defensins. *Front Chem* (2017) 5:45. doi: 10.3389/fchem.2017.00045
97. Eleftherianos I, Zhang W, Heryanto C, Mohamed A, Contreras G, Tettamanti G, et al. Diversity of Insect Antimicrobial Peptides and Proteins - A Functional Perspective: A Review. *Int J Biol Macromol* (2021) 191:277–87. doi: 10.1016/j.ijbiomac.2021.09.082
98. Adamo SA. Estimating Disease Resistance in Insects: Phenoloxidase and Lysozyme-Like Activity and Disease Resistance in the Cricket *Gryllus texensis*. *J Insect Physiol* (2004) 50:209–16. doi: 10.1016/j.jinsphys.2003.11.011
99. Morishima I, Horiba T, Iketani M, Nishioka E, Yamano Y. Parallel Induction of Cecropin and Lysozyme in Larvae of the Silkworm, *Bombyx mori*. *Dev Comp Immunol* (1995) 19:357–63. doi: 10.1016/0145-305x(95)00019-p
100. Erler S, Popp M, Lattorff HMG. Dynamics of Immune System Gene Expression Upon Bacterial Challenge and Wounding in a Social Insect (*Bombus terrestris*). *PLoS One* (2011) 6:e18126. doi: 10.1371/journal.pone.0018126
101. Jo YH, Patnaik BB, Hwang J, Park KB, Ko HJ, Kim CE, et al. Regulation of the Expression of Nine Antimicrobial Peptide Genes by *Tm*IMD Confers Resistance Against Gram-Negative Bacteria. *Sci Rep* (2019) 9:10138. doi: 10.1038/s41598-019-46222-8
102. Mohamed AA, Ali MM, Dorrah MA, Bassal TTM. Mediation of Inducible Nitric Oxide and Immune-Reactive Lysozymes Biosynthesis by Eicosanoid and Biogenic Amines in Flesh Flies. *Int J Trop Insect Sci* (2018) 38:93–104. doi: 10.1017/S1742758417000315
103. Telleria EL, Sant'Anna MRV, Alkurbi MO, Pitaluga AN, Dillon RJ, Traub-Csekö YM. Bacterial Feeding, *Leishmania* Infection and Distinct Infection Routes Induce Differential Defensin Expression in *Lutzomyia longipalpis*. *Parasit Vectors* (2013) 6:12. doi: 10.1186/1756-3305-6-12
104. Takano Y, Sakamoto T, Tabunoki H, Yoshimura J, Iwabuchi K. Integrated Effects of Thermal Acclimation and Challenge Temperature on Cellular Immunity in the Plusiine Moth Larvae *Chrysodeixis eriosoma* (Lepidoptera: Noctuidae). *Physiol Entomol* (2020) 46:52–9. doi: 10.1111/phen.12344
105. Karlsson Green K. The Effects of Host Plant Species and Larval Density on Immune Function in the Polyphagous Moth *Spodoptera littoralis*. *Ecol Evol* (2021) 11:10090–7. doi: 10.1002/ece3.7802

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor IE has declared a past co-authorship with one of the authors (GT) at the time of review.

frontiers in Immunology

MINI REVIEW
published: 06 December 2021
doi: 10.3389/fimmu.2021.771065

Regulation of Plant Immunity by Nuclear Membrane-Associated Mechanisms

Yiling Fang[1,2] *and Yangnan Gu*[1,2*]

[1] Department of Plant and Microbial Biology, University of California, Berkeley, CA, United States, [2] Innovative Genomics Institute, University of California, Berkeley, CA, United States

Unlike animals, plants do not have specialized immune cells and lack an adaptive immune system. Instead, plant cells rely on their unique innate immune system to defend against pathogens and coordinate beneficial interactions with commensal and symbiotic microbes. One of the major convergent points for plant immune signaling is the nucleus, where transcriptome reprogramming is initiated to orchestrate defense responses. Mechanisms that regulate selective transport of nuclear signaling cargo and chromatin activity at the nuclear boundary play a pivotal role in immune activation. This review summarizes the current knowledge of how nuclear membrane-associated core protein and protein complexes, including the nuclear pore complex, nuclear transport receptors, and the nucleoskeleton participate in plant innate immune activation and pathogen resistance. We also discuss the role of their functional counterparts in regulating innate immunity in animals and highlight potential common mechanisms that contribute to nuclear membrane-centered immune regulation in higher eukaryotes.

Keywords: nuclear envelope (NE), nuclear pore complex (NPC), nuclear transport receptors (NTRs), nucleoskeletal proteins, innate immune system, plant Immunity, nucleocytoplasmic continuum, nuclear lamina

OPEN ACCESS

Edited by:
Bostjan Kobe,
The University of Queensland, Australia

Reviewed by:
Mary Christie,
The University of Sydney, Australia
Susan Breen,
University of Warwick, United Kingdom

***Correspondence:**
Yangnan Gu
guyangnan@berkeley.edu

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *05 September 2021*
Accepted: *18 November 2021*
Published: *06 December 2021*

Citation:
Fang Y and Gu Y (2021) Regulation of Plant Immunity by Nuclear Membrane-Associated Mechanisms. Front. Immunol. 12:771065. doi: 10.3389/fimmu.2021.771065

INTRODUCTION

The innate immune system in plants and animals arose independently but converged with a similar set of molecular tools for pathogen perception (1, 2). Higher plants possess an enormous number of surface-localized as well as intracellular-distributed immune receptors that recognize a variety of immunological signals associated with pathogen infections. Surface-localized plant immune receptors consist mostly of plasma membrane-anchored receptor kinases and receptor-like proteins that are conceptually analogous to Toll-like receptors in animals. They are referred to as pattern recognition receptors (PRRs) and are able to detect specific microbe-associated molecular patterns (MAMPs, e.g. bacterial flagellin and lipopolysaccharides) or host-derived, damage-associated molecular patterns (DAMPs, e.g. cutin and apoplastic peptide fragments) through their extracellular leucine-rich repeat domains (3, 4). Once activated, PRRs engage diverse signaling cascades including mitogen-activated protein kinase (MAPK)-, Ca^{2+}-, and reactive oxygen species (ROS)-mediated signaling to activate pattern-triggered immunity (PTI), a prominent and first layer of active immune response in plants (5–7). However, some pathogens evolved functionally versatile effector proteins, which are delivered into plant cells to target critical immune regulators and compromise PTI signaling, leading to effector-triggered susceptibility (ETS) (8). For example, the effector protein AvrPto from

bacterial pathogen *Pseudomonas syringae* binds PRRs (e.g. FLAGELLIN SENSITIVE 2) and their signaling partner (e.g. BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED KINASE 1) to block PTI signaling (9, 10). To counteract ETS, a group of intracellular immune receptors that belong to the nucleotide-binding, leucine-rich repeat (NLR) superfamily evolved to activate the second layer of plant immunity. Pathogen effectors can be directly or indirectly recognized by cognate NLRs, which activate a strong and robust immune response termed effector-triggered immunity (ETI) (11, 12). Activated plant NLRs are assembled into resistosomes (13, 14), which form permeable calcium channels on the plasma membrane to activate defense (15) or mediate NAD+ cleavage to promote cell death (16–18).

PTI and ETI appear to have distinct early signaling but share many common downstream immune regulators to coordinate transcriptional reprogramming toward defense activation. Recent advances also revealed that PTI and ETI are inter-dependent and mutually enhanced to achieve full resistance against pathogens (19–21). Immune activation in local infected cells or tissues can further induce systemic acquired resistance (SAR), an immune mechanism that primes neighboring and distal cells and tissues for an enhanced and broad-spectrum resistance against future pathogen infection. Induction of SAR relies on phytohormone signaling including salicylic acid (SA) and jasmonic acid (JA) (11, 22). Activation of PTI, ETI, and SAR all requires efficient and highly regulated nucleocytoplasmic exchange of signals. As the major communication interface between the cytoplasm and the nucleus, the nuclear envelope (NE) evolved as a critical platform to integrate, decode, transmit, and respond to immune signals. In this review, we focus on discussing the functions of NE-associated proteins in regulating plant innate immunity and comparing their roles in NE-based immune regulatory mechanisms between plants and animals.

EXTENSIVE INVOLVEMENTS OF NUCLEOPORINS IN PLANT IMMUNE REGULATION

Nuclear pore complexes (NPC) perforate the double-layered NE and provide physical access for molecules to travel in or out of the nucleus. Each NPC is a mega protein complex and is assembled by 500~1,000 nucleoporin proteins (Nups) of ~40 different kinds (23–25). These nucleoporins build different NPC modules, including the inner ring complex (IRC) and the outer ring complex (ORC) that together form an octagonal symmetric core scaffold, the membrane ring composed by transmembrane nucleoporins that anchors the core scaffold to the nuclear pore membrane, the central channel barrier composed by phenylalanine-glycine (FG)-rich nucleoporins that fills the core scaffold and mediate selective cargo transport, and the nuclear basket and cytoplasmic filaments that extrude from the core scaffold and establish connections with the nucleoplasmic and cytoplasmic contents, respectively (26–28) (**Figure 1**). Extensive evidence supports that NPC integrity is essential for activating innate immune responses in both plants and animals. Interestingly though, different nucleoporins appear to play distinct roles in immune regulation and show functional specificity.

The Arabidopsis *SUPPRESSOR OF NPR1-1, CONSTITUTIVE 1 (SNC1)* gene encodes an NLR protein that undergoes nuclear translocation for immune signaling, and enhanced nuclear accumulation of SNC1 leads to autoimmune activation (29, 30). Elegant genetic studies showed that several conserved ORC scaffolding nucleoporins, including MOS3 (MODIFIER OF SNC1 3)/Nup96, Nup160, and Seh1, are required for SNC1-mediated autoimmune induction as well as basal resistance against *P. syringae* (31, 32) (**Figure 1**). In animals, Nup96 was shown to be important for both innate and adaptive immune pathways. Defects in Nup96 were reported to impair interferon-γ-mediated induction of major histocompatibility complexes (MHC I and MHC II), which are essential for antigen presentation. Mutations in *Nup96* also diminished MHC-dependent T cell proliferation in mice (33). Although molecular mechanisms behind Nup96-dependent immune regulation remain largely obscure in both animals and plants, the ORC is essential for the proper assembly of the NPC core scaffold and bulk mRNA export (34–36). It is conceivable that loss of ORC nucleoporins may lead to major structural and functional defects in the NPC, which compromise nucleocytoplasmic exchange of critical immune signals and the newly transcribed immune-related mRNA population.

The FG nucleoporins that contain intrinsically disordered FG-rich regions distribute along the central channel of the NPC and form the molecular barrier to enable selective cargo transport. At least six FG nucleoporins have been reported to play a role in regulating the innate immune response in Arabidopsis (37). Among them, Nup98, a conserved FG nucleoporin, has been characterized in detail. Arabidopsis encodes two Nup98 paralogs, Nup98a and Nup98b, both of which interact with a conserved non-FG nucleoporin Nup88/MOS7. A partial loss of function mutation in Arabidopsis *Nup88* (*mos7-1*) results in compromised resistance against both bacterial and fungal pathogens. The *mos7-1* mutation specifically attenuates the nuclear accumulation of a series of important plant immune regulators, including a key ETI signaling protein ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1), the master regulator of SA-mediated defense response NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1 (NPR1), MITOGEN-ACTIVATED PROTEIN KINASE 3 (MPK3), and SNC1, but not other nuclear proteins such as histone protein and transcription factor CELL DIVISION CYCLE 5 (CDC5) and TGA2 (38, 39) (**Figure 1**). It was hypothesized that the *mos7-1* mutation disrupts the Nup98-Nup88 interaction, which alters the NPC permeability and leads to disruption of the nuclear transport of immune-related cargoes. Consistent with this hypothesis, both *nup98a* and *nup98b* mutants are more susceptible to the fungal pathogen *Botrytis cinerea* than WT plants (39), suggesting a specialized function of Nup98 in plant immunity. In line with the observation in Arabidopsis, a rice Nup98 homolog, APIP12, was shown to be important for basal resistance against the rice

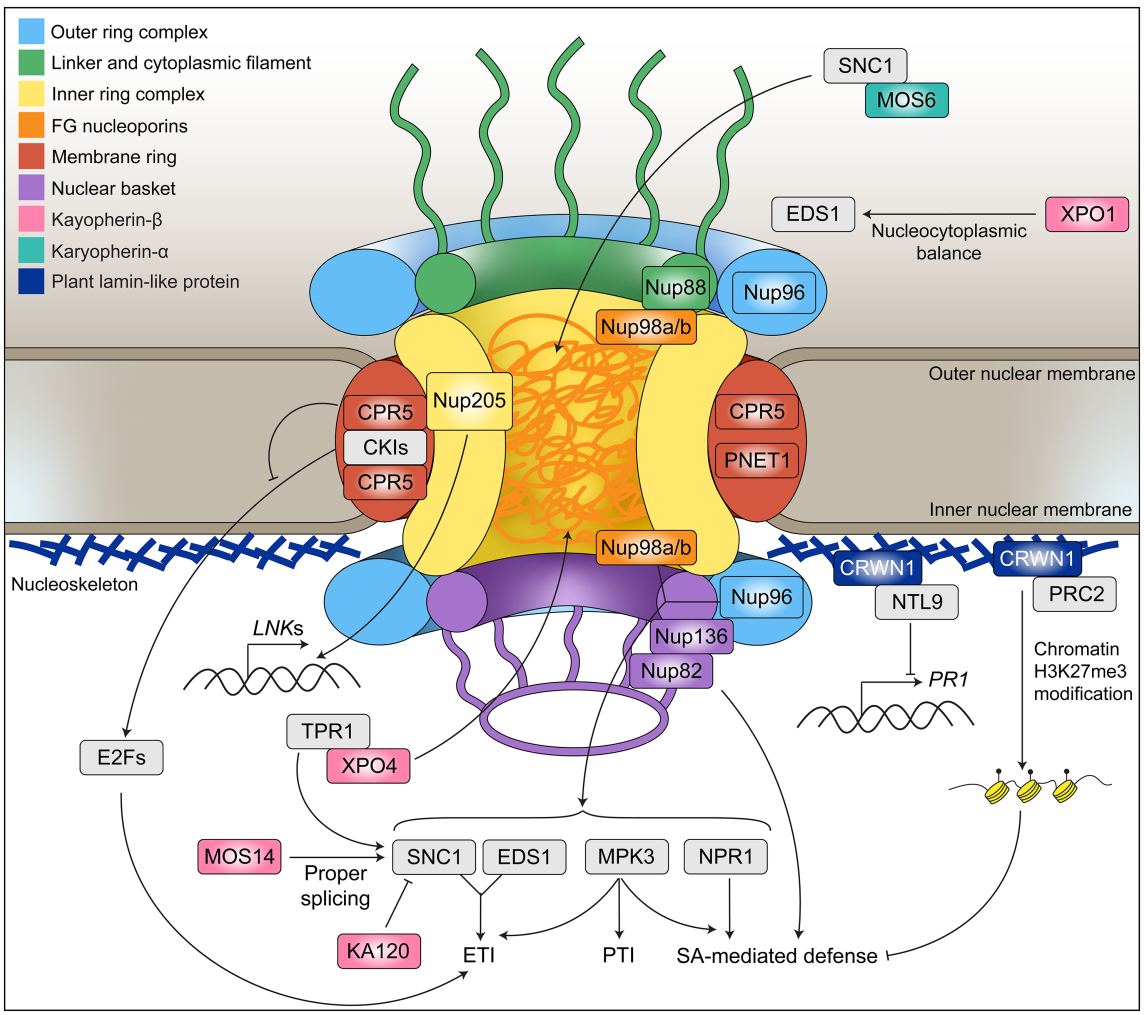


FIGURE 1 | Summary of the regulatory roles of nuclear membrane-associated proteins in plant immune pathways. Critical plant immune regulators that are subject to regulation by nuclear membrane-associated mechanisms include the NLR-type immune receptor SNC1 that activates effector-triggered immunity (ETI), the critical ETI signaling component EDS1 that acts downstream of SNC1, transcription corepressor TPL and TPRs that work together with SNC1 to activate defense gene expression, cell cycle transcription factor E2Fs that activate ETI in a noncanonical manner, the master regulator of SA-mediated immunity NPR1, the mitogen-activated protein kinase MPK3 that widely participates in various types of plant immune responses including PTI, ETI, and basal resistance, core clock regulator LNKs that are required for basal resistance, and the transcription factor NTL9 that represses defense gene (e.g. *PR1*) expression. The outer ring complex (ORC) of the nuclear pore complex (NPC) is required for SNC1-mediated immune activation and basal resistance. As components of the central channel barrier, FG nucleoporin Nup98 and Nup88 are required for nuclear accumulation of SNC1, EDS1, NPR1, and MPK3. The membrane nucleoporin CPR5 (part of the NPC membrane ring) gates ETI activation by sequestering cyclin kinase inhibitor SIM/SMR1 and preventing E2F activation. Nup205 is required for the proper expression of *LNK* family genes and activation of basal resistance. Nup82 and Nup136 are involved in SA-mediated defense activation. Karyopherin-α MOS6 serves as the possible importin adapter for multiple NLR proteins, including SNC1 and TN13. Karyopherin-β KA120 is required for suppressing the nuclear activity of SNC1 and prevents SNC1 autoimmune activation. XPO4 mediates TPR1 nuclear export and negatively regulates SA-mediated immune amplification during ETI. MOS14 is required for proper splicing of NLR genes, including SNC1. The nucleoskeleton protein CRWN1 plays a role in suppressing both SA responses through interacting with transcription factor NTL9 that represses defense gene *PR1* expression and interacting with epigenetic regulator PRC2 that facilitates the H3K27me3 of genes promoting SA synthesis.

blast fungus *Magnaporthe oryzae*. Notably, an *M. oryzae* effector AvrPiz-t evolved to specifically target APIP12 to enhance pathogen virulence (40), highlighting its functional significance in basal defense activation. Knocking out *APIP12* results in an increase in lesion density associated with *M. oryzae* infection. In Drosophila, Nup98 was reported to play a role in defense against RNA viral infection. However, rather than functioning at the NPC, *Dm*Nup98 is re-located from the NPC to the nucleoplasm upon viral infection and directly binds promoters of antiviral genes to facilitate their expression (41).

Both Nup98 and Nup96 discussed above are functionally conserved nucleoporins and are encoded by the same gene and

post-translationally processed into two proteins in animals and fungi (42, 43). However, in plants, except *Chlamydomonas reinhardtii* and *Physcomitrella patens*, Nup98 and Nup96 are encoded by separate genes, which are, more often than not, located on different chromosomes. Arabidopsis encodes one Nup96 and two Nup98 proteins as mentioned above, and *Oryza sativa* encodes one Nup96 and three AtNup98a-like proteins, including APIP12. Duplication of *Nup98* likely occurred from mosses as *Physcomitrella patens* contains two copies of *Nup98*, one encoded with *Nup96* and the other located on a separate chromosome. Therefore, it appears that *Nup98* not only separated from *Nup96* but also gradually expanded in the plant lineage during evolution. Although Arabidopsis *Nup98* paralogs were reported to be functionally redundant in regulating flowering time and starch degradation (44, 45), the duplication of *Nup98* may potentially drive differential assembly of the NPC and allow specific regulation of immune-related cargo transport under distinct environmental/stress conditions. Indeed, Arabidopsis *Nup98a* and *Nup98b* display partially different expression patterns under various developmental and pathological contexts (e-Northern Expression Browser in bar.utoronto.ca) (46). However, whether different Nup98 paralogs are collaboratively or differentially incorporated into the NPC awaits future investigation.

In addition to Nup98 and Nup96, a recent study by de Leone et al. reported that perturbing the IRC nucleoporin Nup205 attenuates plant immunity by affecting the expression of a core clock gene family *NIGHT LIGHT-INDUCIBLE AND CLOCK-REGULATED (LNK*) during pathogen infection (47) (**Figure 1**). This finding is consistent with the emerging concept that the circadian rhythm is integrated with plant immune regulatory pathways (48, 49) and underscores a role of the NPC in this process. The nucleoporin's influence on clock genes was also observed in Drosophila. Two interacting nuclear basket nucleoporins, Tpr (Megator) and Nup153, set the pace of the clock in Drosophila by regulating the protein ubiquitylation/stability and nuclear translocation of the core oscillator PERIOD and TIMELESS (50, 51). Though the roles of Tpr and Nup153 during immune activation in flies are not clear, they appear to be involved in the insertion of HIV-1 DNA into the human genome region that is close to the nuclear pore, a location known to associate with transcriptionally active chromatin (52). Consistently, depletion of Tpr inhibited HIV gene expression in infected cells (53). Additionally, the HIV-1 capsid is physically associated with Nup153, which is required to facilitate the viral import into the nucleus (53, 54).

Although the structure and composition of the NPC are highly conserved among eukaryotes, plants have evolved several specific nucleoporins, which, interestingly, all appear to play a role in plant immune responses. Constitutive Expressor of Pathogenesis-Related Genes 5 (CPR5), a plant unique nucleoporin localized in the NPC membrane module, is involved in gating ETI activation in Arabidopsis. CPR5 switches from oligomer to monomer upon NLR immunoreceptor activation and releases the cyclin-dependent kinase inhibitors SIAMESE and SIAMESE-RELATED 1 to trigger a noncanonical activation of cell cycle transcription factor E2Fs that can drive defense gene expression (55–57) (**Figure 1**). CPR5 was recently found to interact with PLANT NUCLEAR ENVELOPE TRANSMEMBRANE 1 (PNET1), a novel plant nucleoporin that also exists in animals (58). Interestingly, the function of the human PNET1 homolog has also been linked with the cell cycle and was found to promote the proliferation of lung cancer cells (59). Cell cycle regulators are involved in both effector-triggered cell death in plants and apoptosis in mammals (60, 61). Because both CPR5 and PNET1 connect with cell cycle regulatory pathways, the interaction between these two proteins would be interesting for further investigation to reveal the interplay between the cell cycle and immune pathway at the nuclear pore. Another angiosperm-specific nucleoporin, Nup82, interacts with its homolog Nup136 at the nuclear basket. They are redundantly required for SA-dependent immune responses. The *nup82 nup136* double mutant plants are impaired in benzothiadiazole (an analog of SA)-induced resistance to *P. syringae* (62). It is worth noting that the NPC nuclear basket is intimately associated with the nucleoskeleton (63) and that mutations in nucleoskeleton genes also cause altered SA responses in Arabidopsis (64). The influence of nucleoskeleton on plant immune will be further discussed later.

THE ROLE OF NUCLEAR TRANSPORT RECEPTORS AND THEIR CARGO SPECIFICITY IN MODULATING PLANT IMMUNITY

The selective transport of macromolecules through the NPC mainly depends on karyopherin proteins, a superfamily of nuclear transport receptors (NTRs), including importins and exportins. Karyopherin-α, also known as importin-α, is an adaptor protein that connects nuclear localization signal (NLS)-containing protein cargo with a karyopherin-β protein (65). Karyopherin-β proteins are capable of directly interacting with FG nucleoporins and mediate the translocation of the importin-α/cargo complex across the NPC (66, 67). Karyopherin-β can also directly interact with some cargo molecules and mediate their nuclear shuttling independent of a karyopherin-α (68–70). Depending on the transport direction, karyopherin-βs are categorized into importins and exportins. The karyopherin-β family was thought to be established before the divergence of eukaryotic life and has been largely maintained since. In contrast, importin-α proteins diverged early in eukaryotes, evolved within lineages, and have expanded dramatically with the complexity of organisms (71). Accumulating evidence suggests that different karyopherins may have evolved to mediate the transport of specific cargo populations and thus participate in regulating distinct cellular processes (65, 69, 72). Multiple karyopherin-β proteins in Arabidopsis have been reported to function in plant immune regulation, and this role seems mostly attributed to their cargo specificity.

Proximity-labeling proteomics identified transcription corepressors TOPLESS (TPL) and TPL-RELATED proteins (TPRs), which are negative regulators of defense activation, as specific cargo of Exportin-4 (XPO4) in Arabidopsis. XPO4 facilitates the nuclear export of TPR1 in the presence of high levels of SA to negatively regulate the SA-mediated immune amplification (**Figure 1**). Consistently, loss of *XPO4* dramatically enhances *cpr5*-dependent autoimmune activation in Arabidopsis (73). Intriguingly, *xpo4* single mutants are compromised in basal resistance and NLR-mediated immune activation, suggesting distinct regulatory roles of this exportin at multiple levels of plant immune induction.

Mutations in another karyopherin-β, *KA120*, leads to a typical NLR-dependent auto-immune response in the absence of pathogens. KA120 is required to restrict the nuclear activity of SNC1 and is essential to prevent SNC1 autoactivation in the nucleus (**Figure 1**). Overexpression of KA120 efficiently suppresses SNC1-dependent autoimmunity and knocking down *SNC1* partially rescues the *ka120*-associated autoimmune phenotype (74). However, whether SNC1 is a direct cargo of KA120 is yet to be determined. The human homolog of KA120, importin-11 (IPO11), has not been reported to be directly involved in immune regulation. However, IPO11 was found to be a tumor suppressor by maintaining the protein level of Phosphatase and Tensin Homologue (PTEN) (75). Loss of *IPO11* leads to degradation of PTEN, and deletion, mutation, and suppression of *PTEN* are often associated with cancers in various tissues (76).

MOS14 is a predicted importin that transports serine/arginine-rich (SR) proteins (77), which are required for the recognition of splice sites and spliceosome assembly. The loss-of-function *mos14* mutation alters the splicing pattern of multiple NLR genes, including *SNC1*, which leads to compromised immunity (78) (**Figure 1**). Aside from importing SR proteins, the human homolog of MOS14, TNPO3, is also responsible for HIV nuclear import. TNPO3 can interact with the HIV1 integrase, which facilitates host integration of HIV-1 DNA, and disrupting this integrase-TNPO3 interaction can block HIV nuclear import (79, 80).

XPO1 is a conserved exportin across the eukaryotes that binds the leucine-rich nuclear export signal (NES) to mediate protein nuclear export. In animals, XPO1 can interact with NF-κB, a protein complex that is important to activate immune response during infection. The nuclear export of NF-κB is an essential step for preventing autoimmune diseases by controlling the level of activated NF-κB in the nucleus (81). In Arabidopsis, XPO1 is required for maintaining the nucleocytoplasmic balance of EDS1, which is important for ETI activation (82). Recently, Zhang etal. (83) reported that XPO1 is a susceptibility factor in Arabidopsis and *Nicotiana benthamiana* for resistance against the turnip mosaic virus (TuMV). XPO1 promotes TuMV infection by mediating the nuclear export of the viral replicase to facilitate its accumulation in the viral replication complexes outside of the nucleus. XPO1 also exports other sumoylated host factors to attenuate host immune responses (83).

Besides karyopherin-βs, certain importin-αs also show a specific role in immune regulation. For example, importin-α3/MOS6 has been implicated as a specific import adapter for SNC1 (84) (**Figure 1**). Loss of *MOS6*, but not any of the other eight importin-αs encoded by Arabidopsis, partially suppresses the autoimmune phenotype induced by SNC1. Moreover, SNC1 preferably interacts with MOS6 among all importin-αs in coimmunoprecipitation assays performed using a heterologous transient expression system (84). However, SNC1 harbors only weakly predicted NLS sequences, and whether these sequences are required for SNC1-MOS6 interaction is not clear. MOS6 also interacts with another NLR protein TIR-NB13 (TN13), implying a specialized function of MOS6 in mediating the nuclear transport of NLR proteins (85).

In animals, targeting NTRs has been used therapeutically to alleviate symptoms or enhance immunity. For example, importin-α3 is responsible for transporting the transcription factor NF-κB into the nucleus to coordinate expression of inflammatory genes, and vitamin D supplementation can reduce the airway hyperresponsiveness and allergic airway inflammation in ovalbumin-sensitized and -challenged mice by downregulating the protein level of importin-α3 (86, 87). Inhibiting human XPO1 can enhance innate immunity by increasing nuclear accumulation of p62, which promotes expression of innate immune-related genes. This eventually leads to reduced Kaposi's sarcoma-associated herpesvirus (KSHV) lytic replication (88). Therefore, NTRs play a vital role in selectively maintaining the protein homeostasis and activity of key immune regulatory proteins in both plants and animals. We envision that targeted chemical or genetic manipulation of functionally specialized plant NTRs may enable us to specifically modulate nuclear transport kinetics and signaling strength to achieve desired plant immune outcomes without affecting the bulk nuclear transport.

THE NUCLEAR LAMINA IS AN EMERGING PLAYER IN NUCLEAR MEMBRANE-ASSOCIATED IMMUNE REGULATION

Lamin and lamin-like proteins together with other proteins associated with the nucleoskeleton form a dense fibrous protein meshwork that is located beneath the inner nuclear membrane (INM) and called the nuclear lamina (NL). NL components not only regulate nuclear structure and mechanics but also play essential roles in a variety of cellular processes, including cellular signaling, gene transcription, epigenetic regulation, cell cycle progression, and cell differentiation (89–91). Lamins and nuclear matrix constituent proteins (NMCPs) are the major constituent of the nucleoskeleton in animals and plants respectively. Lamins and NMCPs share a similar domain architecture with a central coiled-coil rod domain flanked by a short N-terminal head and a long C-terminal tail domain that contains an NLS sequence (92). However, lamins and NMCPs are likely to have evolved independently as part of the nucleoskeleton, as the origin of NMCPs occurred with the divergence of Charophyta (93). In Arabidopsis, NMCPs are

also known as CROWED NUCLEI (CRWN) proteins, including CRWN1, CRWN2, CRWN3, and CRWN4 (94, 95). Both lamins and CRWNs have been reported to play a role in regulating immune responses.

In Arabidopsis, mutations in *CRWN1* result in accumulation of JA, up-regulated JA-responsive gene expression, and enhanced resistance against necrotrophic pathogen *B. cinerea*. Interestingly, *crwn1* mutants also display compromised PTI but intact basal resistance against the hemibiotrophic bacterial pathogen *P. syringae*, suggesting a complicated role of CRWN1 in regulating plant responses against different types of pathogens (96). Adding more complexity, higher-order *crwn* mutants, including *crwn1 crwn2* and *crwn1 crwn4*, showed significantly enhanced SA-mediated transcriptional response, enhanced resistance against *P. syringae*, as well as activation of other abiotic stress responses (64). Consistently, a recent study showed that loss of *NMCP* genes in liverwort *Marchantia polymorpha* activates both biotic and abiotic transcriptome responses (97). These studies suggest that the nucleoskeleton integrity is critical for maintaining the homeostasis of plant stress responses. In the Arabidopsis *crwn1 crwn2* double mutant, the levels of histone H3 lysine 27 trimethylation (H3K27me3), which is often linked to transcriptional repression, are reduced near *SAR DEFICIENT 1* (*SARD1*) and *CALMODULIN BINDING PROTEIN 60-LIKE.G* (*CBP60g*), two genes that encode critical transcription factors to promote SA biosynthesis, providing a potential mechanism for the enhanced SA responses in *crwn* mutants (98). Consistently, it was demonstrated that CRWN1 is physically associated with Polycomb Repressive Complex 2 (PRC2), which mediates H3K27me3 modification (99) (**Figure 1**). In addition, the carboxyl terminus of CRWN1 interacts with and enhances the DNA-binding activity of NAC WITH TRANSMEMBRANE MOTIF1-LIKE9 (NTL9), a transcription factor involved in repressing the expression of the *PATHOGENESIS-RELATED1* (*PR1*) gene (100) (**Figure 1**). These findings reveal a critical and previously underappreciated role for the nucleoskeleton in regulating plant immunity at the epigenetic and transcriptional level.

In animals, lamin A/C plays a prominent role in multiple functional aspects of immune cells, including cell development and differentiation, nuclear permeability, and cell migration (101). For example, neutrophils can reduce the nuclear stiffness by almost completely losing their lamin A/C during differentiation (102, 103). This generates lobed nuclei in neutrophils, which facilitate neutrophils passing through narrow spaces during migration (104). On the other hand, overaccumulation of lamin A/C in adipose tissue macrophages promotes NF-κB nuclear translocation and increases the expression of proinflammatory genes that lead to the activation of innate immune response (105). Intriguingly, *P. syringae* infection led to a decline of the CRWN1 protein level within hours, and SA treatment also triggered rapid proteosome-dependent degradation of CRWN1 in Arabidopsis, potentially facilitating defense activation (100). These studies suggest that modulating nucleoskeleton dynamics may be a convergent strategy to orchestrate immune activities in eukaryotes.

CONCLUSIONS AND FUTURE PERSPECTIVE

Although not located at the frontline of host-microbe interactions, the nucleus is one of the most critical downstream convergent points for immune signaling. As the boundary of the nucleus, NE-associated core molecular machinery and mechanisms, including structurally and functionally conserved nucleoporins, nuclear transport receptors, and nucleoskeletal proteins, are intimately involved in coordinating immune activation and amplitude in plants (Summarized in **Figure 1**). Future investigations will further advance our understanding of critical questions that remain in the field, such as how nucleoporins may regulate different immune pathways and defense responses by differentially affecting the NPC function, what is the immune-related substrate spectrum of NTRs and how it is determined and regulated at the molecular level, and how other components of the nuclear lamina may influence chromatin activity at the biochemical and epigenetic level to determine defense gene expression. Moreover, we currently have very little information about how integral plant NE proteins, a population of transmembrane proteins with a variety of prominent functions and activities at the NE (e.g. calcium transport, mechanosensation, and chromatin organization), may participate in plant immune regulation. Recently, PNET2, a conserved INM transmembrane protein in Arabidopsis, was shown to interact with CRWN1 and play a role in regulating higher order chromatin architecture at the nuclear periphery and coordinating the activation of plant stress responses including immunity (106). A more comprehensive understanding of the plant NE protein composition and subsequent detailed functional investigation of NE proteins will provide new insight into the NE-centered immune regulatory mechanism in plants.

AUTHOR CONTRIBUTIONS

YF and YG wrote the article. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Arnon Graduate Fellowship (to YF) and the USDA National Institute of Food and Agriculture (HATCH project CA-B-PLB-0243-H), the National Science Foundation (MCB-2049931), Hellman Fellows Fund, and startup funds from Innovative Genomics Institute and University of California Berkeley (to YG).

ACKNOWLEDGMENTS

We apologize to colleagues whose research could not be cited due to space limitations.

REFERENCES

1. Jones JD, Dangl JL. The Plant Immune System. *Nature* (2006) 444:323–9. doi: 10.1038/nature05286
2. Jones JD, Vance RE, Dangl JL. Intracellular Innate Immune Surveillance Devices in Plants and Animals. *Science* (2016) 354(6316):aaf6395. doi: 10.1126/science.aaf6395
3. Couto D, Zipfel C. Regulation of Pattern Recognition Receptor Signalling in Plants. *Nat Rev Immunol* (2016) 16:537–52. doi: 10.1038/nri.2016.77
4. Boutrot F, Zipfel C. Function, Discovery, and Exploitation of Plant Pattern Recognition Receptors for Broad-Spectrum Disease Resistance. *Annu Rev Phytopathol* (2017) 55:257–86. doi: 10.1146/annurev-phyto-080614-120106
5. Bigeard J, Colcombet J, Hirt H. Signaling Mechanisms in Pattern-Triggered Immunity (PTI). *Mol Plant* (2015) 8:521–39. doi: 10.1016/j.molp.2014.12.022
6. Hou S, Liu Z, Shen H, Wu D. Damage-Associated Molecular Pattern-Triggered Immunity in Plants. *Front Plant Sci* (2019) 10:646. doi: 10.3389/fpls.2019.00646
7. Wang W, Feng B, Zhou JM, Tang D. Plant Immune Signaling: Advancing on Two Frontiers. *J Integr Plant Biol* (2020) 62:2–24. doi: 10.1111/jipb.12898
8. Toruno TY, Stergiopoulos I, Coaker G. Plant-Pathogen Effectors: Cellular Probes Interfering With Plant Defenses in Spatial and Temporal Manners. *Annu Rev Phytopathol* (2016) 54:419–41. doi: 10.1146/annurev-phyto-080615-100204
9. Shan L, He P, Li J, Heese A, Peck SC, Nurnberger T, et al. Bacterial Effectors Target the Common Signaling Partner BAK1 to Disrupt Multiple MAMP Receptor-Signaling Complexes and Impede Plant Immunity. *Cell Host Microbe* (2008) 4:17–27. doi: 10.1016/j.chom.2008.05.017
10. Xiang T, Zong N, Zou Y, Wu Y, Zhang J, Xing W, et al. Pseudomonas Syringae Effector AvrPto Blocks Innate Immunity by Targeting Receptor Kinases. *Curr Biol* (2008) 18:74–80. doi: 10.1016/j.cub.2007.12.020
11. Zhang J, Coaker G, Zhou JM, Dong X. Plant Immune Mechanisms: From Reductionistic to Holistic Points of View. *Mol Plant* (2020) 13:1358–78. doi: 10.1016/j.molp.2020.09.007
12. Zhou JM, Zhang Y. Plant Immunity: Danger Perception and Signaling. *Cell* (2020) 181:978–89. doi: 10.1016/j.cell.2020.04.028
13. Wang J, Hu M, Wang J, Qi J, Han Z, Wang G, et al. Reconstitution and Structure of a Plant NLR Resistosome Conferring Immunity. *Science* (2019) 364(6435):eaav5870. doi: 10.1126/science.aav5870
14. Wang J, Wang J, Hu M, Wu S, Qi J, Wang G, et al. Ligand-Triggered Allosteric ADP Release Primes a Plant NLR Complex. *Science* (2019) 364 (6435):eaav5868. doi: 10.1126/science.aav5868
15. Bi G, Su M, Li N, Liang Y, Dang S, Xu J, et al. The ZAR1 Resistosome Is a Calcium-Permeable Channel Triggering Plant Immune Signaling. *Cell* (2021) 184:3528–3541 e3512. doi: 10.1016/j.cell.2021.05.003
16. Horsefield S, Burdett H, Zhang X, Manik MK, Shi Y, Chen J, et al. NAD(+) Cleavage Activity by Animal and Plant TIR Domains in Cell Death Pathways. *Science* (2019) 365:793–9. doi: 10.1126/science.aax1911
17. Wan L, Essuman K, Anderson RG, Sasaki Y, Monteiro F, Chung EH, et al. TIR Domains of Plant Immune Receptors Are NAD(+)-Cleaving Enzymes That Promote Cell Death. *Science* (2019) 365:799–803. doi: 10.1126/science.aax1771
18. Duxbury Z, Wang S, Mackenzie CI, Tenthorey JL, Zhang X, Huh SU, et al. Induced Proximity of a TIR Signaling Domain on a Plant-Mammalian NLR Chimera Activates Defense in Plants. *Proc Natl Acad Sci USA* (2020) 117:18832–9. doi: 10.1073/pnas.2001185117
19. Ngou BPM, Ahn HK, Ding P, Jones JDG. Mutual Potentiation of Plant Immunity by Cell-Surface and Intracellular Receptors. *Nature* (2021) 592:110–5. doi: 10.1038/s41586-021-03315-7
20. Yuan M, Jiang Z, Bi G, Nomura K, Liu M, Wang Y, et al. Pattern-Recognition Receptors Are Required for NLR-Mediated Plant Immunity. *Nature* (2021) 592:105–9. doi: 10.1038/s41586-021-03316-6
21. Yuan M, Ngou BPM, Ding P, Xin XF. PTI-ETI Crosstalk: An Integrative View of Plant Immunity. *Curr Opin Plant Biol* (2021) 62:102030. doi: 10.1016/j.pbi.2021.102030
22. Spoel SH, Dong X. How do Plants Achieve Immunity? Defence Without Specialized Immune Cells. *Nat Rev Immunol* (2012) 12:89–100. doi: 10.1038/nri3141
23. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, et al. The Molecular Architecture of the Nuclear Pore Complex. *Nature* (2007) 450:695–701. doi: 10.1038/nature06405
24. Tamura K, Hara-Nishimura I. The Molecular Architecture of the Plant Nuclear Pore Complex. *J Exp Bot* (2013) 64:823–32. doi: 10.1093/jxb/ers258
25. Kim SJ, Fernandez-Martinez J, Nudelman I, Shi Y, Zhang W, Raveh B, et al. Integrative Structure and Functional Anatomy of a Nuclear Pore Complex. *Nature* (2018) 555:475–82. doi: 10.1038/nature26003
26. Knockenhauer KE, Schwartz TU. The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* (2016) 164:1162–71. doi: 10.1016/j.cell.2016.01.034
27. Beck M, Hurt E. The Nuclear Pore Complex: Understanding Its Function Through Structural Insight. *Nat Rev Mol Cell Biol* (2017) 18:73–89. doi: 10.1038/nrm.2016.147
28. Hampoelz B, Andres-Pons A, Kastritis P, Beck M. Structure and Assembly of the Nuclear Pore Complex. *Annu Rev Biophys* (2019) 48:515–36. doi: 10.1146/annurev-biophys-052118-115308
29. Mang HG, Qian W, Zhu Y, Qian J, Kang HG, Klessig DF, et al. Abscisic Acid Deficiency Antagonizes High-Temperature Inhibition of Disease Resistance Through Enhancing Nuclear Accumulation of Resistance Proteins SNC1 and RPS4 in Arabidopsis. *Plant Cell* (2012) 24:1271–84. doi: 10.1105/tpc.112.096198
30. Xu F, Cheng YT, Kapos P, Huang Y, Li X. P-Loop-Dependent NLR SNC1 can Oligomerize and Activate Immunity in the Nucleus. *Mol Plant* (2014) 7:1801–4. doi: 10.1093/mp/ssu097
31. Zhang Y, Li X. A Putative Nucleoporin 96 Is Required for Both Basal Defense and Constitutive Resistance Responses Mediated by Suppressor of Npr1-1,Constitutive 1. *Plant Cell* (2005) 17:1306–16. doi: 10.1105/tpc.104.029926
32. Wiermer M, Cheng YT, Imkampe J, Li M, Wang D, Lipka V, et al. Putative Members of the Arabidopsis Nup107-160 Nuclear Pore Sub-Complex Contribute to Pathogen Defense. *Plant J* (2012) 70:796–808. doi: 10.1111/j.1365-313X.2012.04928.x
33. Faria AM, Levay A, Wang Y, Kamphorst AO, Rosa ML, Nussenzveig DR, et al. The Nucleoporin Nup96 Is Required for Proper Expression of Interferon-Regulated Proteins and Functions. *Immunity* (2006) 24:295–304. doi: 10.1016/j.immuni.2006.01.014
34. Harel A, Orjalo AV, Vincent T, Lachish-Zalait A, Vasu S, Shah S, et al. Removal of a Single Pore Subcomplex Results in Vertebrate Nuclei Devoid of Nuclear Pores. *Mol Cell* (2003) 11:853–64. doi: 10.1016/S1097-2765(03)00116-3
35. Walther TC, Alves A, Pickersgill H, Loiodice I, Hetzer M, Galy V, et al. The Conserved Nup107-160 Complex Is Critical for Nuclear Pore Complex Assembly. *Cell* (2003) 113:195–206. doi: 10.1016/S0092-8674(03)00235-6
36. Parry G, Ward S, Cernac A, Dharmasiri S, Estelle M. The Arabidopsis SUPPRESSOR OF AUXIN RESISTANCE Proteins Are Nucleoporins With an Important Role in Hormone Signaling and Development. *Plant Cell* (2006) 18:1590–603. doi: 10.1105/tpc.106.041566
37. Li X, Gu Y. Structural and Functional Insight Into the Nuclear Pore Complex and Nuclear Transport Receptors in Plant Stress Signaling. *Curr Opin Plant Biol* (2020) 58:60–8. doi: 10.1016/j.pbi.2020.10.006
38. Cheng YT, Germain H, Wiermer M, Bi D, Xu F, Garcia AV, et al. Nuclear Pore Complex Component MOS7/Nup88 Is Required for Innate Immunity and Nuclear Accumulation of Defense Regulators in Arabidopsis. *Plant Cell* (2009) 21:2503–16. doi: 10.1105/tpc.108.064519
39. Genenncher B, Wirthmueller L, Roth C, Klenke M, Ma L, Sharon A, et al. Nucleoporin-Regulated MAP Kinase Signaling in Immunity to a Necrotrophic Fungal Pathogen. *Plant Physiol* (2016) 172:1293–305. doi: 10.1104/pp.16.00832
40. Tang M, Ning Y, Shu X, Dong B, Zhang H, Wu D, et al. The Nup98 Homolog APIP12 Targeted by the Effector AvrPiz-T Is Involved in Rice Basal Resistance Against Magnaporthe Oryzae. *Rice (NY)* (2017) 10:5. doi: 10.1186/s12284-017-0144-7
41. Panda D, Pascual-Garcia P, Dunagin M, Tudor M, Hopkins KC, Xu J, et al. Nup98 Promotes Antiviral Gene Expression to Restrict RNA Viral Infection in Drosophila. *Proc Natl Acad Sci USA* (2014) 111:E3890–9. doi: 10.1073/pnas.1410087111
42. Fontoura BM, Blobel G, Matunis MJ. A Conserved Biogenesis Pathway for Nucleoporins: Proteolytic Processing of a 186-Kilodalton Precursor

Generates Nup98 and the Novel Nucleoporin, Nup96. *J Cell Biol* (1999) 144:1097–112. doi: 10.1083/jcb.144.6.1097
43. Rosenblum JS, Blobel G. Autoproteolysis in Nucleoporin Biogenesis. *Proc Natl Acad Sci USA* (1999) 96:11370–5. doi: 10.1073/pnas.96.20.11370
44. Jiang S, Xiao L, Huang P, Cheng Z, Chen F, Miao Y, et al. Nucleoporin Nup98 Participates in Flowering Regulation in a CONSTANS-Independent Mode. *Plant Cell Rep* (2019) 38:1263–71. doi: 10.1007/s00299-019-02442-w
45. Xiao L, Jiang S, Huang P, Chen F, Wang X, Cheng Z, et al. Two Nucleoporin98 Homologous Genes Jointly Participate in the Regulation of Starch Degradation to Repress Senescence in Arabidopsis. *BMC Plant Biol* (2020) 20:292. doi: 10.1186/s12870-020-02494-1
46. Toufighi K, Brady SM, Austin R, Ly E, Provart NJ. The Botany Array Resource: E-Northerns, Expression Angling, and Promoter Analyses. *Plant J* (2005) 43:153–63. doi: 10.1111/j.1365-313X.2005.02437.x
47. de Leone MJ, Hernando CE, Romanowski A, Careno DA, Soverna AF, Sun H, et al. Bacterial Infection Disrupts Clock Gene Expression to Attenuate Immune Responses. *Curr Biol* (2020) 30:1740–47.e1746. doi: 10.1016/j.cub.2020.02.058
48. Wang W, Barnaby JY, Tada Y, Li H, Tor M, Caldelari D, et al. Timing of Plant Immune Responses by a Central Circadian Regulator. *Nature* (2011) 470:110–4. doi: 10.1038/nature09766
49. Zhou M, Wang W, Karapetyan S, Mwimba M, Marques J, Buchler NE, et al. Redox Rhythm Reinforces the Circadian Clock to Gate Immune Response. *Nature* (2015) 523:472–6. doi: 10.1038/nature14449
50. Jang AR, Moravcevic K, Saez L, Young MW, Sehgal A. Drosophila TIM Binds Importin Alpha1, and Acts as an Adapter to Transport PER to the Nucleus. *PLoS Genet* (2015) 11:e1004974. doi: 10.1371/journal.pgen.1004974
51. Szabo A, Papin C, Cornu D, Chelot E, Lipinszki Z, Udvardy A, et al. Ubiquitylation Dynamics of the Clock Cell Proteome and TIMELESS During a Circadian Cycle. *Cell Rep* (2018) 23:2273–82. doi: 10.1016/j.celrep.2018.04.064
52. Wong RW, Mamede JI, Hope TJ. Impact of Nucleoporin-Mediated Chromatin Localization and Nuclear Architecture on HIV Integration Site Selection. *J Virol* (2015) 89:9702–5. doi: 10.1128/JVI.01669-15
53. Lelek M, Casartelli N, Pellin D, Rizzi E, Souque P, Severgnini M, et al. Chromatin Organization at the Nuclear Pore Favours HIV Replication. *Nat Commun* (2015) 6:6483. doi: 10.1038/ncomms7483
54. Matreyek KA, Yucel SS, Li X, Engelman A. Nucleoporin NUP153 Phenylalanine-Glycine Motifs Engage a Common Binding Pocket Within the HIV-1 Capsid Protein to Mediate Lentiviral Infectivity. *PLoS Pathog* (2013) 9:e1003693. doi: 10.1371/journal.ppat.1003693
55. Wang S, Gu Y, Zebell SG, Anderson LK, Wang W, Mohan R, et al. A Noncanonical Role for the CKI-RB-E2F Cell-Cycle Signaling Pathway in Plant Effector-Triggered Immunity. *Cell Host Microbe* (2014) 16:787–94. doi: 10.1016/j.chom.2014.10.005
56. Gu Y, Zebell SG, Liang Z, Wang S, Kang BH, Dong X. Nuclear Pore Permeabilization Is a Convergent Signaling Event in Effector-Triggered Immunity. *Cell* (2016) 166:1526–1538 e1511. doi: 10.1016/j.cell.2016.07.042
57. Gu Y. The Nuclear Pore Complex: A Strategic Platform for Regulating Cell Signaling. *New Phytol* (2018) 219:25–30. doi: 10.1111/nph.14756
58. Tang Y, Huang A, Gu Y. Global Profiling of Plant Nuclear Membrane Proteome in Arabidopsis. *Nat Plants* (2020) 6(7):838–47. doi: 10.1038/s41477-020-0700-9
59. Fujitomo T, Daigo Y, Matsuda K, Ueda K, Nakamura Y. Critical Function for Nuclear Envelope Protein TMEM209 in Human Pulmonary Carcinogenesis. *Cancer Res* (2012) 72:4110–8. doi: 10.1158/0008-5472.CAN-12-0159
60. Polager S, Ginsberg D. P53 and E2f: Partners in Life and Death. *Nat Rev Cancer* (2009) 9:738–48. doi: 10.1038/nrc2718
61. Zebell SG, Dong X. Cell-Cycle Regulators and Cell Death in Immunity. *Cell Host Microbe* (2015) 18:402–7. doi: 10.1016/j.chom.2015.10.001
62. Tamura K, Fukao Y, Hatsugai N, Katagiri F, Hara-Nishimura I. Nup82 Functions Redundantly With Nup136 in a Salicylic Acid-Dependent Defense Response of Arabidopsis Thaliana. *Nucleus* (2017) 8:301–11. doi: 10.1080/19491034.2017.1279774
63. Mermet S, Voisin M, Mordier J, Dubos T, Tutois S, Tuffery P, et al. Evolutionary Conserved Protein Motifs Drive Attachment of the Plant Nucleoskeleton at Nuclear Pores. *bioRxiv* (2021). doi: 10.1101/2021.03.20.435662
64. Choi J, Strickler SR, Richards EJ. Loss of CRWN Nuclear Proteins Induces Cell Death and Salicylic Acid Defense Signaling. *Plant Physiol* (2019) 179:1315–29. doi: 10.1104/pp.18.01020
65. Pumroy RA, Cingolani G. Diversification of Importin-Alpha Isoforms in Cellular Trafficking and Disease States. *Biochem J* (2015) 466:13–28. doi: 10.1042/BJ20141186
66. Percipalle P, Clarkson WD, Kent HM, Rhodes D, Stewart M. Molecular Interactions Between the Importin Alpha/Beta Heterodimer and Proteins Involved in Vertebrate Nuclear Protein Import. *J Mol Biol* (1997) 266:722–32. doi: 10.1006/jmbi.1996.0801
67. Bayliss R, Littlewood T, Stewart M. Structural Basis for the Interaction Between FxFG Nucleoporin Repeats and Importin-Beta in Nuclear Trafficking. *Cell* (2000) 102:99–108. doi: 10.1016/S0092-8674(00)00014-3
68. Lee BJ, Cansizoglu AE, Suel KE, Louis TH, Zhang Z, Chook YM. Rules for Nuclear Localization Sequence Recognition by Karyopherin Beta 2. *Cell* (2006) 126:543–58. doi: 10.1016/j.cell.2006.05.049
69. Chook YM, Suel KE. Nuclear Import by Karyopherin-Betas: Recognition and Inhibition. *Biochim Biophys Acta* (2011) 1813:1593–606. doi: 10.1016/j.bbamcr.2010.10.014
70. Guo L, Fare CM, Shorter J. Therapeutic Dissolution of Aberrant Phases by Nuclear-Import Receptors. *Trends Cell Biol* (2019) 29:308–22. doi: 10.1016/j.tcb.2018.12.004
71. Goto C, Tamura K, Fukao Y, Shimada T, Hara-Nishimura I. The Novel Nuclear Envelope Protein KAKU4 Modulates Nuclear Morphology in Arabidopsis. *Plant Cell* (2014) 26:2143–55. doi: 10.1105/tpc.113.122168
72. Mackmull MT, Klaus B, Heinze I, Chokkalingam M, Beyer A, Russell RB, et al. Landscape of Nuclear Transport Receptor Cargo Specificity. *Mol Syst Biol* (2017) 13:962. doi: 10.15252/msb.20177608
73. Xu F, Jia M, Li X, Tang Y, Jiang K, Bao J, et al. Exportin-4 Coordinates Nuclear Shuttling of TOPLESS Family Transcription Corepressors to Regulate Plant Immunity. *Plant Cell* (2021) 33:697–713. doi: 10.1093/plcell/koaa047
74. Jia M, Shen X, Tang Y, Shi X, Gu Y. A Karyopherin Constrains Nuclear Activity of the NLR Protein SNC1 and Is Essential to Prevent Autoimmunity in Arabidopsis. *Mol Plant* (2021) 14:1733–44. doi: 10.1016/j.molp.2021.06.011
75. Chen M, Nowak DG, Narula N, Robinson B, Watrud K, Ambrico A, et al. The Nuclear Transport Receptor Importin-11 Is a Tumor Suppressor That Maintains PTEN Protein. *J Cell Biol* (2017) 216:641–56. doi: 10.1083/jcb.201604025
76. Vanhaesebroeck B, Stephens L, Hawkins P. PI3K Signalling: The Path to Discovery and Understanding. *Nat Rev Mol Cell Biol* (2012) 13:195–203. doi: 10.1038/nrm3290
77. Kataoka N, Bachorik JL, Dreyfuss G. Transportin-SR, A Nuclear Import Receptor for SR Proteins. *J Cell Biol* (1999) 145:1145–52. doi: 10.1083/jcb.145.6.1145
78. Xu S, Zhang Z, Jing B, Gannon P, Ding J, Xu F, et al. Transportin-SR Is Required for Proper Splicing of Resistance Genes and Plant Immunity. *PLoS Genet* (2011) 7:e1002159. doi: 10.1371/journal.pgen.1002159
79. Christ F, Thys W, De Rijck J, Gijsbers R, Albanese A, Arosio D, et al. Transportin-SR2 Imports HIV Into the Nucleus. *Curr Biol* (2008) 18:1192–202. doi: 10.1016/j.cub.2008.07.079
80. Demeulemeester J, Blokken J, De Houwer S, Dirix L, Klaassen H, Marchand A, et al. Inhibitors of the Integrase-Transportin-SR2 Interaction Block HIV Nuclear Import. *Retrovirology* (2018) 15:5. doi: 10.1186/s12977-018-0389-2
81. Lee SH, Hannink M. The N-Terminal Nuclear Export Sequence of IkappaBalpha Is Required for RanGTP-Dependent Binding to CRM1. *J Biol Chem* (2001) 276:23599–606. doi: 10.1074/jbc.M011197200
82. Garcia AV, Blanvillain-Baufume S, Huibers RP, Wiermer M, Li G, Gobbato E, et al. Balanced Nuclear and Cytoplasmic Activities of EDS1 Are Required for a Complete Plant Innate Immune Response. *PLoS Pathog* (2010) 6:e1000970. doi: 10.1371/journal.ppat.1000970
83. Zhang M, Gong P, Ge L, Chang Z, Cheng X, Zhou X, et al. Nuclear Exportin 1 Facilitates Turnip Mosaic Virus Infection by Exporting the Sumoylated Viral Replicase and by Repressing Plant Immunity. *New Phytol* (2021) 232(3):1382–98. doi: 10.1111/nph.17657
84. Ludke D, Roth C, Kamrad SA, Messerschmidt J, Hartken D, Appel J, et al. Functional Requirement of the Arabidopsis Importin-Alpha Nuclear

Transport Receptor Family in Autoimmunity Mediated by the NLR Protein SNC1. *Plant J* (2021) 105:994–1009. doi: 10.1111/tpj.15082
85. Roth C, Ludke D, Klenke M, Quathamer A, Valerius O, Braus GH, et al. The Truncated NLR Protein TIR-NBS13 Is a MOS6/IMPORTIN-Alpha3 Interaction Partner Required for Plant Immunity. *Plant J* (2017) 92:808–21. doi: 10.1111/tpj.13717
86. Fagerlund R, Kinnunen L, Kohler M, Julkunen I, Melen K. NF-{Kappa}B Is Transported Into the Nucleus by Importin {Alpha}3 and Importin {Alpha}4. *J Biol Chem* (2005) 280:15942–51. doi: 10.1074/jbc.M500814200
87. Agrawal T, Gupta GK, Agrawal DK. Vitamin D Supplementation Reduces Airway Hyperresponsiveness and Allergic Airway Inflammation in a Murine Model. *Clin Exp Allergy* (2013) 43:672–83. doi: 10.1111/cea.12102
88. Meng W, Gao SJ. Targeting XPO1 Enhances Innate Immune Response and Inhibits KSHV Lytic Replication During Primary Infection by Nuclear Stabilization of the P62 Autophagy Adaptor Protein. *Cell Death Dis* (2021) 12:29. doi: 10.1038/s41419-020-03303-1
89. Gruenbaum Y, Margalit A, Goldman RD, Shumaker DK, Wilson KL. The Nuclear Lamina Comes of Age. *Nat Rev Mol Cell Biol* (2005) 6:21–31. doi: 10.1038/nrm1550
90. Burke B, Stewart CL. The Nuclear Lamins: Flexibility in Function. *Nat Rev Mol Cell Biol* (2013) 14:13–24. doi: 10.1038/nrm3488
91. Pawar S, Kutay U. The Diverse Cellular Functions of Inner Nuclear Membrane Proteins. *Cold Spring Harb Perspect Biol* (2021) 13(9):a040477. doi: 10.1101/cshperspect.a040477
92. Ciska M, Moreno Diaz de la Espina S. The Intriguing Plant Nuclear Lamina. *Front Plant Sci* (2014) 5:166. doi: 10.3389/fpls.2014.00166
93. Ciska M, Hikida R, Masuda K, Moreno Diaz de la Espina S. Evolutionary History and Structure of Nuclear Matrix Constituent Proteins, the Plant Analogues of Lamins. *J Exp Bot* (2019) 70:2651–64. doi: 10.1093/jxb/erz102
94. Wang H, Dittmer TA, Richards EJ. Arabidopsis CROWDED NUCLEI (CRWN) Proteins Are Required for Nuclear Size Control and Heterochromatin Organization. *BMC Plant Biol* (2013) 13:200. doi: 10.1186/1471-2229-13-200
95. Blunt EL, Shandler JA, Hughes EJ, Sussman H, Christopherson RC, Richards EJ. Coordination of NMCP1- and NMCP2-Class Proteins Within the Plant Nucleoskeleton. *Mol Biol Cell* (2020) 31:2948–58. doi: 10.1091/mbc.E19-08-0464
96. Jarad M, Mariappan K, Almeida-Trapp M, Mette MF, Mithofer A, Rayapuram N, et al. The Lamin-Like LITTLE NUCLEI 1 (LINC1) Regulates Pattern-Triggered Immunity and Jasmonic Acid Signaling. *Front Plant Sci* (2019) 10:1639. doi: 10.3389/fpls.2019.01639
97. Wang N, Karaaslan ES, Faiss N, Berendzen KW, Liu C. Characterization of a Plant Nuclear Matrix Constituent Protein in Liverwort. *Front Plant Sci* (2021) 12:670306. doi: 10.3389/fpls.2021.670306
98. Choi J, Richards EJ. The Role of CRWN Nuclear Proteins in Chromatin-Based Regulation of Stress Response Genes. *Plant Signal Behav* (2020) 15:1694224. doi: 10.1080/15592324.2019.1694224
99. Mikulski P, Hohenstatt ML, Farrona S, Smaczniak C, Stahl Y, Kalyanikrishna, et al. The Chromatin-Associated Protein PWO1 Interacts With Plant Nuclear Lamin-Like Components to Regulate Nuclear Size. *Plant Cell* (2019) 31:1141–54. doi: 10.1105/tpc.18.00663
100. Guo T, Mao X, Zhang H, Zhang Y, Fu M, Sun Z, et al. Lamin-Like Proteins Negatively Regulate Plant Immunity Through NAC WITH TRANSMEMBRANE MOTIF1-LIKE9 and NONEXPRESSOR OF PR GENES1 in Arabidopsis Thaliana. *Mol Plant* (2017) 10:1334–48. doi: 10.1016/j.molp.2017.09.008
101. Saez A, Herrero-Fernandez B, Gomez-Bris R, Somovilla-Crespo B, Rius C, Gonzalez-Granado JM. Lamin a/C and the Immune System: One Intermediate Filament, Many Faces. *Int J Mol Sci* (2020) 21(17):6109. doi: 10.3390/ijms21176109
102. Yabuki M, Miyake T, Doi Y, Fujiwara T, Hamazaki K, Yoshioka T, et al. Role of Nuclear Lamins in Nuclear Segmentation of Human Neutrophils. *Physiol Chem Phys Med NMR* (1999) 31:77–84.
103. Olins AL, Zwerger M, Herrmann H, Zentgraf H, Simon AJ, Monestier M, et al. The Human Granulocyte Nucleus: Unusual Nuclear Envelope and Heterochromatin Composition. *Eur J Cell Biol* (2008) 87:279–90. doi: 10.1016/j.ejcb.2008.02.007
104. Hoffmann K, Sperling K, Olins AL, Olins DE. The Granulocyte Nucleus and Lamin B Receptor: Avoiding the Ovoid. *Chromosoma* (2007) 116:227–35. doi: 10.1007/s00412-007-0094-8
105. Kim Y, Bayona PW, Kim M, Chang J, Hong S, Park Y, et al. Macrophage Lamin a/C Regulates Inflammation and the Development of Obesity-Induced Insulin Resistance. *Front Immunol* (2018) 9:696. doi: 10.3389/fimmu.2018.00696
106. Tang Y, Dong Q, Wang T, Gong L, Gu Y. PNET2 Is a Component of the Plant Nuclear Lamina and Is Required for Proper Genome Organization and Activity. *Dev Cell* (2022) S1534-5807(21)00887-X. doi: 10.1016/j.devcel.2021.11.002

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Immunology

ORIGINAL RESEARCH
published: 22 December 2021
doi: 10.3389/fimmu.2021.795331

Activin and BMP Signaling Activity Affects Different Aspects of Host Anti-Nematode Immunity in *Drosophila melanogaster*

Yaprak Ozakman, Dhaivat Raval and Ioannis Eleftherianos *

Infection and Innate Immunity Laboratory, Department of Biological Sciences, The George Washington University, Washington, DC, United States

The multifaceted functions ranging from cellular and developmental mechanisms to inflammation and immunity have rendered TGF-ß signaling pathways as critical regulators of conserved biological processes. Recent studies have indicated that this evolutionary conserved signaling pathway among metazoans contributes to the *Drosophila melanogaster* anti-nematode immune response. However, functional characterization of the interaction between TGF-ß signaling activity and the mechanisms activated by the *D. melanogaster* immune response against parasitic nematode infection remains unexplored. Also, it is essential to evaluate the precise effect of entomopathogenic nematode parasites on the host immune system by separating them from their mutualistic bacteria. Here, we investigated the participation of the TGF-ß signaling branches, activin and bone morphogenetic protein (BMP), to host immune function against axenic or symbiotic *Heterorhabditis bacteriophora* nematodes (parasites lacking or containing their mutualistic bacteria, respectively). Using *D. melanogaster* larvae carrying mutations in the genes coding for the TGF-ß extracellular ligands Daw and Dpp, we analyzed the changes in survival ability, cellular immune response, and phenoloxidase (PO) activity during nematode infection. We show that infection with axenic *H. bacteriophora* decreases the mortality rate of *dpp* mutants, but not *daw* mutants. Following axenic or symbiotic *H. bacteriophora* infection, both *daw* and *dpp* mutants contain only plasmatocytes. We further detect higher levels of *Dual oxidase* gene expression in *dpp* mutants upon infection with axenic nematodes and *Diptericin* and *Cecropin* gene expression in *daw* mutants upon infection with symbiotic nematodes compared to controls. Finally, following symbiotic *H. bacteriophora* infection, *daw* mutants have higher PO activity relative to controls. Together, our findings reveal that while *D. melanogaster* Dpp/BMP signaling activity modulates the DUOX/ROS response to axenic *H. bacteriophora* infection, Daw/activin signaling activity modulates the antimicrobial peptide and melanization responses to axenic *H. bacteriophora* infection. Results from this study expand our current understanding of the molecular and mechanistic interplay between nematode parasites and the host immune system, and

OPEN ACCESS

Edited by:
Katherine Buckley, Auburn University, United States

Reviewed by:
Samuel Liegeois, University of Strasbourg, France
Lage Cerenius, Uppsala University, Sweden

***Correspondence:**
Ioannis Eleftherianos
ioannise@gwu.edu

Specialty section:
This article was submitted to Comparative Immunology, a section of the journal Frontiers in Immunology

Received: *14 October 2021*
Accepted: *06 December 2021*
Published: *22 December 2021*

Citation:
Ozakman Y, Raval D and Eleftherianos I (2021) Activin and BMP Signaling Activity Affects Different Aspects of Host Anti-Nematode Immunity in Drosophila melanogaster. Front. Immunol. 12:795331. doi: 10.3389/fimmu.2021.795331

the involvement of TGF-ß signaling branches in this process. Such findings will provide valuable insight on the evolution of the immune role of TGF-ß signaling, which could lead to the development of novel strategies for the effective management of human parasitic nematodes.

Keywords: *Drosophila*, innate immunity, *Heterorhabditis*, *Photorhabdus*, TGF-ß

INTRODUCTION

Transforming growth factor (TGF-ß) signaling is an evolutionary conserved signaling pathway among metazoans (1). Having components present in all animals studied to date, TGF-ß signaling has been regarded as a key pathway in metazoan evolution and in transition to multicellularity (1, 2). In addition, its role has been implicated in various biological processes ranging from development to metabolism (3). Compared to vertebrates, TGF-ß signaling in *Drosophila melanogaster* comprises fewer representatives of each signaling component, yet still regulates diverse functions including axis formation, body patterning, and morphogenesis (4–7). Activin and bone morphogenetic protein (BMP) are the two signaling branches of the *D. melanogaster* TGF-ß pathway, which are characterized by extracellular ligands, type I and type II receptors, and intracellular transducers (8). Signaling is initiated by the binding of ligands to transmembrane receptor complex of serine/threonine kinases, which in turn phosphorylates transcription factors that regulate the activation of downstream genes (9).

In addition to its contribution to wounding response and anti-bacterial immunity, the role of TGF-ß signaling has been implicated in anti-nematode immunity in *D. melanogaster* (10–14). In particular, gene transcript levels of the ligands Daw (activin branch) and Dpp (BMP branch) are upregulated following infection with the parasitic nematodes, *Heterorhabditis gerrardi* and *Heterorhabditis bacteriophora* in adult flies (12). Inactivation of *dpp* leads to increased survival ability and induction of humoral immunity against *H. bacteriophora* infection (11). Also, in response to *H. gerrardi* infection, *daw* mutants have decreased expression of *Dual oxidase* (*Duox*) compared to their background controls (14). However, detailed functional characterization of the interaction between TGF-ß signaling activity and cellular, humoral, and melanization responses in *D. melanogaster* against parasitic nematode infection remains incomplete.

The entomopathogenic nematode (EPN) *H. bacteriophora* and its mutualistic bacterium *P. luminescens* constitute an excellent model to elucidate the interaction between TGF-ß signaling and host anti-nematode immunity. The nematodes and their bacteria, together or separately, can infect and kill a variety of insect species (15). Infective juveniles (IJs), the only free-living stage of the nematodes, enter the insect hemocoel either through natural openings or by using their special appendage to penetrate through the cuticle (16). Upon entry, IJs regurgitate their *Photorhabdus* bacteria into the hemolymph to spread into the host and overcome the insect immune response (17, 18). Once the insect dies and resources in the cadaver are depleted, IJs emerge from the carcass to search for new prey (15) *D. melanogaster* is not a natural host of *H. bacteriophora* (19). However, previous studies have established methods for infecting *D. melanogaster* larvae and adult flies with *Heterorhabditis* nematodes, which can develop efficiently in both stages of the host (19, 20).

Although previous studies have shown that TGF-ß signaling interacts with the *D. melanogaster* larval and adult immune response during infection with the *Heterorhabditis-Photorhabdus* complex, it is crucial to elucidate the precise effect of the parasites on the host immune system by separating them from their mutualistic bacteria (11–14). To this end, we investigated the participation of the two TGF-ß signaling branches, activin and BMP, in host immune function against axenic (lacking *P. luminescens*) or symbiotic (containing *P. luminescens*) *H. bacteriophora* nematodes. Using *D. melanogaster* larvae carrying mutations in the genes coding for the TGF-ß extracellular ligands Daw and Dpp, we explored modifications in cellular, humoral, and melanization responses during infection with these pathogens.

In this study, we show that BMP signaling activity promotes survival to *D. melanogaster* larvae when infected by symbiotic *H. bacteriophora* but increases mortality to larvae infected by axenic nematodes. Also, following axenic *H. bacteriophora* infection, BMP signaling activity leads to reduced *DUOX* expression and increased *Diptericin* and *Drosomycin* antimicrobial gene expression. Similarly, activin signaling activity upregulates *Drosomycin* expression against axenic nematode infection but downregulates *Diptericin* and *Cecropin* expression against symbiotic nematode infection. Finally, we demonstrate that in response to symbiotic *H. bacteriophora* infection, activin signaling activity lowers PO activity in *D. melanogaster* larvae. Taken together our findings provide a detailed characterization of the *D. melanogaster* activin and BMP signaling activity regulation in relation to host immune function against parasitic nematodes deficient of their mutualistic bacteria. The information garnered from these results set the stage for elucidating the evolution of the immune role of TGF-ß signaling against parasitic nematode infection and contribute to the effective use of *Heterorhabditis* nematodes as models for human parasitic diseases.

MATERIALS AND METHODS

Fly and Nematode Stocks

All flies were reared on Bloomington Drosophila Stock Center cornmeal food (Labexpress) supplemented with yeast (Carolina

Biological Supply), maintained at 25°C, and a 12:12-h light:dark photoperiodic cycle. Larvae carrying a spontaneous dpp^{s1} mutation (strain 397, Bloomington, IL, USA) and carrying P-bac insertion Pbac{XP}daw05680 (strain d05680, Exelixis, Boston, MA, USA) were used, as previously described (14). Line w^{1118} (strain 3605, Bloomington, IL, USA) was used as the background control. Infective juveniles (IJs) of *H. bacteriophora* strain TT01 were amplified in larvae of the wax moth *Galleria mellonella* using the water trap technique (21). To generate axenic *H. bacteriophora* nematodes lacking their *P. luminescens* bacteria, a previously established protocol was followed (22). Prior to experiments, axenic nematodes were surface sterilized in 5% bleach solution and rinsed five times with water to remove the bleach residue. Nematodes were used 1-4 weeks after collection.

Larval Infection and Survival Assay

Infections were carried out using 96-well plates containing 100 µL of 1.25% agarose in each well. The suspension of axenic or symbiotic *H. bacteriophora* IJs (100 nematodes in 10 µL of sterile distilled water) was added to a single larva (late second to third instar) in each well. Sterile distilled water (10 µL) served as uninfected control. The plate was covered with a sealing film (USA Scientific, Ocala, FL, USA) and air holes were poked for ventilation. Plates were kept in the dark at room temperature for 24 h. At the 24-h time point, larvae were collected and frozen at −80°C, or immediately used in experiments. The survival of larvae kept in nematode solution or in sterile distilled water was estimated at 12-h intervals and up to 60 h. The survival experiments were replicated three times.

Gene Expression Analysis

RNA was extracted from 4-5 *D. melanogaster* larvae (late second to third instar) using Trizol reagent (Ambion, Life Technologies) and reverse transcription was performed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturers' instructions. Real time PCR experiments were conducted in a CFX96 Real-Time System, C1000 Thermal Cycler (Bio-Rad) with gene-specific primers (**Table 1**) using GreenLink qPCR Mix (BioLink). Each experiment was run in technical triplicates and repeated three times.

Hemocyte Assays

Following axenic or symbiotic *H. bacteriophora* infection, 10 *D. melanogaster* larvae (late second to third instar) were bled into 30 µl of 2.5× protease inhibitor cocktail (Sigma). Hemolymph samples were loaded on a hemocytometer and total numbers and aggregates of cells were counted using 40× magnification on a compound microscope (Olympus CX21). Clumps of two or more hemocytes were considered as aggregates and each clump was counted as one aggregate. In each experiment five technical replicates were used, and each experiment was repeated three times.

Hemocyte Staining and Microscopy

To distinguish the different types of hemocytes following infection with axenic or symbiotic *H. bacteriophora*, six *D. melanogaster* larvae (late second to third instar) were washed in 1xPBS and bled into 30 µl of 1x PBS on a poly-lysine coated microscope slide which was air dried for 20 min. Hemocytes were fixed in 4% paraformaldehyde for 15 min, washed with 1xPBS, permeabilized in PBST (0.3% Triton in 1x PBS) for 20 min, and washed again in 1x PBS. Hemocytes were blocked in PBSTB (8% BSA in PBST) for 15 min, washed in 1x PBS, and then incubated with diluted phalloidin TRITC (1:100) for 20 min before a final wash in 1x PBS. Using ProLong® Gold Antifade reagent with DAPI, hemocytes were mounted and stored at 4°C until imaging. Hemocytes were visualized using Zeiss LSM 510 confocal microscope and their types were distinguished based on morphology, as previously described (23). The experiment was repeated three times.

TABLE 1 | Primers and their sequences used in quantitative qPCR experiments.

Gene	Accession Number	Primer	Sequence (5' - 3')	Tm (°C)
NOS	CG6713	Forward	AACGTTCGACAAATGCGCAA	60
		Reverse	GTTGCTGTGTCTGTGCCTTC	
DUOX	CG3131	Forward	ACGTGTCCACCCAATCGCACGAG	60
		Reverse	AAGCGTGGTGGTCCAGTCAGTCG	
PPO1	CG5779	Forward	CAACTGGCTTCGTTGAGTGA	60
		Reverse	CGGGCAGTTCCAATACAGTT	
PPO2	CG8193	Forward	CCCGCCTATACCGAGA	59
		Reverse	CGCACGTAGCCGAAAC	
PPO3	CG42640	Forward	GGCGAGCTGTTCTACT	58
		Reverse	GAGGATACGCCCTACTG	
Diptericin	CG12763	Forward	GCTGCGCAATCGCTTCTACT	60
		Reverse	TGGTGGAGTTGGGCTTCATG	
Cecropin	CG1365	Forward	TCTTCGTTTTCGTCGCTCTC	57
		Reverse	CTTGTTGAGCGATTCCCAGT	
Drosomycin	CG10810	Forward	GACTTGTTCGCCCTCTTCG	60
		Reverse	CTTGCACACACGACGACAG	
Metchnikowin	CG8175	Forward	TCTTGGAGCGATTTTTCTGG	56
		Reverse	AATAAATTGGACCCGGTCTTG	
RpL32	CG7939	Forward	GATGACCATCCGCCCAGCA	60
		Reverse	CGGACCGACAGCTGCTTGGC	

Nitric Oxide and Aconitase Assays

To assess levels of Nitric Oxide (NO) and aconitase activity, six late second to third instar *D. melanogaster* larvae were collected 24-h post-infection with axenic or symbiotic *H. bacteriophora*. For NO estimation, larvae were homogenized in PBS and centrifuged at 10,000 x g for 10 min at 4°C. The resultant supernatant was used to quantify proteins as previously described and then mixed 1:1 with Griess reagent (Sigma) (1). Absorbance was measured at 595 nm using a plate reader (BioTek). To calculate NO levels, a silver nitrite standard curve was used. For the estimation of aconitase activity, larvae were homogenized in aconitase assay buffer and processed following the manufacturer's instructions (MAK051-1KT; Sigma). Absorbance was measured at 450 nm and levels of aconitase activity were calculated from an isocitrate standard curve. Experiments were conducted in technical duplicates and repeated three times.

Phenoloxidase Assay

To collect hemolymph following infection with axenic or symbiotic *H. bacteriophora*, 10 larvae from each *D. melanogaster* line were bled into 30 µl of 2.5x protease inhibitor. Hemolymph samples were added to Pierce® Spin Columns and spun at 4°C and 13,000 rpm for 10 min. Protein concentrations were measured using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). A mixture containing 15 µg of protein, 5 mM $CaCl_2$, and 2.5× protease inhibitor was added to 160 µl of fresh L-DOPA solution (in phosphate buffer, pH 6.6) in a clear microplate well. Absorbance values were measured at 29°C and 492 nm for 60 min. Absorbance of the blank was subtracted from the absorbance of each sample. Each experiment was repeated three times.

Assessment of Melanization

To observe the ability of hemolymph to melanize following infection with axenic or symbiotic *H. bacteriophora*, hemolymph samples were collected from 10 larvae of each *D. melanogaster* line in 1xPBS, added to a Pierce® Spin Column and spun at 4°C and 13,000 rpm for 10 min. Hemolymph plasma samples were transferred to a 96 well microtiter plate. After 3 h incubation at room temperature, the presence of melanization in the wells was noted.

Statistical Analysis

GraphPad Prism 8 was used for data plotting and statistical analyses. Statistics for the results from the survival experiments were carried out using log-rank (Mantel-Cox) test. One-way analysis of variance (ANOVA) and Tukey *post-hoc* tests were used for analyzing the results from the rest of the experiments.

RESULTS

BMP Branch Activity Promotes Survival of *D. melanogaster* Larvae Upon Symbiotic *H. bacteriophora* Infection

To understand whether activin and BMP TGF-ß signaling branches contribute to the survival ability of *D. melanogaster* larvae against infection with *H. bacteriophora* containing or lacking their mutualistic bacteria, we challenged *daw* and *dpp* mutants with either type of nematode to assess their time-course survival rate following infection. For this, we monitored larval survival every 12 h and up to 48 h post nematode infection. We did not observe any significant differences between *daw* mutants and their background controls (w^{1118}) following axenic or symbiotic *H. bacteriophora* infection (**Figure 1A**). However, we found that while *dpp* mutants showed higher survival rates upon axenic *H. bacteriophora* infection, their survival ability to symbiotic nematode infection decreased compared to w^{1118} controls (**Figure 1B**). These results indicate that the BMP signaling activity in *D. melanogaster* larvae promotes survival against symbiotic *H. bacteriophora* infection but limits survival against axenic *H. bacteriophora* infection.

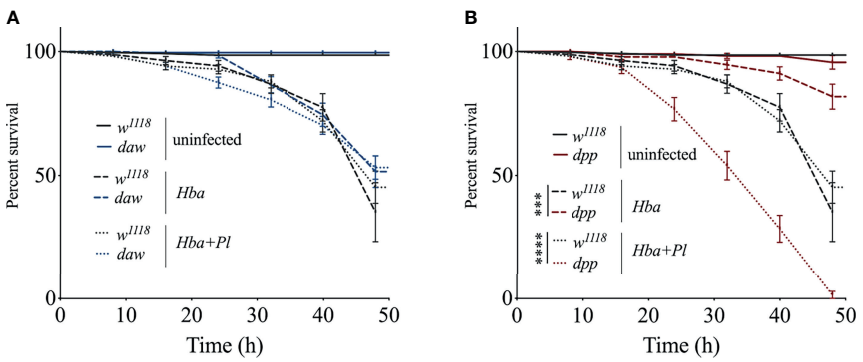


FIGURE 1 | Survival analysis of *Drosophila melanogaster* TGF-ß mutant larvae upon infection with axenic (*Hba*) or symbiotic (*Hba+Pl*) *Heterorhabditis bacteriophora*. *Daw* and *dpp* mutant larvae and their background controls (w^{1118}) were infected with either *H. bacteriophora* axenic or symbiotic nematodes. Treatment with water served as negative control (uninfected). Larval survival was estimated every 12 h and up to 48 h after infection. Significance levels were assessed using log-rank (Mantel-Cox) test. Survival ability of *daw* **(A)** and *dpp* **(B)** mutants following infection with axenic or symbiotic *H. bacteriophora* (***p = 0.0002, ****p < 0.0001).

Activin and BMP Signaling Activity Do Not Participate in the *D. melanogaster* Larval Cellular Immune Response to *H. bacteriophora* Infection

The cellular immune response is mediated by the hemocytes and forms an integral part of the *D. melanogaster* host defense because it acts in conjunction with humoral immune mechanisms to oppose infection (24, 25). Previous studies have linked the number of hemocytes to immune system capacity in *D. melanogaster* (26–29). To determine whether activin and BMP signaling activity modulate the number of hemocytes in *D. melanogaster* larvae responding to parasitic nematode infection, we counted the number of circulating hemocytes in *daw* and *dpp* mutants following infection with axenic or symbiotic *H. bacteriophora*. Of note, here we focused only on circulating hemocytes, but there is a possibility that some of the sessile hemocytes may have become circulating to act on the anti-nematode response. We found that the number of hemocytes was not significantly affected in TGF-ß mutants when infected with either type of nematode compared to their background controls. However, hemocyte numbers decreased in *dpp* mutants following infection with axenic *H. bacteriophora* relative to uninfected *dpp* individuals suggesting a link between hemocyte proliferation and BMP signaling activity (**Figure 2A**). Wasp parasitization in *D. melanogaster* larvae induces the differentiation of lamellocytes, which are specialized type of hemocytes (29–32). Previous evidence suggests that infection with the EPN *Steinernema carpocapsae* also leads to lamellocyte differentiation in *D. melanogaster* wild-type larvae (23). To elucidate whether activin and BMP signaling branches in *D. melanogaster* larvae are involved in the differentiation of lamellocytes during *H. bacteriophora* infection, we stained hemocytes in *daw* and *dpp* mutants and their w^{1118} background controls with phalloidin-TRITC and DAPI and visualized them using confocal microscopy (**Figure 2B**). We only detected plasmatocytes in either TGF-ß mutant following infection with axenic or symbiotic *H. bacteriophora* nematodes. Interestingly, we noticed the formation of a higher number of hemocyte aggregates in uninfected *daw* mutants compared to their background controls. To quantitively assess whether inactivating TGF-ß signaling branches leads to changes in hemocyte aggregation in the context of nematode infection, we counted the number of hemocyte aggregates in *daw* and *dpp* mutants following infection with axenic or symbiotic *H. bacteriophora*. Consistent with the initial observation, we found that in the absence of infection, there was a higher number of hemocyte aggregates in *daw* mutants compared to those detected in w^{1118} individuals. In addition, following axenic *H. bacteriophora* infection, the number of cell aggregates decreased in *daw* mutant larvae compared to their uninfected counterparts (**Figure 2C**). These findings suggest that activin and BMP signaling activity do not affect cellular immune processes in *D. melanogaster* larvae, such as total hemocyte numbers,

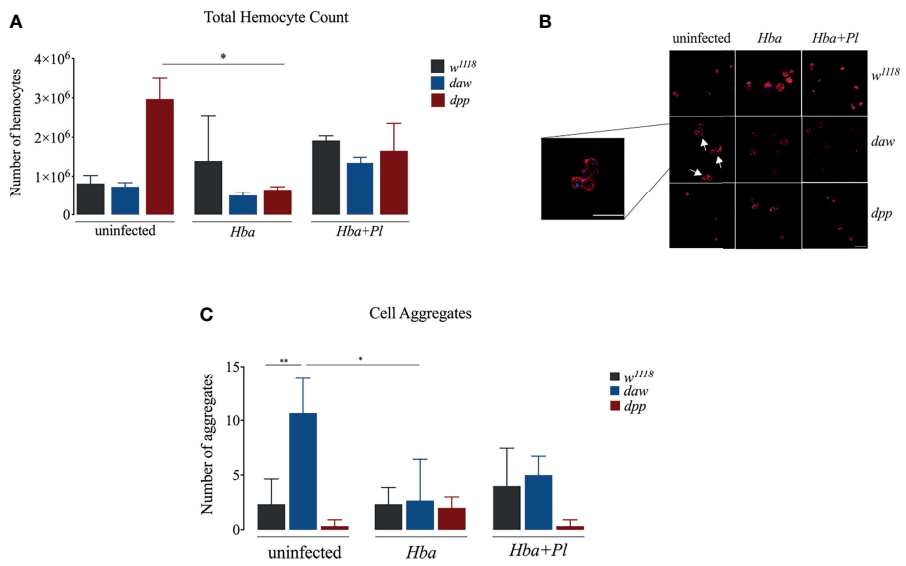


FIGURE 2 | Cellular immune response of *Drosophila melanogaster* TGF-ß mutant larvae against axenic (*Hba*) or symbiotic (*Hba+Pl*) *Heterorhabditis bacteriophora* nematode infection. **(A)** Total number of circulating hemocytes in *D. melanogaster daw* and *dpp* mutant larvae and their background controls (w^{1118}) upon infection with either axenic or symbiotic nematodes. Hemolymph samples were collected from 10 larvae at 24 h following infection and the number of circulating hemocytes was assessed using a hemocytometer under a light microscope (*p = 0.0343). **(B)** Confocal microscopy images of hemocytes stained with Phalloidin-TRITC (red) and DAPI (blue) in *D. melanogaster daw* and *dpp* mutants and their background controls (w^{1118}) following axenic or symbiotic *H. bacteriophora* nematode infection. Scale bar is 20 µm. White arrows indicate hemocyte aggregates. **(C)** Assessment of the number of hemocyte aggregates in *D. melanogaster daw* and *dpp* mutant larvae and their background controls (w^{1118}) following *H. bacteriophora* axenic or symbiotic nematode infection (*p = 0.0123, **p = 0.0086). Significance levels were assessed using one-way analysis of variance (ANOVA).

differentiation of lamellocytes, and formation of hemocyte aggregates upon *H. bacteriophora* infection.

BMP Signaling Activity in *D. melanogaster* Larvae Lowers Expression of *DUOX* Upon Axenic *H. bacteriophora* Infection

Similar to vertebrates, intermediates of both oxygen (ROI) and nitrogen (RNO) in *D. melanogaster* constitute a prominent defense strategy against bacterial infections (33). *D. melanogaster* has a single NADPH, Dual oxidase (DUOX), which acts as a principal factor in the production of reactive oxygen species (ROS) (34, 35). Recently, it was shown that *D. melanogaster daw* mutant larvae express *DUOX* at higher levels upon infection with *H. gerrardi* nematodes (14). To elucidate whether activin and BMP branches are involved in the *D. melanogaster* DUOX mediated ROS response to *H. bacteriophora* infection, we estimated the gene expression level of *DUOX* in *daw* and *dpp* mutant larvae following infection with either axenic or symbiotic nematodes. We found no significant difference in *DUOX* expression between *daw* mutants and w^{1118} controls when infected with either type of nematode (**Figure 3A**). Interestingly, *dpp* mutants expressed *DUOX* at higher levels compared to w^{1118} individuals during axenic but not symbiotic *H. bacteriophora* infection. This finding indicates that the BMP branch mediates ROS response through reducing the expression of *DUOX* upon infection with *P. luminescens*-deficient *H. bacteriophora* nematodes. We have also assessed ROS levels by measuring the relative aconitase activity, yet we have not observed any significant differences between *daw* and *dpp* mutants and w^{1118} controls upon infection with axenic or symbiotic *H. bacteriophora* (**Figure S1A**). We then investigated the association between nitric oxide (NO) response, which is controlled by the enzyme NO synthase (NOS), and TGF-ß signaling activity in the *D. melanogaster* anti-nematode response. For this, we performed qRT-PCR to determine the expression level of *NOS* and spectrophotometrically quantified nitrite protein levels in *daw* and *dpp* mutant larvae infected by either axenic or symbiotic *H. bacteriophora*. There were no statistically significant differences in *NOS* expression or nitrite protein levels between nematode infected TGF-ß mutants and the w^{1118} controls (**Figures 3B** and **S1B**). This outcome suggests that TGF-ß signaling activity in *D. melanogaster* larvae is not involved in the NO response to parasitic nematode infection.

Activin Signaling Activity Lowers Expression of *Diptericin* and *Cecropin* in *D. melanogaster* Larvae Following Infection With Symbiotic *H. bacteriophora*

The hallmark of the *D. melanogaster* host defense is the inducible synthesis and secretion of antimicrobial peptides (AMPs) through NF-κB signaling pathways (36). Also, using GFP reporter transgenes, infection of *D. melanogaster* larvae with symbiotic *H. bacteriophora* results in higher percentage of fly larvae expressing the AMP encoding genes *Attacin*, *Diptericin*, *Drosomycin*, and *Metchnikowin* (19). Previous studies have demonstrated a connection between TGF-ß signaling and synthesis of AMPs in the context of parasitic nematode infection (19, 37). In addition, NF-κB/Rel signaling interferes with the expression of Activinβ (activin signaling ligand), while NF-κB/Dif signaling interferes with the expression of glass bottom boat (BMP signaling ligand) in *D. melanogaster* larvae infected by *H. gerrardi* nematodes (37). To further study the link between TGF-ß signaling activity and expression of AMPs following infection with axenic or symbiotic *H. bacteriophora* nematodes, we used qRT-PCR and gene-specific primers to determine the expression levels of *Diptericin*, *Cecropin*, *Drosomycin*, and *Metchnikowin* in *D. melanogaster daw* and *dpp* mutants. We found that following symbiotic *H. bacteriophora* infection, expression of *Diptericin* and *Cecropin* were higher in *daw* mutant larvae compared to w^{1118} larvae (**Figures 4A, B**). *Diptericin* and *Cecropin* are mainly induced by Gram-negative bacterial infections (38). Therefore, our results indicate that these two AMPs were possibly induced by *P. luminescens* bacteria which are mutualistically associated with *H. bacteriophora* nematodes. Conversely, we found that *Drosomycin* expression significantly decreased in both *daw* and *dpp* mutants compared to w^{1118} controls upon infection with

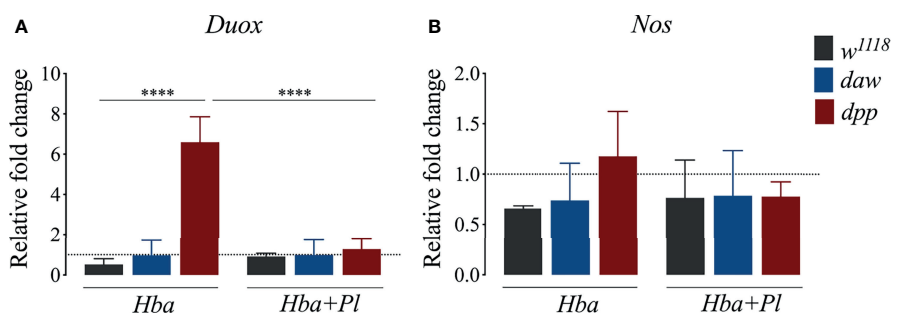


FIGURE 3 | Quantitative PCR analysis of *dual oxidase* (*DUOX*) **(A)** and *nitric oxide synthase* (*NOS*) **(B)** expression in *Drosophila melanogaster daw* and *dpp* mutant larvae and their background controls (w^{1118}) responding to infection by either axenic (*Hba*) or symbiotic (*Hba + Pl*) *Heterorhabditis bacteriophora* nematodes. Dotted line at 1.0 indicates normalization of fold change relative to uninfected controls. Significance levels were assessed using one-way analysis of variance (ANOVA, ****$p < 0.0001$).

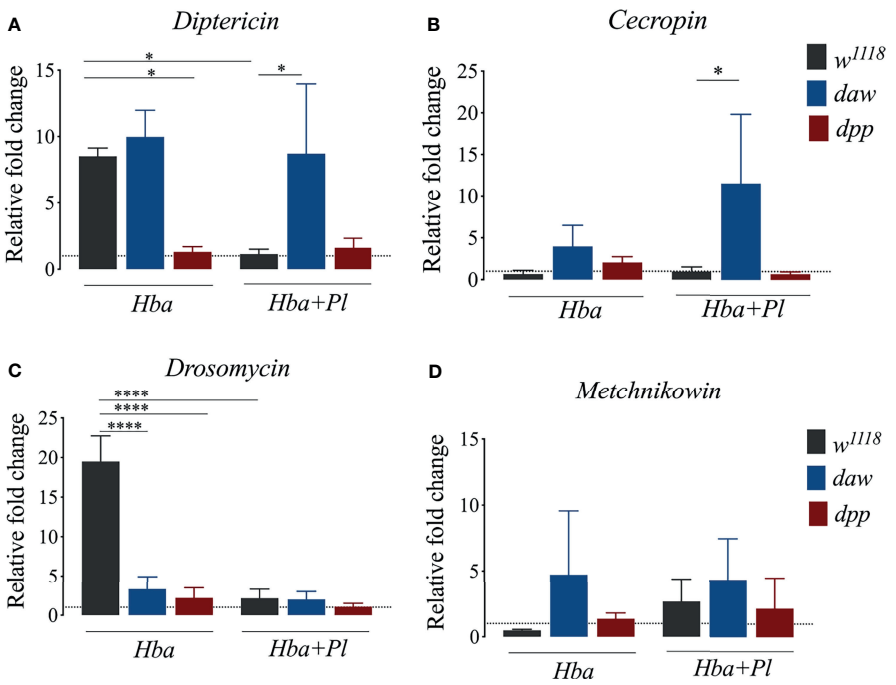


FIGURE 4 | Expression of the antimicrobial peptide encoding genes *Diptericin, Cecropin*, *Drosomycin*, and *Metchnikowin* in *Drosophila melanogaster daw* and *dpp* mutant larvae and their background controls (w^{1118}) following infection with either axenic (*Hba*) or symbiotic (*Hba+Pl*) *Heterorhabditis bacteriophora*. Dotted line at 1.0 indicates normalization of fold change relative to uninfected controls. Significance levels were assessed using one-way analysis of variance (ANOVA). **(A)** *Diptericin* (*p = 0.0180), **(B)** *Cecropin* (*p = 0.0318) **(C)** *Drosomycin* (****p < 0.0001), and **(D)** *Metchnikowin* expression *via* qPCR.

axenic *H. bacteriophora* (**Figure 4C**). We have not found any significant differences in *Metchnikowin* expression between TGF-ß mutants and their w^{1118} controls when infected with either axenic or symbiotic nematodes (**Figure 4D**). Taken together, these findings suggest that activin signaling activity interacts with the Imd pathway through decreasing the expression of *Diptericin* and *Cecropin* in *D. melanogaster* larvae during infection with *H. bacteriophora* nematodes harboring their mutualistic *P. luminescens* bacteria.

Activin Signaling Activity in *D. melanogaster* Larvae Reduces the Melanization Response Upon Symbiotic *H. bacteriophora* Infection

Melanization mediated by the enzyme phenoloxidase (PO) is an immediate and essential response to pathogen infection in *D. melanogaster* (39). When Daw/activin signaling is impaired in the absence of infection, melanotic tumors are observed in adult flies and PO enzyme activity increases in uninfected larvae (10, 14). Interestingly, *daw* mutant larvae have lower PO activity when challenged by *H. gerrardi* nematodes, suggesting an association between the activin branch and regulation of the melanization response against parasitic nematode infection (14). To examine the interaction between TGF-ß signaling activity in *D. melanogaster* and the melanization response to axenic or symbiotic *H. bacteriophora*, we estimated the expression levels of the inactive precursors of PO *prophenoloxidase* (*PPO*) genes *PPO1*, *PPO2*, and *PPO3* in *daw* and *dpp* mutant larvae infected with either type of nematode (39). We did not observe any significant differences in *PPO1* and *PPO2* expression between the TGF-ß mutants and their w^{1118} controls following axenic or symbiotic *H. bacteriophora* infection (**Figures S2A–C**). However, expression of *PPO3* increased significantly in *dpp* mutants compared to their background control following axenic nematode infection. The activity of PO enzyme in larval hemolymph can be measured spectrophotometrically through the detection of dopachrome which is converted by PO from L-DOPA (40). Using this method, we determined PO activity in the hemolymph plasma of *daw* and *dpp* mutants following infection with axenic or symbiotic *H. bacteriophora*. *Daw* mutants exhibited significantly increased PO activity compared to w^{1118} controls when infected with symbiotic *H. bacteriophora*, but not with axenic nematodes (**Figure 5A**). Extraction of hemolymph followed by incubation of the plasma at room temperature leads to blackening due to the production of melanin (41). We observed that hemolymph plasma from *daw* mutant larvae, which had previously been infected with symbiotic *H. bacteriophora*, melanized at room temperature (**Figure 5B**). However, we have not observed melanization in hemolymph plasma from *dpp* mutants or background controls following symbiotic *H. bacteriophora* infection. These findings collectively suggest that activin signaling activity regulates the melanization response in *D. melanogaster* larvae through reducing the levels of PO following symbiotic *H. bacteriophora* infection.

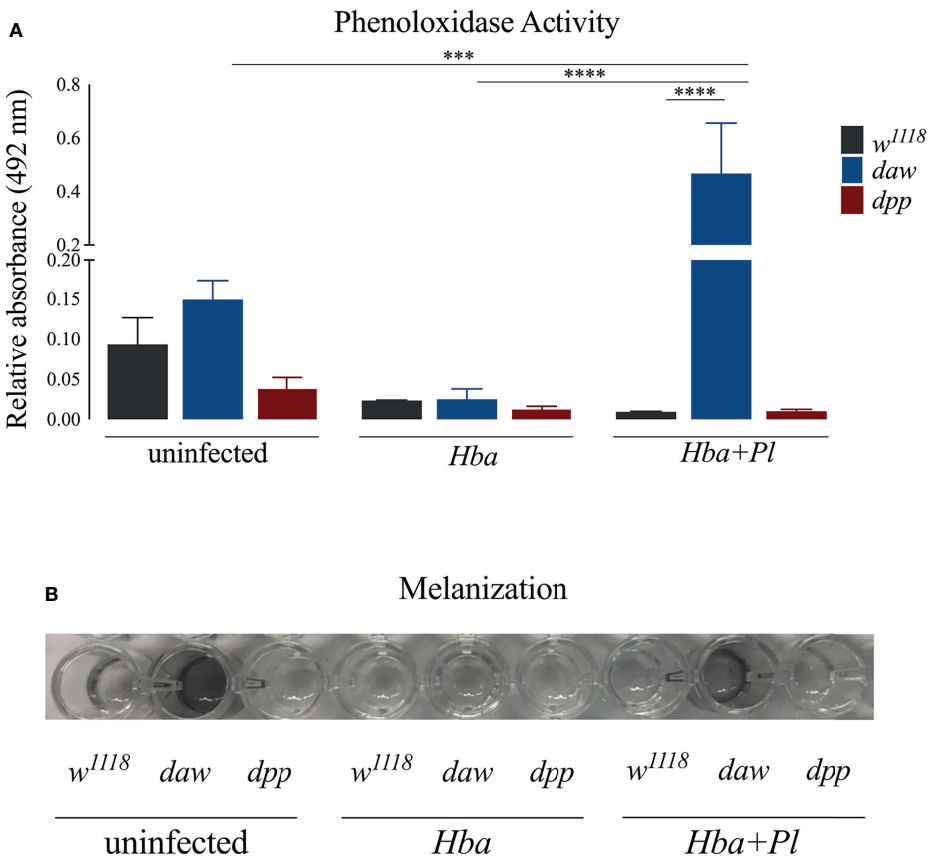


FIGURE 5 | Assessment of phenoloxidase (PO) activity and melanization response in *Drosophila melanogaster* TGF-ß mutant larvae following infection with axenic (*Hba*) or symbiotic (*Hba+Pl*) *Heterorhabditis bacteriophora*. **(A)** PO activity in the hemolymph plasma of *daw* and *dpp* mutant larvae and their background controls (w^{1118}) infected with either *H. bacteriophora* axenic or symbiotic nematodes. Significance levels were assessed using one-way analysis of variance (ANOVA), (***p = 0.0003; ****p < 0.0001). **(B)** Melanization of hemolymph plasma of *daw* and *dpp* mutant larvae and their background controls (w^{1118}) infected with *H. bacteriophora* axenic or symbiotic nematodes one hour post incubation at room temperature.

DISCUSSION

Previous studies have demonstrated that the conserved TGF-ß signaling pathway is induced in *D. melanogaster* larvae and adult flies during infection with EPNs (12, 13). Also, activin signaling activity decreases *DUOX* expression and increased PO activity in *D. melanogaster* larvae following infection with *H. gerrardi* symbiotic nematodes (14). Therefore, it is important to explore any additional roles that activin as well as BMP signaling may play in regulating cellular, humoral, and melanization immune responses in *D. melanogaster* during parasitic nematode infection. Identification of the precise molecular mechanisms underlying these processes, specifically against infection with the nematode parasites without any input from their mutualistic bacteria will contribute to a more comprehensive understanding of the host defense capability against parasitic diseases. Here, we used symbiotic together with axenic *H. bacteriophora* nematodes to examine the participation of the activin and the BMP signaling branches in the regulation of the *D melanogaster* anti-nematode immune response. We report that BMP signaling activity promotes survival against symbiotic *H. bacteriophora* infection but reduces survival against axenic nematode infection. Also, BMP signaling activity reduces *DUOX* expression but increases *Diptericin* and *Drosomycin* expression upon infection of *D. melanogaster* larvae with axenic *H. bacteriophora*. In addition, a functional activin pathway increases *Drosomycin* expression upon axenic nematode infection and decreases *Diptericin* and *Cecropin* expression upon symbiotic nematode infection. Finally, we report that activin signaling activity modulates the melanization response by reducing the PO enzyme levels upon symbiotic *H. bacteriophora* nematode parasitism (**Figure 6**).

Heterorhabditis nematode invasion through the cuticle causes injury to the insect host and the wounding response, which is characterized by the loss of tissue integrity, is observed throughout the infection (42). Previous work has indicated that in *D. melanogaster* larvae, *dpp* is transcriptionally induced by injury and might have a role in attenuating immune responses following wounding and infection (10). In this study, we found that *dpp* mutant larvae survive better to axenic *H. bacteriophora* infection relative to their background controls. A possible

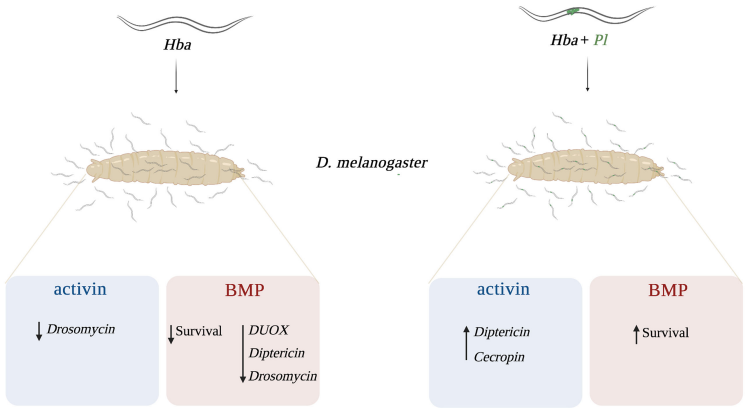


FIGURE 6 | Proposed model of the interaction between activin and BMP signaling branches of the TGF-ß signaling pathway and *Drosophila melanogaster* immunity against *Heterorhabditis bacteriophora* parasitic nematodes. In response to axenic (*Hba*) *H. bacteriophora* infection, activin signaling activity reduces the expression of *Drosomycin* while BMP signaling activity reduces the survival ability and the expression of *dual oxidase* (*DUOX*), *Diptericin*, and *Drosomycin*. Following symbiotic (*Hba* + *Pl*) *H. bacteriophora* infection, activin signaling activity promotes the expression of *Diptericin* and *Cecropin* whereas BMP signaling activity promotes the survival of *D. melanogaster* larvae.

explanation for this observation could be that the injury caused by IJs during entry into *D. melanogaster* larvae, which may lead to the activation of humoral and cellular immune reactions.

Humoral immunity in *D. melanogaster* is characterized by the synthesis and secretion of AMPs into the hemolymph through the activation of two NF-κB signaling pathways, Toll and Imd (36). While the Toll signaling is predominantly activated by Gram-positive bacteria and fungi, the Imd signaling is mainly activated by Gram-negative bacteria (43–46). The Imd pathway has also been previously implicated in the immune response of *D. melanogaster* to *H. bacteriophora*. More precisely, *Diptericin*, the readout AMP-encoding gene for the Imd pathway, is upregulated at 6 h post-injection of *D. melanogaster* adult flies and larvae with excreted secreted products (ESPs) isolated from *H. bacteriophora* axenic nematodes (22, 36). In addition, feeding of *D. melanogaster* larvae with *P. luminescens* bacteria also leads to upregulation of *Diptericin* (19). Interestingly, we have found that expression of *Diptericin* is lower in w^{1118} larvae (background control) following symbiotic *H. bacteriophora* infection compared to those infected with axenic nematodes. It is possible that at the 24 h time point, molecules produced by the nematode-bacteria pair during infection are able to suppress *Diptericin* expression. Another plausible explanation could be that *Diptericin* expression starts to decrease in larvae at 6 h post symbiotic *H. bacteriophora* infection, similar to *Diptericin* expression decrease in adult flies infected with *Escherichia coli* or *Micrococcus luteus* (47). We also demonstrate that *Diptericin* and *Cecropin* are upregulated following symbiotic *H. bacteriophora* infection in activin deficient larvae compared to the background control line. This implies that activin signaling activity interacts with the Imd pathway in *D. melanogaster* larvae during infection with the *H. bacteriophora-P. luminescens* complex to modify the AMP response. However, additional work is required to fully elucidate if Imd and activin signaling in *D. melanogaster* work synergistically against parasitic nematode infection. It would be interesting to distinguish the factors produced by the nematodes and their mutualistic bacteria to determine those that lead to *Diptericin* upregulation in *daw* mutants. The two virulence factors secreted by *H. bacteriophora* nematodes, Ecdysteroid glycosyltransferase Hba_07292 and Serine carboxypeptidase (Hba_11636) are able to suppress *Diptericin* expression in *D. melanogaster* larvae (40, 48). Investigating whether injection of these factors into *daw* mutant larvae is associated with any changes in *Diptericin* expression would provide insight about the impact of ESPs on activin-Imd interaction.

Previous work has shown that in the absence of infection, BMP signaling is regulated by Toll signaling activity *via* binding of Dorsal to *Dpp* and suppressing its expression in the ventral domain of the embryo (49). The interplay between these two signaling pathways is essential for *D. melanogaster* development (50). Here, we suggest that interaction between TGF-ß signaling branches and Toll signaling could be further extended to the immune response against parasitic nematode infection. This is particularly due to our observation of the reduced expression of *Drosomycin* in both *daw* and *dpp* mutants upon axenic *H. bacteriophora* infection. Our findings provide a basis for future investigations aimed at elucidating interactions between activin and BMP signaling activity with the Toll pathway during the *D. melanogaster* response to *H. bacteriophora*. Toll signaling receptors, GNBP1, PGRP-SA and PGRP-SD are upregulated following infection with symbiotic *H. bacteriophora* nematodes in *D. melanogaster* adult flies (36, 51). To identify potential interactions between TGF-ß and Toll signaling at the level of recognition, changes in the expression of *daw* and *dpp* following infection with *H. bacteriophora* nematodes or injection with ESPs can be studied in flies deficient of GNBP1, PGRP-SA or PGRP-SD. Moreover, to identify the interactions at downstream expression, flies deficient of the Toll pathway transcription factor Dif could be used to assess the expression of *daw* and *dpp* or

other intracellular TGF-ß signaling components in the context of response to parasitic nematode infection (36). Lessons we learn through studying these interactions could be applicable to higher organisms, including humans, considering the similarities between innate immunity in *D. melanogaster* and mammals and the phylogenetic relationship between *H. bacteriophora* nematodes and vertebrate parasitic nematodes including *Necator*, *Dictyocaulus*, and *Oslerus* (52–54).

In our study we found that *D. melanogaster* activin deficient larvae have increased levels of PO as well as melanin formation in their hemolymph plasma compared to their background controls following infection with symbiotic *H. bacteriophora*. This finding suggests that activin signaling activity reduces the melanization response in *D. melanogaster* larvae upon symbiotic nematode infection. In contrast, we have not found any significant differences in *daw* mutants following axenic *H. bacteriophora* infection suggesting that activin signaling activity reduces the melanization response only when nematodes infect *D. melanogaster* larvae together with their mutualistic bacteria. Interestingly, a recent study has shown that following infection of *daw* mutant larvae with symbiotic *H. gerrardi* nematodes, PO activity remains at low levels (14). It is therefore possible that the property of functional activin signaling to decrease the melanization response in *D. melanogaster* larvae is restricted only to the response against the *H. bacteriophora-P. luminescens* complex. It would be interesting to further explore how other parasitic nematodes influence the interaction between activin signaling activity and the melanization response by testing the response of *D. melanogaster* to other EPNs such as *S. carpocapsae*. Infection of *D. melanogaster* larvae with *S. carpocapsae* nematodes leads to increased PO activity in the hemolymph (23). In contrast, inoculation of the *S. carpocapsae* mutualistic bacteria, *Xenorhabdus nematophila*, *via* pricking into *D. suzukii* larvae results in lower PO activity in the hemolymph (55). Based on this information, it would be interesting to test whether activin signaling activity has a role in regulating the *D. melanogaster* PO activity and melanization response against the *S. carpocapsae-X. nematophila* complex. In addition, considering the role of melanization in clot formation, another field of inquiry would be to study whether activin signaling interacts with the clotting response in *D. melanogaster* against parasitic nematode infection (56). Clotting factors such as Fondue, Eig71Ee, and transglutaminase protect *D. melanogaster* larvae against symbiotic *H. bacteriophora* infection (57, 58). Future studies could focus on understanding if this protection extends to the mutants of activin signaling suggesting an interaction between clotting and TGF-ß signaling activity during the *D. melanogaster* response to EPN infection. This is a particularly interesting avenue of research because not only it can provide important insight about the regulation of clotting response in mammals against infection with nematode parasites but also it can give clues about the evolutionary basis of anti-parasitic immune response.

Although we cannot exclude the possibility that the current observations may be due to the different genetic background of the mutants as compared to the control flies and rescue experiments may be needed to confirm the observed phenotypes, our findings demonstrate novel functions for activin and BMP signaling activity in regulating the *D. melanogaster* immune response to parasitic nematode infection. In addition to paving the way to a better understanding of host-parasite interactions and the evolution of the immune role of TGF-ß signaling, results obtained from this study could have broader impacts. Parasitic nematodes pose a major threat to human health, and thus identification of the key immune signaling components that oppose infection against parasitic nematodes will provide important clues for the development of novel treatment strategies against parasitic diseases.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YO designed and conducted the experiments, analyzed the data, constructed the figures, interpreted the results, and wrote drafts of the manuscript. DR conducted parts of the experiments. IE designed the experiments, interpreted the results, and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the National Institute of Allergy and Infectious Diseases (grants 1R01AI110675 and 1R56AI110675) and the National Science Foundation (grant 2019869).

ACKNOWLEDGMENTS

We thank Zachary Stickelman for maintaining and amplifying the laboratory fly lines and members of the Department of Biological Sciences at George Washington University for providing feedback to the project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2021.795331/full#supplementary-material

REFERENCES

1. Huminiecki L, Goldovsky L, Freilich S, Moustakas A, Ouzounis C, Heldin CH. Emergence, Development and Diversification of the TGF-β Signalling Pathway Within the Animal Kingdom. *BMC Evol Biol* (2009) 9:28. doi: 10.1186/1471-2148-9-28
2. Pang K, Ryan JF, Baxevanis AD, Martindale MQ. Evolution of the TGF-β Signaling Pathway and its Potential Role in the Ctenophore, *Mnemiopsis leidyi*. *PLoS One* (2011) 6:e24152. doi: 10.1371/journal.pone.0024152
3. Tzavlaki K, Moustakas A. TGF- β Signaling. *Biomolecules* (2020) 3:487. doi: 10.3390/biom10030487
4. Masucci JD, Miltenberger RJ, Hoffmann FM. Pattern-Specific Expression of the *Drosophila* Decapentaplegic Gene in Imaginal Disks Is Regulated by 3' Cis-Regulatory Elements. *Genes Dev* (1990) 4:2011–23. doi: 10.1101/gad.4.11.2011
5. Montanari MP, Tran NV, Shimmi O. Regulation of Spatial Distribution of BMP Ligands for Pattern Formation. *Dev Dyn* (2021). doi: 10.1002/dvdy.397
6. Dobens LL, Raftery LA. Drosophila Oogenesis: A Model System to Understand TGF-ß/Dpp Directed Cell Morphogenesis. *Ann NY Acad Sci* (1998) 857:245–7. doi: 10.1111/j.1749-6632.1998.tb10123.x
7. Campbell G. Distalization of the *Drosophila* Leg by Graded EGF-Receptor Activity. *Nature* (2002) 418:781–5. doi: 10.1038/nature00971
8. Upadhyay A, Moss-Taylor L, Kim M-J, Ghosh AC, O'Connor MB. TGF-β Family Signaling in *Drosophila*. *Cold Spring Harb Perspect Biol* (2017) 9: a022152. doi: 10.1101/cshperspect.a022152
9. Zi Z, Chapnick DA, Liu X. Dynamics of TGF-ß/Smad Signaling. *FEBS Lett* (2012) 586:1921–8. doi: 10.1016/j.febslet.2012.03.063
10. Clark RI, Woodcock KJ, Geissmann F, Trouillet C, Dionne MS. Multiple TGF-β Superfamily Signals Modulate the Adult *Drosophila* Immune Response. *Curr Biol* (2011) 21:1672–7. doi: 10.1016/j.cub.2011.08.048
11. Patrnogic J, Heryanto C, Eleftherianos I. Wounding-Induced Upregulation of the Bone Morphogenic Protein Signaling Pathway in *Drosophila* Promotes Survival Against Parasitic Nematode Infection. *Gene* (2018) 673:112–8. doi: 10.1016/j.gene.2018.06.052
12. Eleftherianos I, Castillo JC, Patrnogic J. TGF-β Signaling Regulates Resistance to Parasitic Nematode Infection in *Drosophila melanogaster*. *Immunobiol* (2016) 221:1362–8. doi: 10.1016/j.imbio.2016.07.011
13. Patrnogic J, Heryanto C, Eleftherianos I. Transcriptional Up-Regulation of the TGF-β Intracellular Signaling Transducer Mad of *Drosophila* Larvae in Response to Parasitic Nematode Infection. *Innate Immun* (2018) 24:349–56. doi: 10.1177/1753425918790663
14. Ozakman Y, Eleftherianos I. TGF-β Signaling Interferes With the *Drosophila* Innate Immune and Metabolic Response to Parasitic Nematode Infection. *Front Physiol* (2019) 10:716. doi: 10.3389/fphys.2019.00716
15. Ozakman Y, Eleftherianos I. Nematode Infection and Antinematode Immunity in *Drosophila*. *Trends Parasitol* (2021) 37:1002–13. doi: 10.1016/j.pt.2021.06.001
16. Ciche TA, Ensign JC. For the Insect Pathogen *Photorhabdus luminescens*, Which End of a Nematode Is Out? *Appl Environ Microbiol* (2003) 69:1890–7. doi: 10.1128/AEM.69.4.1890-1897.2003
17. Stock SP, Blair HG. Entomopathogenic Nematodes and Their Bacterial Symbionts: The Inside Out of a Mutualistic Association. *Symbiosis* (2008) 46:65–75.
18. Castillo JC, Reynolds SE, Eleftherianos I. Insect Immune Responses to Nematode Parasites. *Trends Parasitol* (2011) 27:537–47. doi: 10.1016/j.pt.2011.09.001
19. Hallem EA, Rengarajan M, Ciche TAA, Sternberg PW. Nematodes, Bacteria, and Flies: A Tripartite Model for Nematode Parasitism. *Curr Biol* (2007) 17:898–904. doi: 10.1016/j.cub.2007.04.027
20. Castillo JC, Shokal U, Eleftherianos I. A Novel Method for Infecting *Drosophila* Adult Flies With Insect Pathogenic Nematodes. *Virulence* (2012) 3:339–47. doi: 10.4161/viru.20244
21. White GF. A Method for Obtaining Infective Nematode Larvae From Cultures. *Science* (1927) 66:302–3. doi: 10.1126/science.66.1709.302-a
22. Kenney E, Hawdon JM, O'Halloran D, Eleftherianos I. *Heterorhabditis bacteriophora* Excreted-Secreted Products Enable Infection by *Photorhabdus luminescens* Through Suppression of the Imd Pathway. *Front Immunol* (2019) 10:2372. doi: 10.3389/fimmu.2019.02372
23. Cooper D, Wuebbolt C, Heryanto C, Eleftherianos I. The Prophenoloxidase System in *Drosophila* Participates in the Anti-Nematode Immune Response. *Mol Immunol* (2019) 109:88–98. doi: 10.1016/j.molimm.2019.03.008
24. Honti V, Csordás G, Kurucz É, Márkus R, Andó I. The Cell-Mediated Immunity of *Drosophila melanogaster*: Hemocyte Lineages, Immune Compartments, Microanatomy and Regulation. *Dev Comp Immunol* (2014) 42:47–56. doi: 10.1016/j.dci.2013.06.005
25. Gold KS, Brückner K. Macrophages and Cellular Immunity in *Drosophila melanogaster*. *Semin Immunol* (2015) 27:357–68. doi: 10.1016/j.smim.2016.03.010
26. Eleftherianos I, More K, Spivack S, Paulin E, Khojandi A, Shukla S. Nitric Oxide Levels Regulate the Immune Response of *Drosophila melanogaster* Reference Laboratory Strains to Bacterial Infections. *Infect Immun* (2014) 82:4169–81. doi: 10.1128/IAI.02318-14
27. Vlisidou I, Wood W. *Drosophila* Blood Cells and Their Role in Immune Responses. *FEBS J* (2015) 282:1368–82. doi: 10.1111/febs.13235
28. Shokal U, Kopydlowski H, Eleftherianos I. The Distinct Function of Tep2 and Tep6 in the Immune Defense of *Drosophila melanogaster* Against the Pathogen *Photorhabdus*. *Virulence* (2017) 8:1668–82. doi: 10.1080/21505594.2017.1330240
29. Lavine MD, Strand MR. Insect Hemocytes and Their Role in Immunity. *Insect Biochem Mol Biol* (2002) 32:1295–309. doi: 10.1016/S0965-1748(02)00092-9
30. Rizki TM, Rizki RM. Lamellocyte Differentiation in *Drosophila* Larvae Parasitized by Leptopilina. *Dev Comp Immunol* (1992) 16:103–10. doi: 10.1016/0145-305X(92)90011-Z
31. Williams MJ. *Drosophila* Hemopoiesis and Cellular Immunity. *J Immunol* (2007) 178:4711–6. doi: 10.4049/jimmunol.178.8.4711
32. Rizki RM, Rizki TM. Effects of Lamellolysin From a Parasitoid Wasp on *Drosophila* Blood Cells *In Vitro*. *J Exp Zool* (1991) 2:236–44. doi: 10.1002/jez.140257021
33. Myers AL, Harris CM, Choe K-M, Brennan CA. Inflammatory Production of Reactive Oxygen Species by *Drosophila* Hemocytes Activates Cellular Immune Defenses. *Biochem Biophys Res Commun* (2018) 505:726–32. doi: 10.1016/j.bbrc.2018.09.126
34. Kim S-H, Lee W-J. Role of DUOX in Gut Inflammation: Lessons From *Drosophila* Model of Gut-Microbiota Interactions. *Front Cell Infect Microbiol* (2014) 3:116. doi: 10.3389/fcimb.2013.00116
35. Lee K-A, Cho K-C, Kim B, Jang I-H, Nam K, Kwon YE, et al. Inflammation-Modulated Metabolic Reprogramming Is Required for DUOX-Dependent Gut Immunity in *Drosophila*. *Cell Host Microbe* (2018) 23:338–352.e5. doi: 10.1016/j.chom.2018.01.011
36. Ferrandon D, Imler J-L, Hetru C, Hoffmann JA. The *Drosophila* Systemic Immune Response: Sensing and Signalling During Bacterial and Fungal Infections. *Nat Rev Immunol* (2007) 7:862–74. doi: 10.1038/nri2194
37. Patrnogic J, Heryanto C, Ozakman Y, Eleftherianos I. Transcript Analysis Reveals the Involvement of NF-κb Transcription Factors for the Activation of TGF-β Signaling in Nematode-Infected *Drosophila*. *Immunogenetics* (2019) 71:501–10. doi: 10.1007/s00251-019-01119-8
38. Imler JL, Bulet P. Antimicrobial Peptides in *Drosophila*: Structures, Activities and Gene Regulation. *Chem Immunol Allergy* (2005) 86:1–21. doi: 10.1159/000086648
39. Binggeli O, Neyen C, Poidevin M, Lemaitre B. Prophenoloxidase Activation Is Required for Survival to Microbial Infections in *Drosophila*. *PLoS Path* (2014) 10:e1004067. doi: 10.1371/journal.ppat.1004067
40. Kenney E, Yaparla A, Hawdon JM, O' Halloran DM, Grayfer L, Eleftherianos I. A Putative Lysozyme and Serine Carboxypeptidase From *Heterorhabditis bacteriophora* Show Differential Virulence Capacities in *Drosophila melanogaster*. *Dev Comp Immunol* (2021) 114:103820. doi: 10.1016/j.dci.2020.103820
41. Dudzic JP, Hanson MA, Iatsenko I, Kondo S, Lemaitre B. More Than Black or White: Melanization and Toll Share Regulatory Serine Proteases in *Drosophila*. *Cell Rep* (2019) 27:1050–61.e3. doi: 10.1016/j.celrep.2019.03.101
42. Dziedziech A, Shivankar S, Theopold U. High-Resolution Infection Kinetics of Entomopathogenic Nematodes Entering *Drosophila melanogaster*. *Insects* (2020) 1:60. doi: 10.3390/insects11010060
43. Michel T, Reichhart J-M, Hoffmann JA, Royet J. *Drosophila* Toll Is Activated by Gram-Positive Bacteria Through a Circulating Peptidoglycan Recognition Protein. *Nature* (2001) 414:756–9. doi: 10.1038/414756a
44. Gottar M, Gobert V, Matskevich AA, Reichhart J-M, Wang C, Butt TM, et al. Dual Detection of Fungal Infections in *Drosophila via* Recognition of Glucans

and Sensing of Virulence Factors. *Cell* (2006) 127:1425–37. doi: 10.1016/j.cell.2006.10.046
45. Lemaitre B, Kromer-Metzger E, Michaut L, Nicolas E, Meister M, Georgel P, et al. A Recessive Mutation, Immune Deficiency (Imd), Defines Two Distinct Control Pathways in the *Drosophila* Host Defense. *Proc Natl Acad Sci USA* (1995) 92:9465–9. doi: 10.1073/pnas.92.21.9465
46. Kleino A, Silverman N. The *Drosophila* IMD Pathway in the Activation of the Humoral Immune Response. *Dev Comp Immunol* (2014) 42:25–35. doi: 10.1016/j.dci.2013.05.014
47. Lemaitre B, Reichhart J, Hoffmann JA. *Drosophila* Host Defense: Differential Induction of Antimicrobial Peptide Genes After Infection by Various Classes of Microorganisms. *Proc Natl Acad Sci USA* (1997) 26:14614–9. doi: 10.1073/pnas.94.26.14614
48. Kenney E, Yaparla A, Hawdon JM, O'Halloran DM, Grayfer L, Eleftherianos I. A Putative UDP-Glycosyltransferase From *Heterorhabditis bacteriophora* Suppresses Antimicrobial Peptide Gene Expression and Factors Related to Ecdysone Signaling. *Sci Rep* (2020) 10:12312. doi: 10.1038/s41598-020-69306-212
49. Xiao C, Shim J, Klüppel M, Zhang SS, Dong C, Flavell RA, et al. Ecsit Is Required for Bmp Signaling and Mesoderm Formation During Mouse Embryogenesis. *Genes Dev* (2003) 23:2933–49. doi: 10.1101/gad.1145603.lated
50. Pechmann M, Kenny NJ, Pott L, Heger P, Lynch J, Roth S, et al. Striking Parallels Between Dorsoventral Patterning in *Drosophila* and *Gryllus* Reveal a Complex Evolutionary History Behind a Model Gene Regulatory Network. *eLife* (2021) 10:e68287. doi: 10.7554/eLife.68287
51. Castillo JC, Creasy T, Kumari P, Shetty A, Shokal U, Tallon LJ, et al. *Drosophila* Anti-Nematode and Anti- Bacterial Immune Regulators Revealed by RNA-Seq. *BMC Genomics* (2015) 16:1–21. doi: 10.1186/s12864-015-1690-2
52. Plotkin S, Diemert D, Bethony JM, Hotez PJ. Hookworm Vaccines. *Clin Infect Dis* (2008) 46:282–8. doi: 10.1086/524070
53. Panuska C. Lungworms of Ruminants. *Vet Clin North Am Food Anim Pract* (2006) 22:583–93. doi: 10.1016/j.cvfa.2006.06.002
54. Vieson MD, Piñeyro P, LeRoith T. A Review of the Pathology and Treatment of Canine Respiratory Infections. *Vet Med* (2012) 3:25–39. doi: 10.2147/VMRR.S25021
55. Garriga A, Mastore M, Morton A, Del Pino FG, Brivio MF. Immune Response of *Drosophila suzukii* Larvae to Infection With the Nematobacterial Complex *Steinernema carpocapsae–Xenorhabdus nematophila*. *Insects* (2020) 11:210. doi: 10.3390/insects11040210
56. Bidla G, Lindgren M, Theopold U, Dushay MS. Hemolymph Coagulation and Phenoloxidase in *Drosophila* Larvae. *Dev Comp Immunol* (2005) 8:669–79. doi: 10.1016/j.dci.2004.11.007
57. Wang Z, Wilhelmsson C, Hyrsl P, Loof TG, Dobes P, Klupp M, et al. Pathogen Entrapment by Transglutaminase – A Conserved Early Innate Immune Mechanism. *PLoS Pathog* (2010) 2:e1000763. doi: 10.1371/journal.ppat.1000763
58. Hyrsl P, Dobesa P, Wang Z, Haulinga T, Wilhelmssona C, Theopold U. Clotting Factors and Eicosanoids Protect Against Nematode Infections. *J Innate Immun* (2011) 3:65–70. doi: 10.1159/000320634

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership