

HOST AND MICROBE ADAPTATIONS IN THE EVOLUTION OF IMMUNITY

EDITED BY: Larry J. Dishaw and Gary W. Litman
PUBLISHED IN: Frontiers in Immunology





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-022-6

DOI 10.3389/978-2-88963-022-6

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: researchtopics@frontiersin.org

HOST AND MICROBE ADAPTATIONS IN THE EVOLUTION OF IMMUNITY

Topic Editors:

Larry J. Dishaw, University of South Florida St. Petersburg, United States

Gary W. Litman, University of South Florida, United States

The evolution of metazoans has been accompanied by new interfaces with the microbial environment that include biological barriers and surveillance by specialized cell types. Increasingly complex organisms require increased capacities to confront pathogens, achieved by co-evolution of recognition mechanisms and regulatory pathways. Two distinct but interactive forms of immunity have evolved. Innate immunity, shared by all metazoans, is traditionally viewed as simple and non-specific. Adaptive immunity possesses the capacity to anticipate new infectious challenges and recall previous exposures; the most well-understood example of such a system, exhibited by lymphocytes of vertebrates, is based on somatic gene alterations that generate extraordinary specificity in discrimination of molecular structures. Our understanding of immune phylogeny over the past decades has tried to reconcile immunity from a vertebrate standpoint. While informative, such approaches cannot completely address the complex nature of selective pressures brought to bear by the complex microbiota (including pathogens) that co-exist with all metazoans.

In recent years, comparative studies (and new technologies) have broadened our concepts of immunity from a systems-wide perspective. Unexpected findings, e.g., genetic expansions of innate receptors, high levels of polymorphism, RNA-based forms of generating diversity, adaptive evolution and functional divergence of gene families and the recognition of novel mediators of adaptive immunity, prompt us to reconsider the very nature of immunity. Even fundamental paradigms as to how the jawed vertebrate adaptive immune system should be structured for “optimal” recognition potential have been disrupted more than once (e.g., the discovery of the multicluster organization and germline joining of immunoglobulin genes in sharks, gene conversion as a mechanism of somatic diversification, absence of IgM or MHC II in certain teleost fishes). Mechanistically, concepts of innate immune memory, often referred to as “trained memory,” have been realized further, with the development of new discoveries in studies of epigenetic regulation of somatic lineages. Immune systems innovate and adapt in a taxon-specific manner, driven by the complexity of interactions with microbial symbionts (commensals, mutualists and pathogens). Immune systems are shaped by selective forces that reflect consequences of dynamic interactions with microbial environments as well as a capacity for rapid change that can be facilitated by genomic instabilities. We have learned that characterizing receptors and receptor interactions is not necessarily the most significant component in understanding the evolution of immunity. Rather, such a subject needs to be understood from a more global perspective and will necessitate re-consideration of the physical barriers that afford protection and the developmental processes that create them.

By far, the most significant paradigm shifts in our understanding of immunity and the infection process has been that microbes no longer are considered to be an automatic cause or consequence of illness, but rather integral components of normal physiology and homeostasis. Immune phylogeny has been shaped not only by an arms race with pathogens but also perhaps by mutualistic interactions with resident microbes. This Research Topic updates and extends the previous eBook on Changing Views of the Evolution of Immunity and contains peer-reviewed submissions of original research, reviews and opinions.

Citation: Dishaw, L. J., Litman, G. W., eds. (2019). Host and Microbe Adaptations in the Evolution of Immunity. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88963-022-6

Table of Contents

- 06** *Profiling of Human Molecular Pathways Affected by Retrotransposons at the Level of Regulation by Transcription Factor Proteins*
Daniil Nikitin, Dmitry Penzar, Andrew Garazha, Maxim Sorokin, Victor Tkachev, Nicolas Borisov, Alexander Poltorak, Vladimir Prassolov and Anton A. Buzdin
- 20** *Regular Wounding in a Natural System: Bacteria Associated With Reproductive Organs of Bedbugs and Their Quorum Sensing Abilities*
Oliver Ott, Peter Deines, Katrin Hammerschmidt and Klaus Reinhardt
- 31** *F-Type Lectins: A Highly Diversified Family of Fucose-Binding Proteins With a Unique Sequence Motif and Structural Fold, Involved in Self/Non-Self-Recognition*
Gerardo R. Vasta, L. Mario Amzel, Mario A. Bianchet, Matteo Cammarata, Chiguang Feng and Keiko Saito
- 49** *Long Non-Coding RNAs: Emerging and Versatile Regulators in Host–Virus Interactions*
Xing-Yu Meng, Yuzi Luo, Muhammad Naveed Anwar, Yuan Sun, Yao Gao, Huawei Zhang, Muhammad Munir and Hua-Ji Qiu
- 59** *The Intra-Dependence of Viruses and the Holobiont*
Juris A. Grasis
- 66** *An Organismal Model for Gene Regulatory Networks in the Gut-Associated Immune Response*
Katherine M. Buckley and Jonathan P. Rast
- 74** *Specific Pathogen Recognition by Multiple Innate Immune Sensors in an Invertebrate*
Guillaume Tetreau, Silvain Pinaud, Anaïs Portet, Richard Galinier, Benjamin Gourbal and David Duval
- 86** *The Value of a Comparative Approach to Understand the Complex Interplay Between Microbiota and Host Immunity*
Norma M. Morella and Britt Koskella
- 97** *Neutrophil Evolution and Their Diseases in Humans*
Jennifer W. Leiding
- 105** *Alkaline Phosphatase, an Unconventional Immune Protein*
Bethany A. Rader
- 111** *Patterns of Early-Life Gut Microbial Colonization During Human Immune Development: An Ecological Perspective*
Isabelle Laforest-Lapointe and Marie-Claire Arrieta
- 124** *The SpTransformer Gene Family (Formerly Sp185/333) in the Purple Sea Urchin and the Functional Diversity of the Anti-Pathogen rSpTransformer-E1 Protein*
L. Courtney Smith and Cheng Man Lun
- 144** *Evolution and Function of Thioester-Containing Proteins and the Complement System in the Innate Immune Response*
Upasana Shokal and Ioannis Eleftherianos
- 153** *Immunity in Protochordates: The Tunicate Perspective*
Nicola Franchi and Lorian Ballarin

- 169** *Dscam1 in Pancrustacean Immunity: Current Status and a Look to the Future*
Sophie A. O. Armitage, Joachim Kurtz, Daniela Brites, Yuemei Dong, Louis Du Pasquier and Han-Ching Wang
- 182** *Leaky Gut as a Danger Signal for Autoimmune Diseases*
Qinghui Mu, Jay Kirby, Christopher M. Reilly and Xin M. Luo
- 192** *Under Pressure: Interactions Between Commensal Microbiota and the Teleost Immune System*
Cecelia Kelly and Irene Salinas
- 201** *The Recombinant Sea Urchin Immune Effector Protein, rSpTransformer-E1, Binds to Phosphatidic Acid and Deforms Membranes*
Cheng Man Lun, Robin L. Samuel, Susan D. Gillmor, Anthony Boyd and L. Courtney Smith
- 216** *Immunometabolic Phenotype Alterations Associated With the Induction of Disease Tolerance and Persistent Asymptomatic Infection of Salmonella in the Chicken Intestine*
Michael H. Kogut and Ryan J. Arsenault
- 223** *Evidence of an Antimicrobial Peptide Signature Encrypted in HECT E3 Ubiquitin Ligases*
Ivan Lavander Candido-Ferreira, Thales Kronenberger, Raphael Santa Rosa Sayegh, Isabel de Fátima Correia Batista and Pedro Ismael da Silva Junior
- 237** *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*
Arjun Balakrishnan, Markus Schnare and Dipshikha Chakravorty



Profiling of Human Molecular Pathways Affected by Retrotransposons at the Level of Regulation by Transcription Factor Proteins

Daniil Nikitin^{1,2}, Dmitry Penzar³, Andrew Garazha^{2,4}, Maxim Sorokin^{4,5,6}, Victor Tkachev⁴, Nicolas Borisov^{4,5}, Alexander Poltorak⁷, Vladimir Prassolov¹ and Anton A. Buzdin^{1,2,4,5*}

¹ Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia, ² D. Rogachev Federal Research Center of Pediatric Hematology, Oncology and Immunology, Moscow, Russia, ³ The Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia, ⁴ OmicsWay Corp., Walnut, CA, United States, ⁵ National Research Centre Kurchatov Institute, Centre for Convergence of Nano-, Bio-, Information and Cognitive Sciences and Technologies, Moscow, Russia, ⁶ Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, ⁷ Program in Immunology, Sackler Graduate School, Tufts University, Boston, MA, United States

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Harold Charles Smith,
University of Rochester School
of Medicine and Dentistry,
United States
Yoshinao Kubo,
Nagasaki University, Japan

*Correspondence:

Anton A. Buzdin
buzdin@oncobox.com

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 19 October 2017

Accepted: 04 January 2018

Published: 30 January 2018

Citation:

Nikitin D, Penzar D, Garazha A, Sorokin M, Tkachev V, Borisov N, Poltorak A, Prassolov V and Buzdin AA (2018) Profiling of Human Molecular Pathways Affected by Retrotransposons at the Level of Regulation by Transcription Factor Proteins. *Front. Immunol.* 9:30. doi: 10.3389/fimmu.2018.00030

Endogenous retroviruses and retrotransposons also termed retroelements (REs) are mobile genetic elements that were active until recently in human genome evolution. REs regulate gene expression by actively reshaping chromatin structure or by directly providing transcription factor binding sites (TFBSs). We aimed to identify molecular processes most deeply impacted by the REs in human cells at the level of TFBS regulation. By using ENCODE data, we identified ~2 million TFBS overlapping with putatively regulation-competent human REs located in 5-kb gene promoter neighborhood (~17% of all TFBS in promoter neighborhoods; ~9% of all RE-linked TFBS). Most of REs hosting TFBS were highly diverged repeats, and for the evolutionary young (0–8% diverged) elements we identified only ~7% of all RE-linked TFBS. The gene-specific distributions of RE-linked TFBS generally correlated with the distributions for all TFBS. However, several groups of molecular processes were highly enriched in the RE-linked TFBS regulation. They were strongly connected with the immunity and response to pathogens, with the negative regulation of gene transcription, ubiquitination, and protein degradation, extracellular matrix organization, regulation of STAT signaling, fatty acids metabolism, regulation of GTPase activity, protein targeting to Golgi, regulation of cell division and differentiation, development and functioning of perception organs and reproductive system. By contrast, the processes most weakly affected by the REs were linked with the conservative aspects of embryo development. We also identified differences in the regulation features by the younger and older fractions of the REs. The regulation by the older fraction of the REs was linked mainly with the immunity, cell adhesion, cAMP, IGF1R, Notch, Wnt, and integrin signaling, neuronal development, chondroitin sulfate and heparin metabolism, and endocytosis. The younger REs regulate other aspects of immunity, cell cycle progression and apoptosis, PDGF, TGF beta, EGFR, and p38 signaling, transcriptional repression, structure of nuclear lumen, catabolism of phospholipids, and heterocyclic

molecules, insulin and AMPK signaling, retrograde Golgi-ER transport, and estrogen signaling. The immunity-linked pathways were highly represented in both categories, but their functional roles were different and did not overlap. Our results point to the most quickly evolving molecular pathways in the recent and ancient evolution of human genome.

Keywords: endogenous retrovirus, transcription factor binding site, retrotransposon, retroelement, molecular pathway, immunity, evolution, human genome evolution

INTRODUCTION

Retrotransposable elements (REs) are mobile genetic elements that self-reproduce in the host DNA. For proliferation of their copies, they use a specific molecular mechanism based on RNA-dependent synthesis of DNA by an enzyme termed reverse transcriptase (RT) (1). Taken together, REs occupy ~40% of human DNA. They are represented by three major classes: human endogenous retroviruses/LTR retrotransposons (HERV/LRs) and LINE and SINE retrotransposons (2). The first group shaped ~8% of human genome, whereas LINEs and SINEs ~20 and 13%, respectively. HERV/LRs are thought to be remnants of multiple previous retroviral infections (3, 4). Unlike many common infectious retroviruses, they became inheritable because their insertions occurred in the ancestral germ cells (5). By contrast, LINEs and SINEs are non-infective retrotransposons. HERV/LRs and LINEs are called autonomous mobile elements because they encode RT, and SINEs—non-autonomous because for their life cycle they use foreign, LINE-encoded enzymes (6).

The studies of evolutionary dynamics of REs revealed that they were actively proliferating in human DNA until the most recent events in human speciation (7). All groups of REs include transcription of their genomic copies as the necessary step in their life cycle. Therefore, RE sequences are enriched in transcription factor binding sites (TFBSs) and other regulatory motifs (8–11). Moreover, most of the RE copies accumulated mutations and could strengthen their regulatory repertoire. For example, the HERV/LRs include promoters (12), enhancers (13, 14), polyadenylation signals (5), chromatin folding reshapers (15), and binding sites for various nuclear proteins (16). In human genome, REs are represented by millions of individual elements that can be found in the vicinity of any known gene. Therefore, the REs are considered among the major factors of evolution of gene expression regulatory networks. For example, ~30% of all transcriptional factor p53 binding sites in the human genome fall within the HERV/LR elements (17). We recently showed that functional TFBS within the human-specific endogenous retroviruses may control expression of schizophrenia-linked gene *PRODH* in human hippocampus (14).

Transcription factor binding sites denote regulatory fragments of DNA that can bind transcription factors and influence gene expression. Congruently, mapping DNaseI hypersensitivity sites (DHS) became a golden standard for the identification of regulatory loci of an open chromatin (18). Recent studies evidence that huge numbers of DHS and TFBS in the human genome are located within the TEs. For example, totally, ~155,000 and ~320,000 HERV/LR-derived DHS and TFBS were identified, respectively

(19). For the HERV/LR elements, ~110,000 inserts (~15%) had at least 2 TFBS and ~140,000 individual inserts (~19%)—at least 1 DHS, as shown in the previous report (19). Finally, at least ~31% of all mapped human transcription start sites were identified within the REs (20).

Like never before, high-throughput mapping of functional genomic features such as TFBS, DHS, and different types of histone binding sites provides opportunity to explore RE influence on gene expression in a comprehensive way. Besides individual affected genes, their functional groups can be assayed, including gene families and molecular pathways. Intracellular molecular pathways are involved in all major events in the living organisms. The major groups are metabolic, cell signaling, cytoskeleton reorganization, and DNA repair pathways (21, 22).

The pathways may include tens or hundreds of nodes and aggregate up to several hundreds of different gene products (23, 24). Remarkably, each node in a pathway is typically built not by just a single-gene product, but rather by their groups. Those can be formed by the homologous families of similarly functionally charged proteins, or by the various protein subunits which may be all needed to execute a function required for the pathway activity (25, 26).

For few decades, the molecular pathways are still on the forefront of biomedical sciences (27–30). Hundreds of thousands of molecular interactions and thousands of molecular pathways have been discovered by the molecular biologists and cataloged in different databases (31–37).

On the other hand, the gene products can be sorted according to their functional role in the cell and with reference to the molecular or supramolecular processes they are involved. This way of data aggregation does not require knowledge of the particular chains of molecular interactions, as for the above group of the pathway databases. For example, the gene ontology (GO) database provides functional and structural labels to the gene products or their groups.¹ By uploading a specific set of gene products, one can find it out whether this list is statistically significantly enriched in certain types of functional gene families. For example, in certain applications this enables to make a quick overview of the differentially expressed and most frequently mutated groups of genes (38).

In this study, we aimed to identify molecular processes most deeply regulated by the RE inserts in the human cells. To this end, we mapped the available TFBS data on the individual human REs for K562 cells. We found that in the close gene neighborhood,

¹www.geneontology.org.

~17% of TFBS overlap with the RE sequences, of them 44% belong SINEs, 33%—to LINEs, and 23%—to LR/ERVs. Most of the REs hosting TFBS were highly diverged repeats, and for the evolutionary young (0–8% diverged) elements we identified only ~7% of all RE-specific TFBS. Among them, SINEs hosted ~68%, LINEs ~15%, and LR/ERVs ~17% of TFBS.

Depending on the number of RE-mapped TFBS in the vicinities of the particular genes, we calculated a score for each gene positively reflecting the RE impact on gene regulation. Based on the scores for the individual genes, for the first time we could identify the molecular processes most strongly impacted by the RE regulatory features. To this end, we applied and modified bioinformatic method Oncofinder that has been used before only for the analysis of gene or microRNA expression profiles (39) and could effectively reduce experimental noise caused by different experimental platforms and batch effects (40, 41). In the initial version, this method makes it possible to calculate the quantitative value reflecting molecular pathway activation, called pathway activation strength (PAS). The absolute value of PAS reflects the extent of a molecular pathway perturbation. Negative PAS values indicate downregulation of molecular pathways, positive values mean upregulation, whereas 0 values represent non-significant difference with the control samples (42). Previously PAS values were calculated only based on the gene expression profiles (high-throughput mRNA or protein levels). Here, we for the first time applied this rationale to quantitatively measure the impact of REs on the evolution of human molecular pathways with the input data on TFBS distribution.

We found that the gene-specific distributions of the RE-linked TFBS generally correlated with the distributions for all the TFBS. However, several groups of molecular processes were highly enriched in the RE-linked TFBS regulation. They were strongly connected with the immunity and response to pathogens, with the negative regulation of gene transcription, ubiquitination, and protein degradation, extracellular matrix organization, regulation of STAT signaling, fatty acids metabolism, regulation of GTPase activity, protein targeting to Golgi, regulation of cell division and differentiation, and with development and functioning of perception organs and the reproductive system. By contrast, the processes most weakly implicated by the REs were linked mainly with the embryonic development.

We also found that both the gene- and pathway-specific scores significantly correlated for the evolutionary *young* and *all* RE-linked TFBS, thus evidencing that the major evolutionary trends in RE-linked TFBS regulation are largely conserved in the evolution. However, we identified many differences in the regulation features by the younger and older fractions of the REs. The regulation by the older fraction of the REs was linked mainly with the immunity, cell adhesion, Notch, Wnt, and integrin signaling, neuronal development and sensing, chondroitin sulfate and heparin metabolism, cAMP metabolism, endocytosis, and IGF1R signaling.

By contrast, the younger REs were regulating the other aspects of immunity, cell cycle progression and apoptosis attenuation, PDGF, TGF beta, EGFR, and p38 signaling, histone deacetylation and DNA methylation interplay, structure of nuclear lumen, metabolism (primarily catabolism) of phospholipids and

heterocyclic nitrogen-containing molecules, insulin and AMPK signaling, retrograde Golgi-ER transport, estrogen signaling, and oocyte maturation. The immunity-linked pathways were highly represented in both categories (recently and long-term evolving), but their functional characteristics were different and did not overlap. Our results shed light on the evolution of regulatory network in humans and point to the most quickly evolving molecular pathways in higher primates.

RESULTS

Mapping of RE-Specific Human TFBS

From the ENCODE database, we extracted TFBS information for the human myelogenous leukemia cell line K562. The TFBS data for different transcription factor proteins were based on the sequencing of immunoprecipitated DNA fragments (43, 44). The cell line K562 was chosen because it was assayed for the maximum number of transcription factor proteins (225 versus only 120 in the next by abundance cell line GM12878). The TFBSs for all available transcription factor proteins were then mapped onto genomic sequences of the human REs. To identify a fraction of TFBS most likely involved in the regulation of gene expression, we took the elements located in the 5-kb neighborhood of the transcription start sites of known protein-coding genes. A total of 13,029,963 TFBS reads were identified close to transcription start sites. Among them, 2,232,273 (~17%) overlapped with the RE sequences and were referred as the RE-specific fraction of TFBS. Among them, ~44% were attributed to SINEs; ~33%—to LINEs, and 23%—to LR/ERVs. Most of the REs hosting TFBS were highly diverged repeats. For the evolutionary younger REs (0–8% diverged from their consensus sequence), we identified 154,275 TFBS (~7% of all RE-specific TFBS). Among them, SINEs hosted ~68%, LINEs ~15%, and LR/ERVs ~17% of the RE-specific TFBS (**Figure 1B**). The analogous distribution of RE-linked TFBS outside the gene promoter neighborhoods (the rest of the genome) is shown on **Figure 1A**. Interestingly, our data strongly suggest that there is a bias against active TFBS within the evolutionary young LINE elements located close to the gene promoters (**Figure 1B**).

For the same 5-kb neighborhood, we next calculated relative concentration of RE-linked TFBS per kilobase for different RE classes (**Table 1**). It may be seen that for the *young* elements, their ability to provide functional TFBS is generally ~14 times lower than for the group of *all* REs. For the LR/ERVs this factor is also ~14-fold lower, whereas for the LINEs and SINEs ~32- and 9-fold lower, respectively. The extent of this suppression was different for the different types of REs varying from ~9-fold for SINEs till ~32-fold for LINEs, with the median level for LR/ERVs (**Table 1**). The absolute concentrations for the REs were also different, varying from ~0.1 for LR/ERVs and LINEs till 0.4 for SINEs.

Identification of Human Genes Impacted by RE-Linked TFBS

For every individual gene, we calculated its enrichment score for the RE-linked TFBS. We introduced the value termed *Gene RE-linked TFBS enrichment score (TES)* or *GRE score* (**Figure 2**).

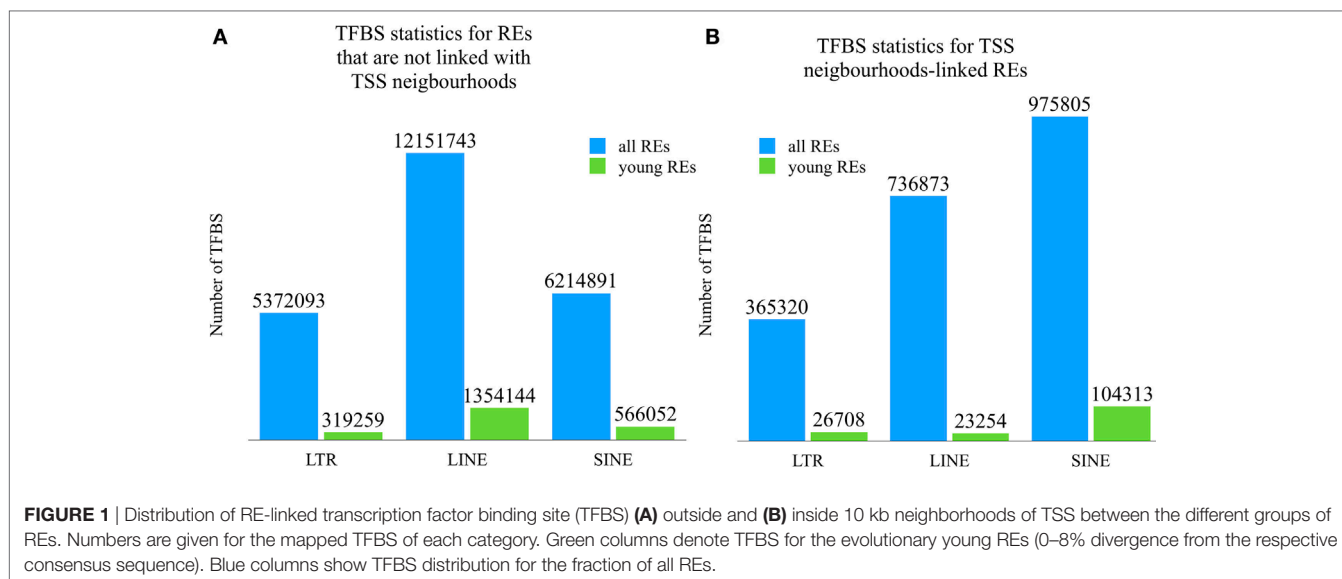


TABLE 1 | Relative concentration of RE-linked transcription factor binding site (TFBS) in 5-kb neighborhood of human gene transcription start sites.

Concentration of TFBS per kilobase

Class of RE	All	Young	Fold change
LINE	2.939	0.093	−31.6
SINE	3.892	0.416	−9.4
LR/ERV	1.457	0.107	−13.6
Total REs	8.287	0.616	−13.5

GRE is the sum of RE-specific TFBS reads mapped close to the individual gene's transcriptional start site, normalized on the average sum of RE-specific TFBS reads for all genes. For every individual gene, GRE score is calculated according to the formula:

$$GRE_g = \frac{TES_g}{\frac{1}{n} \sum_{i=1}^n TES_i},$$

where GRE_g is the GRE score for a gene g ; TES_g is the number of RE-linked TFBS reads for a gene g ; i is gene index and TES_i is the number of RE-linked TFBS reads for a gene i ; and n is the total number of genes.

For every gene, the GRE score makes it possible to measure the extent of enrichment by the RE-linked regulatory elements. For example, $GRE = 1$ means average impact on the regulation of a gene. $GRE > 1$ means that the individual gene is enriched in RE-specific TFBS. Contrarily, $GRE < 1$ means that the gene has lower than average number of RE-specific TFBS.

Our results suggest that there is a fraction of human genes highly enriched in the content of RE-specific TFBS in the regulatory regions, which is reflected by high GRE scores of up to 5 for the protein-coding genes (Table S1 in Supplementary Material; Figure 2). By contrast, many other genes had close to 0 GRE values (Figure 2).

While GRE provides an integral assessment of TFBS impact belonging to all 225 TFs studied here, we also elucidated how strongly each specific TF affects expression of each specific

human gene *via* gene-linked REs. For each gene, i and TF j an entry with indices (i, j) is number of RE-linked TFBS of this TF in the neighborhood of this gene. Our results suggest that most human genes are affected by RE-linked TFBS of various different TFs (Table S2 in Supplementary Material).

Identification of Molecular Pathways Impacted by RE-Linked TFBS

To assess the impact of RE-linked TFBS on the regulation of molecular pathways, we introduced a quantitative metric termed *Pathway Involvement Index (PII)* that is calculated according to the following formula:

$$PII_p = \frac{\sum_{i=1}^n GRE_i}{n},$$

where PII_p is the PII score for a pathway p ; GRE_i is the GRE score for a gene i ; and n is the number of genes in a pathway p . To avoid misleading higher PII values for bigger pathways, PII_p value is normalized on the number of genes in a pathway.

For information about gene products forming molecular pathways, we used the databases BioCarta, KEGG, NCI, Reactome, and Pathway Central. For our profiling, we used 1,749 molecular pathways covering ~11,000 human genes.

The biggest PII scores suggested the highest impact of RE-linked TFBS on the regulation of the whole molecular pathway, and *vice versa*. Zero PII score means no impact on the regulation of the molecular pathway. Similarly to the figure observed for the individual genes, the distribution of PII scores suggests that many molecular pathways are enriched in the regulatory motifs contributed by the REs. We next attempted to characterize the most strongly impacted individual genes and molecular pathways.

Genes Impacted by the RE-Linked TFBS Regulation

The human genes were sorted according to their GRE scores. For different genes, they varied from 0 to 16.4 (Figure 2B; Table S1

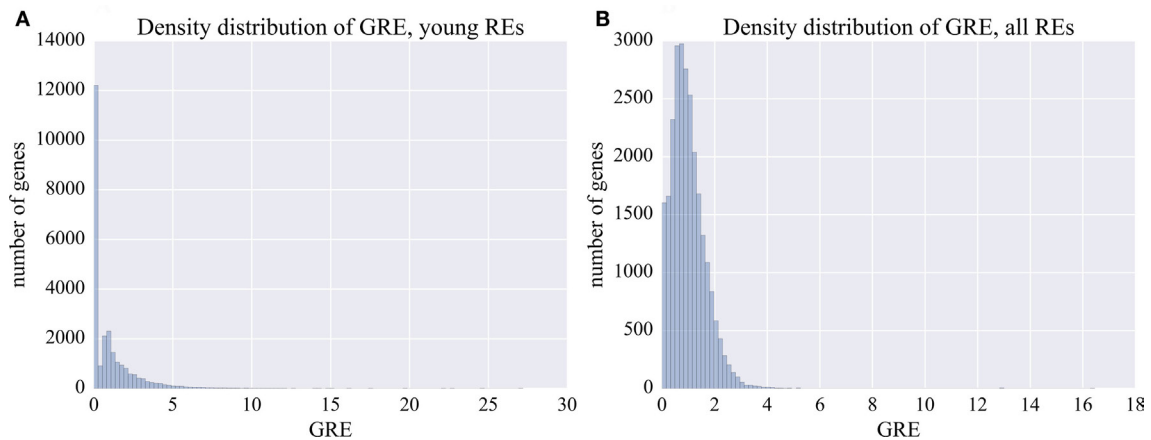


FIGURE 2 | Distribution of GRE score among the known human genes. **(A)** Distribution of GRE for the young fraction of REs (0–8% divergence from the respective consensus sequence). **(B)** Distribution of GRE for the total fraction of all REs.

in Supplementary Material). The top and the bottom 6% of the genes with the highest and the lowest GRE scores profiled for all REs were next analyzed using GO annotation terms and DAVID software.

Top Genes

For the top 6% genes, we identified 48 significantly enriched annotation clusters (Table S3 in Supplementary Material). Among them, 8 (17%) were connected with ribosome biogenesis and translation, 7 (15%) with protein complex assemblies, 5 (10%) with chromatin organization and maintaining structure of the nucleus, 5 (10%) with cell stress and innate immune response mechanisms, 3 (6%) with microtubules and organization of mitotic spindle, 3 (6%) with the regulation of programmed cell death, 3 (6%) with oxidoreductase activity involving purine nucleosides, 2 (4%) with DNA replication and repair, 2 (4%) with formation of mitochondrial outer membrane complexes, and 2 (4%) with the regulation of autophagy. One cluster represented p53-regulated signal transduction, another one—maintaining nucleolus structure. Other features were also presented by minor number of clusters.

Bottom Genes

For the least RE-impacted genes with close to zero RE scores (bottom 6%), quite distinct set of 96 annotation clusters was observed (Table S3 in Supplementary Material). Among them, notably high proportion was taken by 80 (83%) clusters directly linked with embryonic development. Among the others, 8% represented different transcription factor binding assemblies, 2% neuronal axon development, 2% cell–cell adhesion, and signaling, and 2% positive regulation of cell proliferation.

Molecular Pathways Impacted by the RE-Linked TFBS Regulation

We next ranked the molecular pathways by their enrichment with the RE-linked TFBS. For the analysis, we took the

molecular pathways each including at least 10 gene products. The pathways were ranked according to their PII scores (Table S4 in Supplementary Material). We analyzed 65 top (pathways with the highest PII score) and 65 bottom (pathways with the lowest PII score) molecular pathways.

Top Pathways

The following groups of top molecular pathways were identified: 15 (24%) pathways linked with DNA replication and repair, 19% for ribosome and translation, 11% for cytoskeleton remodeling and cell migration, 10% for nuclear transport of mRNA, 10% for other types of nuclear trafficking, 6% for cell stress and innate immune response, 6% for cellular export machinery and vesicle trafficking, 3% for regulation of microtubules and mitotic spindle assembly, 3% for mRNA decay mechanisms, and 8% for the other processes (Table S5 in Supplementary Material).

The major featured molecular processes dealt with protein translation, cell stress and innate immune response, cytoskeleton remodeling, and DNA replication and repair.

Bottom Pathways

The following groups of molecular pathways had the lowest PII scores (Table S5 in Supplementary Material): 18 (30%) for extra-cellular matrix and cell migration, 16% for interleukin-related cell signaling, 21% for neurogenesis, 15% for embryogenesis and morphogenesis, 3% for PTEN signaling, 3% related to G protein coupled receptors (GPCR) signaling, 3% for fatty acids metabolism, and 9% for the other processes.

Comparison of RE-Linked and Non-RE-Linked TFBS Profiles

However, it appeared unclear whether the genes/pathways were enriched in RE-linked TFBS congruently with the overall (not RE-specific) TFBS distribution. To characterize total TFBS distribution trends, we introduced a relative value termed GTE

(Gene TFBS Enrichment). GTE is expressed by the following formula:

$$\text{GTE}_g = \frac{\text{TTS}_g}{\text{TTS}_m},$$

where TTS_g is total number of TFBS reads mapped in the 5-kb neighborhood of a gene g and TTS_m is the mean TTS for all genes under investigation. To define RE-specific enrichment in the regulation of an individual gene, a relative value termed NGRE was introduced:

$$\text{NGRE}_g = \text{GRE}_g / \text{GTE}_g.$$

Bigger NGRE value means bigger impact of RE-specific regulation of certain gene, and *vice versa*.

Another set of values was introduced to estimate the relative RE-specific impact in the regulation of molecular pathways. We added a metric termed PGI (Pathway Gene-based TFBS Index) to assess the impact of total TFBS on the regulation of molecular pathways:

$$\text{PGI}_p = \frac{\sum_{i=1}^n \text{GTE}_i}{n},$$

where PGI_p is the PGI score for a pathway p ; GTE_i is the GTE score for a gene i ; and n is the number of genes in a pathway p .

In turn, the normalized PII called NPPII determines enrichment in RE-specific TFBS regulation of a molecular pathway:

$$\text{NPPII}_p = \text{PII}_p / \text{PGI}_p.$$

where PII_p is a Pathway RE-based Involvement Index for a pathway p and PGI_p is the Pathway Gene-based TFBS Index for a pathway p .

At the level of individual genes, we observed statistically significant correlations between the GRE (based on RE-linked TFBS) and GTE (based on all TFBS) scores (**Figure 3**, Pearson correlation coefficient = 0.47, p -value < 0.001; Table S6 in Supplementary Material). The respective lists of top and bottom GO annotation terms were also highly interconnected featuring protein translation, chromatin remodeling and DNA replication as the most strongly regulated processes, whereas neurogenesis, GPCR signaling, and developmental programs were the most weakly regulated aspects. Taken together, these data evidence that the abundance of RE-linked TFBS roughly (correlation = 0.47) follows overall trend of all TFBS accumulation near gene promoter regions.

Genes and Pathways under Strong Regulation by the REs

To assess the specific trends in RE-dependent regulation of gene expression, we analyzed distributions of the NGRE scores, which characterize the impact of RE-specific TFBS normalized on the regulation by all TFBS for the individual genes (Table S7 in Supplementary Material). The most strongly specifically regulated protein-coding genes were *USP176L26*, *USP17L13*, and *USP17L12* for ubiquitin-specific peptidases. We next analyzed the lists of 6% top and bottom genes sorted according to NGRE

(**Table 2**). The top GO features were linked with immunity and response to pathogens (64/295 terms, or 32%), 7% for organ development, 6% for negative regulation of gene transcription, 6% for chromatin assembly, 6% for protein targeting to Golgi, 4% for ubiquitination and protein degradation, 4% for extracellular matrix organization, 4% for regulation of STAT signaling, 4% for perception organ development and functioning, 4% for negative regulation of macromolecule metabolism, 3% for peptide modifications, 3% for regulation of GTPase activity, 3% for reproductive systems development and functioning, 3% for negative regulation of cell differentiation and positive regulation of cell division, 2% for regulation of body fluids, and 9% was for the other processes (Table S8 in Supplementary Material).

For the group of the bottom 6% of genes, the least regulated features were linked to embryonic development and stem cell differentiation (44/98, or 45%), 16% for transcription and processing of RNA, 16% for nuclear chromatin organization, 8% for ribosome functioning and protein translation, 2% for regulation of apoptosis, 2% for ubiquitin binding, 2% for steroid receptor signaling, 2% for regulation of cell proliferation, and 7% for the other activities (**Table 3**; Table S8 in Supplementary Material).

Similar tendencies were seen at the level of molecular pathways (**Table 3**; Table S9 in Supplementary Material). NPPII scores were calculated that reflect the RE-specific impact on the regulation of molecular pathways normalized to the impact by all TFBS. The top 65 pathways sorted according to NPPII score were linked with fatty acids metabolism (19%), immunity and pathogen recognition (15%), nuclear transport (9%), maturation of mRNA (6%), DNA repair and replication (6%), synuclein A signaling (5%), small RNA biogenesis and function (3%), protein ubiquitination (3%), protein trafficking to Golgi (3%), and other pathways.

The major bottom pathways were involved in the regulation of nerve growth and neuronal signaling (24%), cell adhesion (19%), cytokine networks (14%), other developmental programs (14%), IGF signaling, and regulation of glucose metabolism (9%) (**Table 3**; Table S9 in Supplementary Material).

We next compared the NGRE score distribution at the gene level and NPPII score distribution at the pathway level for the fractions of *all* REs and evolutionary younger REs (*young*; 0–8% diverged from their consensus sequence).

In general, NPPII and NGRE scores were statistically significantly correlated for the *young* and *all* REs, but the pathway-linked NPPII scores showed bigger correlation (**Figure 4A**, Pearson correlation coefficient = 0.38, p -value < 0.001; **Figure 4B**, Pearson correlation coefficient = 0.57, p -value < 0.001). These data are congruent with the previous findings that the data aggregation at the level of molecular pathways frequently provides more congruent results compared with the single-gene level of analysis (41), especially in the case of cancer (45) and neurodegenerative diseases (46, 47).

Although there was a 0.38–0.57 correlation (**Figure 4**), some regulatory features were different between the *young* and *all* REs. To analyze the differences in pathway regulation by *all* and *young* REs, we calculated *ratio* of *all* and *young* REs separately for the NGRE and the NPPII scores. Bigger values here mean greater regulation changes in a long-term rather than recent evolution; lower values mean greater changes in the recent evolution (**Table 4**; Tables S10 and S11 in Supplementary Material for *all/young*

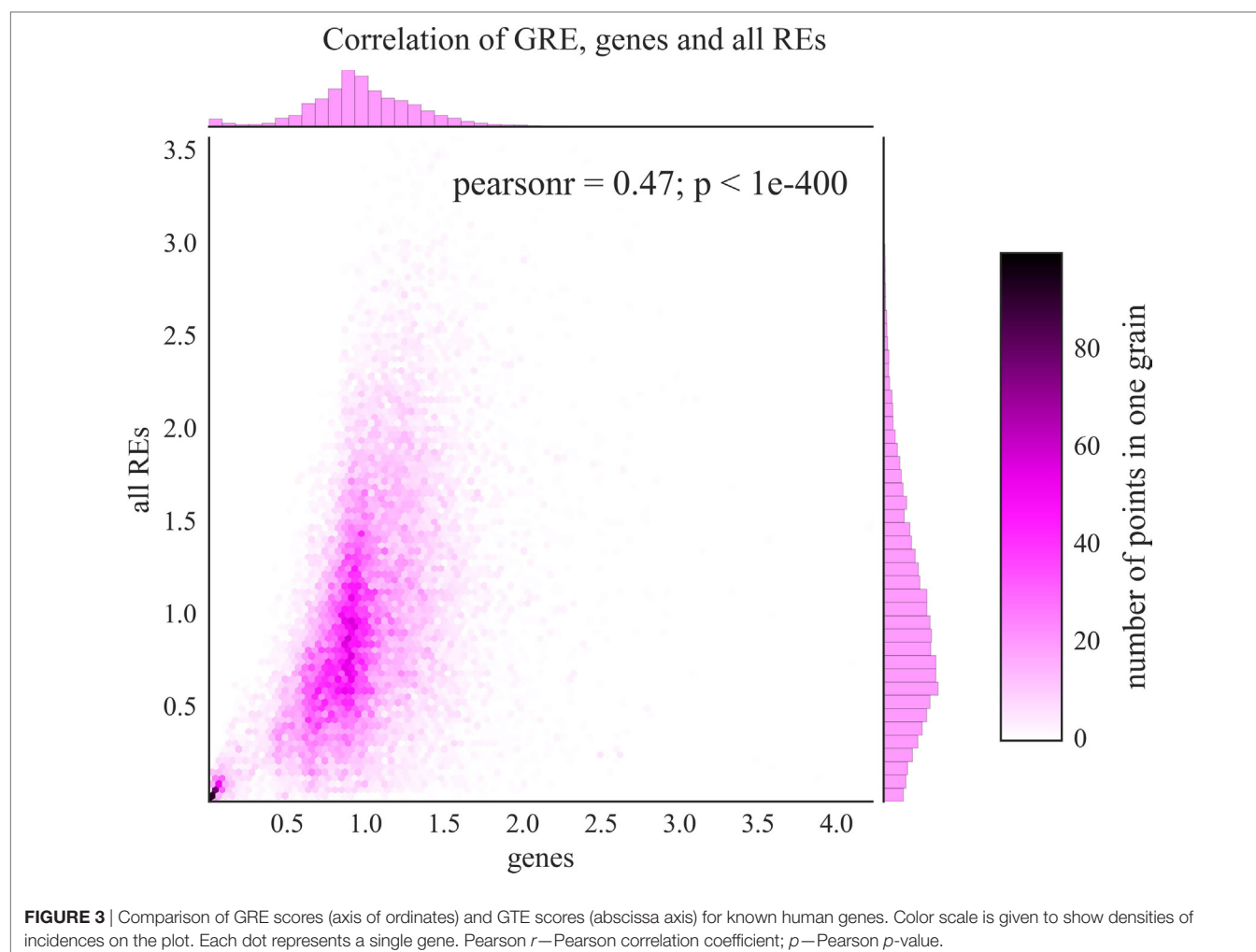


TABLE 2 | Gene ontology (GO) functional annotation clusters in top and bottom 6% of human genes sorted by their NGRE scores.

Cluster, GO terms	Percentage of clusters	
	Top 6%	Bottom 6%
Immunity and response to pathogens	32	—
Organ development and embryogenesis	7	45
Gene transcription and negative regulation	6	—
Chromatin assembly	6	16
Protein targeting to Golgi	6	—
Ubiquitination and protein degradation	4	—
Extracellular matrix organization	4	—
Regulation of STAT signaling	4	—
Perception organ development	4	—
Negative regulation of metabolism	4	—
Peptide modifications	3	—
Regulation of GTPases	3	—
Reproductive system development	3	—
Regulation of differentiation and cell proliferation	3	2
Regulation of body fluids	2	—
Transcription and processing of RNA	—	16
Ribosome assembly and protein translation regulation	—	8
Regulation of apoptosis	—	2
Proteasomal degradation	—	2
Steroid receptor signaling	—	2

TABLE 3 | Functional groups of top and bottom molecular pathways sorted by their NP11 scores.

Functional group	Percentage	
	Top 6%	Bottom 6%
Fatty acids metabolism	19	—
Immunity and response to pathogens	15	14
Nuclear transport	9	—
Maturation of RNA (mRNA and small RNAs)	9	—
DNA repair and replication	6	—
Alpha-synuclein signaling	5	—
Ubiquitination and protein degradation	3	—
Protein targeting to Golgi	3	—
Nerve growth and neuronal signaling	—	24
Organ development, embryogenesis, and cell adhesion	—	35
IGF1R signaling and regulation of glucose metabolism	—	9

ratio of NGRE and NP11, respectively). In the long-term (but not short-term) perspective, the *top* 65 pathways sorted according to NP11 ratio were dealing mainly with cell adhesion, Notch, Wnt, and integrin signaling (20%), immunity and cytokine signaling (20%), neuronal development and sensing (17%), chondroitin

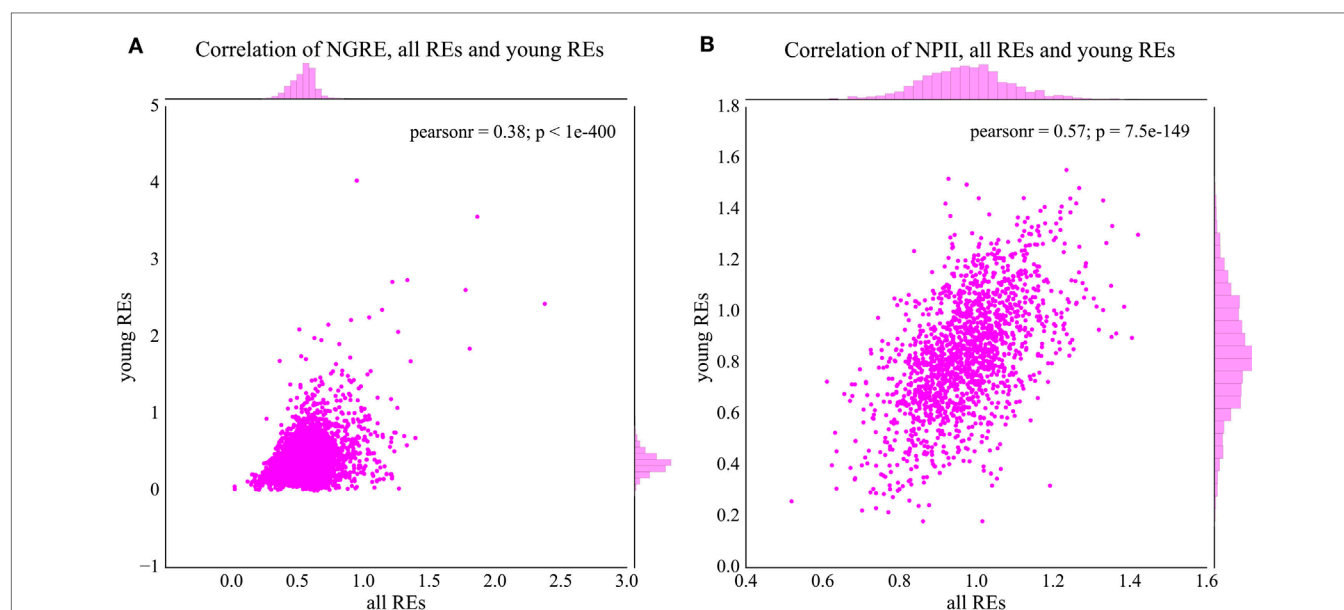


FIGURE 4 | Comparison of normalized transcription factor binding site distributions between the young (0–8% divergence from the respective consensus sequence) and total fractions of REs. **(A)** Comparison of NGRE scores (gene level of regulation), each dot represents a single gene. **(B)** Comparison of NPIL scores (pathway level of regulation), each dot represents a single molecular pathway. Pearson r —Pearson correlation coefficient; p —Pearson p -value.

TABLE 4 | Functional groups of top and bottom molecular pathways sorted by the ratios of NPIL scores for the *all* and *young* RE-linked transcription factor binding sites.

Functional group	Percentage	
	Top 6%	Bottom 6%
Cell adhesion, Notch, Wnt, and integrin signaling	20	–
Immunity and response to pathogens	20	17
Nerve growth and neuronal signaling	17	–
Metabolism of chondroitin sulfate and heparin	8	–
Metabolism of cAMP	6	–
Endocytosis	3	–
IGF1R signaling and regulation of glucose metabolism	3	–
Cell cycle progression and regulation of apoptosis	–	21
PDGF, TGF beta, EGFR, and p38 signaling	–	12
Histone deacetylation and DNA methylation	–	10
Phospholipid metabolism	–	9
Insulin and AMPK signaling	–	6
Protein targeting to Golgi	–	3
Estrogen signaling and oocyte maturation	–	3

sulfate and heparin metabolism (8%), cAMP metabolism (6%), endocytosis (3%), and IGF1R signaling (3%).

The *lower scoring* pathways (most quickly evolving in the recent evolution) were linked mainly with the general cell cycle progression and apoptosis attenuation mechanisms (21%), immunity (17%), PDGF, TGF beta, EGFR, and p38 signaling (12%), histone deacetylation and DNA methylation interplay (10%), phospholipid metabolism (9%), insulin and AMPK signaling (6%), retrograde Golgi-ER transport (3%), and estrogen signaling and oocyte maturation (3%).

Sorting according to NGRE ratio had no sense for the *top* individual genes because there were too many 0 values on the denominator for the NGRE scores calculated for the *young* REs. However, the list of the *bottom* 6% of genes was successfully

generated presumably including the most quickly evolving genes in the recent human evolution (according to RE-linked TFBS acquisition). These genes were mostly involved in the catabolism and synthesis of heterocyclic nitrogen-containing molecules and phospholipids metabolism (50/163, or 31%), nuclear lumen structure (8%), mRNA splicing and processing (7%), ribosome assembly and translation (7%), DNA and histone methylation (4%), and DNA repair (2%).

MATERIALS AND METHODS

Identification of RE-Specific TFBSs

Complete genome binding profiles of 225 transcription factor proteins were extracted from the ENCODE database² for human cell line K562 according to the standard ENCODE ChIP-seq protocol (43). The reference human genome assembly 2009 (hg19) was indexed using Burrows–Wheeler algorithm using BWA software.³ Concatenation of fastq files with single-end or pairwise reads, alignment to the reference genome, and filtering were done using BWA, Samtools, Picard, Bedtools, and Phantompeakqualtools software.⁴ The aligned TFBSs were mapped on the RE sequences annotated by RepeatMasker⁵ and downloaded from the USCS Browser⁶ (RepeatMasker table). TFBS occurrence data were extracted from the bedGraph files.⁷ The fold change over control profiles for TFBS, as well as the profiles for p -value to reject the

²https://www.encodeproject.org/chip-seq/transcription_factor/.

³<https://www.encodeproject.org/pipelines/ENCPL220NBH/>.

⁴<https://www.encodeproject.org>.

⁵<http://www.repeatmasker.org>.

⁶<https://genome.ucsc.edu/cgi-bin/hgTables>.

⁷<https://genome.ucsc.edu/goldenpath/help/bedgraph.html>.

null hypothesis that the signal at that location, is present in the control were built using Macs software⁸ based on the alignment data. The list of transcription factors investigated and the raw data obtained from the ENCODE web site are shown on the Tables S12 and S13 in Supplementary Material.

For every individual mapped RE, we calculated the TES according to the formula:

$$TES = \sum_{i=1}^{225} b_i,$$

where b_i is the number of TFBS reads for transcription factor i mapped on the individual RE.

Measuring Gene Enrichment by the RE-Linked TFBS

The coordinates of human protein-coding genes were downloaded from the USCS Browser.⁹ For each gene, all individual REs overlapping with the 5-kb long neighborhood of its reference transcription start site were selected for further analysis. The 5-kb neighborhood covered an interval starting 5 kb upstream and ending 5 kb downstream the transcription start site. The selected REs were classified according to their structure (HERV/LR, LINE, and SINE) and divergence from the consensus sequence for the respective RE family. The REs with the divergence less than 8% were considered “young” elements. We introduced an integral enrichment score to calculate the RE-linked TFBS enrichment specific to every individual gene (GRE score):

$$GRE_g = \frac{TES_g}{\frac{1}{n} \sum_{i=1}^n TES_i},$$

where GRE_g is the GRE score for a gene g ; TES_g is the sum of TES values for all the RE types for the REs located in the 5-kb neighborhood of a gene g ; n is the number of all genes; and i is gene index and TES_i is the sum of TES values for all the RE types for the REs located in the 5-kb neighborhood of a gene i . Alternatively, specific GRE values can be calculated for every specific type of the REs, when only the TFBS related to the respective RE type are taken into account, e.g., $GRE_{LR/ERV}$, GRE_{LINE} , and GRE_{SINE} .

To separately assess RE-linked TFBS for each of 225 different TF, we created a table for all human genes and all 225 TFs studied here (Table S2 in Supplementary Material). For each gene, i and TF j an entry with indices (i, j) is number of RE-linked TFBS of this TF in the neighborhood of this gene.

For every individual gene g , analogous value termed GTE (Gene TFBS Enrichment) was calculated according to the following formula:

$$GTE_g = \frac{TTS_g}{\frac{1}{n} \sum_{i=1}^n TTS_i},$$

where TTS_g is total number of TFBS reads mapped in the 5-kb neighborhood of a gene g ; n is the number of all genes; i is gene index and TTS_i is the sum of TFBS reads mapped in the 5-kb neighborhood of a gene i .

Alternatively, to assess the relative enrichment in RE-linked TFBS for a certain gene compared with the total number of TFBS for the same gene, the normalized value termed NGRE was introduced:

$$NGRE_g = GRE_g / GTE_g.$$

Measuring Pathway Enrichment by the RE-Linked TFBS

The gene structures of the human molecular pathways were extracted from the following databases: BioCarta,¹⁰ KEGG,¹¹ NCI,¹² Reactome,¹³ and Pathway Central.¹⁴ For each pathway, the PII was calculated according to the formula:

$$PII_p = \frac{\sum_{i=1}^n GRE_i}{n},$$

where PII_p is the PII score for a pathway p ; GRE_i is the GRE score for a gene i ; and n is the number of genes in a pathway p . PII_p value is normalized on the number of genes in a pathway to avoid artificially higher values for larger pathways.

PGI (Pathway Gene-based TFBS involvement Index) is expressed by the formula:

$$PGI_p = \frac{\sum_{i=1}^n GTE_i}{n},$$

where PGI_p is the PGI score for a pathway p , GTE_i is the GTE score for a gene i , and n is the number of genes in a pathway p .

The normalized enrichment in RE-linked TFBS for regulation of a certain molecular pathway termed NPII was calculated as follows:

$$NPII_g = PII_g / PGI_g.$$

GO Enrichment Analysis

Gene ontology analysis of the top and the bottom 6% of the genes by GRE scores profiled for all REs was performed using DAVID software.¹⁵ The p -values specifying the significance of observed GO terms and Annotation Clusters enrichment were calculated using a modified Fisher's exact test (38). The cutoff for p -values was set to be equal to 0.05. The enrichment values of GO terms and Annotation Clusters were calculated as fold changes of their occurrence in the sample versus their occurrence in the human genome (38).

Testing the Significance of the Observed Correlations

The statistical significance of correlations was computed as Pearson correlation coefficient with p -value using the Seaborn package.¹⁶

⁸<https://www.encodeproject.org/pipelines/ENCPL138KID/>.

⁹<https://genome.ucsc.edu/cgi-bin/hgTables>.

¹⁰https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways.

¹¹<http://www.genome.jp/kegg/>.

¹²<https://cactus.nci.nih.gov/ncicadd/about.html>.

¹³<http://reactome.org>.

¹⁴<http://www.sabiosciences.com/pathwaycentral.php>.

¹⁵<https://david.ncifcrf.gov/>.

¹⁶<http://seaborn.pydata.org/>.

DISCUSSION

Our data strongly evidence that the evolutionary changes in transcriptional regulation of gene expression by REs are tightly associated with the gene functions. From the ENCODE database, we extracted TFBS information for the human leukemia cell line K562. For our analysis, we took the TFBS located in the 5-kb neighborhood of the transcription start sites of known protein-coding genes (**Figure 5**). Approximately 13 millions TFBS reads were identified meeting these criteria. Among them, ~17% overlapped with the RE sequences and were referred as the RE-specific fraction of TFBS. They were formed by the three major RE classes: ~44% of them were attributed to SINES; ~33%—to LINES, and 23%—to LR/ERVs. Some REs are known to be transpositionally competent in the human genome and theoretically could generate a cell line-specific population of the RE inserts. However, they only form a negligible proportion of the RE content and could only exert a minor influence on an overall figure of RE-linked TFBS.

Most of the REs hosting TFBS were highly diverged repeats, and for the evolutionary younger elements (0–8% diverged from their consensus sequence), we identified only ~7% of all RE-specific TFBS. Among them, SINES covered ~68%, LINES ~15%, and LR/ERVs ~17% of TFBS (**Figure 1**). This suggests that in the recent evolutionary horizon SINES were approximately four times more active than LINES and LR/ERVs in providing functional TFBS. For the same gene neighborhood, the *young* REs provided functional TFBS generally ~14 times less frequently than the group of *all* REs. These data are congruent with the previously published hypothesis that upon insertion into the host DNA, the newly integrated REs are heavily suppressed. This block is held until they accumulate sufficient number of mutations (48). We show here that the extent of this suppression is different for different RE types varying from ~9-fold for SINES till ~32-fold for LINES, with the median level for LR/ERVs. The absolute concentrations for the REs were also different, varying from ~0.1 for LR/ERVs and LINES till 0.4 for SINES.

Moreover, LINES-linked TFBS are more numerous than the SINES-linked ones outside the gene neighborhoods, whereas the reverse situation is observed near the genes (**Figure 1**). Taken together, these data are also supportive toward another hypothesis that the recent genomic inserts of LINES and LR/ERVs are significantly more deleterious for the human genome than for the SINES (49, 50).

We calculated the absolute RE-linked TESs for the individual genes and for the molecular pathways. The most strongly affected genes and pathways were implicated in the major processes such as cell stress and immune response, ribosome biogenesis and translation, chromatin remodeling and DNA replication, and organization of mitotic spindle and cell cycle progression. On the other hand, the most weakly regulated genes and pathways were mostly dealing with the embryonic development and neurogenesis (Tables S3 and S4 in Supplementary Material). We next showed that the distribution of RE-linked TFBS generally followed the same trend as the total distribution of all TFBS (**Figure 3**, Pearson correlation coefficient = 0.47, p -value < 0.001). The respective lists of top and bottom implicated processes were also highly interconnected for RE-linked and all TFBS, featuring most

strongly regulated protein translation, chromatin remodeling and DNA replication versus most weakly regulated embryonic development and neurogenesis (Table S3 in Supplementary Material). It should be noted that TFBS abundance most likely depends on the importance of a given gene/pathway for the cell type under investigation. For example, for the intensively proliferating leukemia K562 cells investigated here, the programs of embryonic development and neurogenesis can be of an especially low priority, in contrast to DNA replication, protein translation and cell cycle progression (top processes). However, the correlations between all TFBS and RE-linked TFBS features were statistically significant yet not very high (**Figure 3**). This means that there are many fields where the RE-mediated TFBS regulation is different from the general TFBS distribution rule.

The processes specifically enriched in RE-linked TFBS regulation may be thought the most quickly evolving because RE-linked TFBS are generally not conservative among the different species, unlike those located on the unique segments of DNA (51–53). We next attempted to identify the RE-specific trends in the regulation of gene expression and pathway activation. To this end, we analyzed the relative values of RE-specific TFBS profiles normalized on all TFBS profiles for the same genes (Table S7 in Supplementary Material). Of note, the most strongly specifically RE-regulated protein-coding genes were three different genes for the ubiquitin-specific peptidases, which underline relatively faster evolution of the enclosing molecular processes. The top RE-regulated features were strongly connected with the immunity and response to pathogens, and also with the negative regulation of gene transcription, protein targeting to Golgi, ubiquitination and protein degradation, extracellular matrix organization, regulation of STAT signaling, development and functioning of perception organs and reproductive system, fatty acids metabolism, regulation of GTPase activity, negative regulation of cell differentiation and positive regulation of cell division, and with regulation of body fluids (**Tables 3 and 4**).

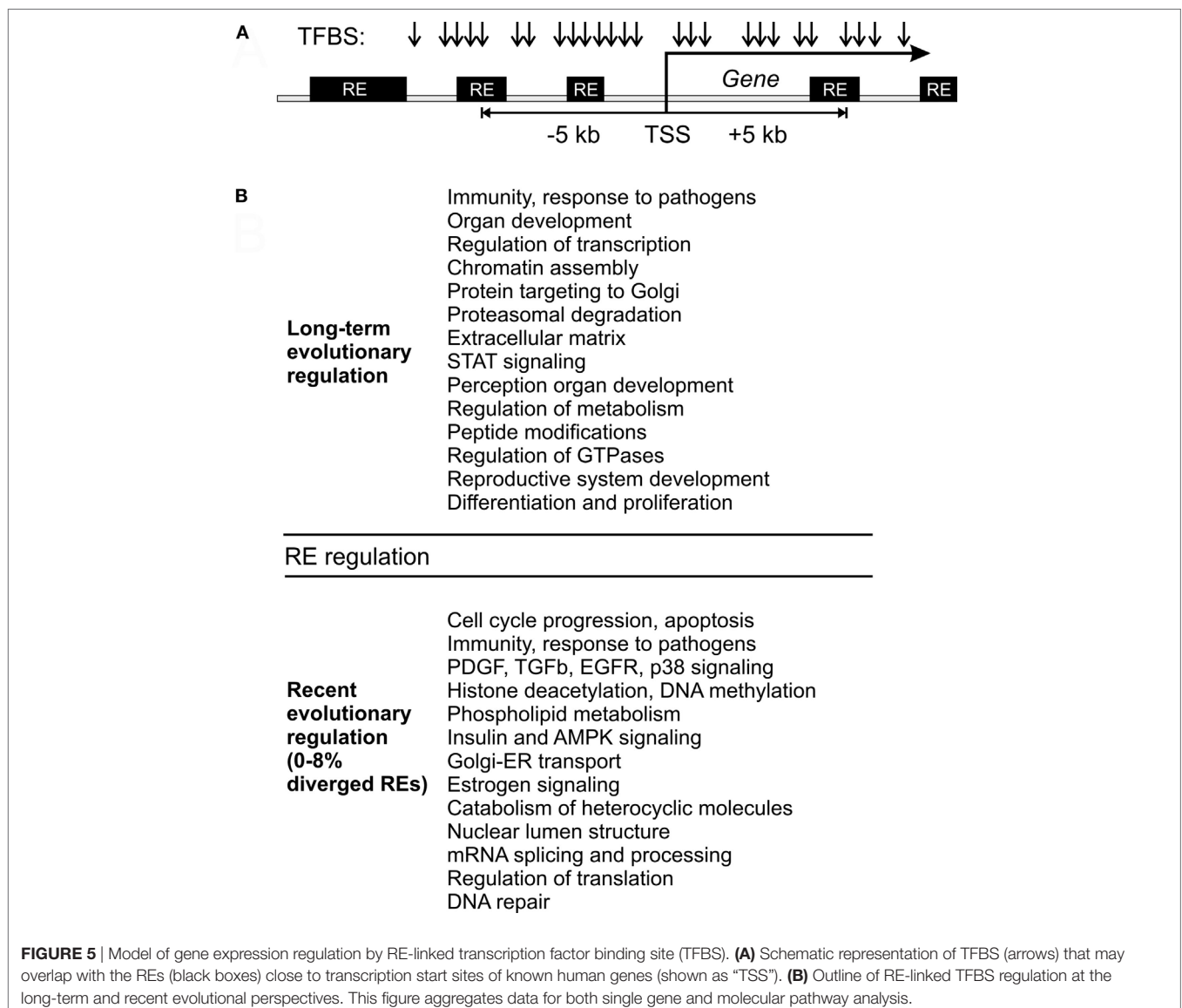
By contrast, the processes most weakly regulated by the REs were linked mostly with the embryonic development, stem cell differentiation, nerve growth and neuronal signaling, cytokine signaling networks, transcription and processing of RNA, nuclear chromatin organization, ribosome assembly and protein translation, IGF1R signaling, and regulation of glucose metabolism (**Tables 3 and 4**).

Moreover, the RE-specific TESs can be calculated for the different fractions of the REs. Here, we analyzed their distributions for the evolutionary *young* fraction of the REs (diverged less than 8%), and for *all* REs. The regulation features in the *all* RE fraction demonstrate long-term tendencies in RE-specific accumulation of TFBS, whereas the *young* fraction may serve as the marker for the relatively recent trends in the human genome evolution, starting roughly since the radiation of Old World monkeys (7, 54). Both gene- and pathway-specific scores statistically significantly correlated for the *young* and *all* RE-linked TFBS (**Figure 4**). This suggests that the major evolutionary trends in RE-linked TFBS regulation are largely conserved. Interestingly, the pathway-specific score was correlated stronger than the gene-specific score (**Figure 4A**, Pearson correlation coefficient = 0.38, p -value < 0.001; **Figure 4B**, Pearson correlation coefficient = 0.57,

p -value < 0.001), which is in line with the previous findings that the data aggregation at the level of molecular pathways provides more stable results and may enhance correlations compared with the single-gene level of analysis (41).

To analyze differences in gene and pathway regulation by *all* and *young* REs, we calculated ratios of the above gene- and pathway-specific scores for *all* and *young* REs. Bigger values here mean greater regulation changes in a long-term rather than recent evolution, lower values—by contrast, greater changes in the recent rather than long-term evolution (**Table 4**; Tables S10 and S11 in Supplementary Material). In the long-term, but not short-term perspective, the top evolving pathways were linked mainly with the immunity and cytokine signaling, cell adhesion, Notch, Wnt, and integrin signaling, neuronal development and sensing, chondroitin sulfate and heparin metabolism, cAMP metabolism, endocytosis, and IGF1R signaling.

By contrast, the most quickly recently evolving processes were linked mainly with the immunity, cell cycle progression and apoptosis attenuation, PDGF, TGF beta, EGFR, and p38 signaling, histone deacetylation and DNA methylation interplay, structure of nuclear lumen, metabolism (primarily catabolism) of phospholipids and heterocyclic nitrogen-containing molecules, insulin and AMPK signaling, retrograde Golgi-ER transport, estrogen signaling, and oocyte maturation (**Figure 5**). The immunity-linked pathways were highly represented in both categories (recently and long-term evolving), but their functional characteristics were different and did not overlap (**Table 5**). These pathways are mostly connected with inflammation, pathogen recognition of innate immunity and cytokine signaling. Our findings concerning the RE impact of the both long-term and short-term evolution of human immune system are in accord with recent experimental findings that HERV have dispersed



numerous IFN-inducible enhancers regulating essential innate immune functions (10, 11).

Patterns of genes mostly impacted by transposons are generally consistent with universals of genome evolution (55). Our findings of RE-impacted changes in human molecular pathways are also generally in line with both ancient and recent trends in the evolution of human lineage. Retrotransposon insertion is an abrupt event that can drastically affect expression of neighboring genes by regulatory innovation and direct mutation (9). A general hypothesis was proposed that genes that are highly expressed in all tissues (mostly cytoplasmic and housekeeping) cannot tolerate regulatory and mutational pressure imposed by transposons without fitness loss (56, 57) because the toxic effects of protein misfolding and stoichiometric imbalance of subunits are thought to be most severe for highly abundant proteins (58). Here, we show that human RE impact mainly the pathways linked with immunity, signal transduction, proliferation, cell interaction and communication both on the recent and the long-term time scales, whereas cytoplasmic and housekeeping molecular pathways are weaker affected.

Moreover, evolutionary history of human lineage most likely includes series of time-periods with the accelerated evolution of some particular molecular systems, i.e., due to evolutionary arms race (59), run-away processes of sexual selection (60), and classical positive selection, e.g., selection for the ability to accept new types of food (61). Interestingly, regulatory innovations were probably the major source of changes throughout the recent human evolution (62). First, evolutionary arms-race between human ancestors and various pathogens has driven the changes of adaptive immune response (63) and is still shaping

human immunity nowadays (64). Here, we show that such shaping is mediated also through RE insertions and exaptation of their TFBS to regulate expression of immunity-linked genes. Interestingly, long-term and short-term evolutionary pressures onto the human immune system sometimes appear disjoined, e.g., because of encountering new pathogens, reflected by the fact that different modules of immunity were affected by REs on different time scales (Table 5). Second, evolution of human brain was largely affected by sexual selection under a trend toward monogamy, lowering male competition, and increasing female choice (65). Our study suggests that REs had been affecting human nerve system for a long time (Table 4) that may accounts for multiple events in the evolution of mammalian brain. Third, recent human evolution after divergence with chimpanzee imposed several dietary transitions, such as increased meat-eating that occurred ~2 mya simultaneously with massive usage of stone and fire (61). Therefore, recent changes in the catabolism of heterocyclic molecules and phospholipid catabolism can be at least partly connected with this kind of food speciation of great apes and hominids. Fourth, rapid recent RE-affected evolution of histone deacetylation and DNA methylation interplay can be at least partly connected with gradual diversification of transposon-repressing KRAB zinc finger TFs (66), reflecting intragenome evolutionary arms race between REs and host genes.

In this study, we analyzed in depth RE-linked TFBS signatures for a unique human cell line where the high-throughput TFBS profile is currently available. Further accumulation of high-throughput data on TFBS distribution will make it possible to build a more robust model of RE influence on human molecular

TABLE 5 | Top and bottom immunity-linked molecular pathways sorted by the ratios of NPII scores for the *all* and *young* RE-linked transcription factor binding site.

Pathway name	All NPII	Young NPII	Ratio (A/Y)
Top ratio pathways			
NCI downstream signaling in naive CD8 T cells pathway (pathway regulation of survival gene product expression <i>via</i> IL2RG)	1.02	0.18	5.59
NCI downstream signaling in naive CD8 T cells main pathway	0.93	0.43	2.16
Reactome toll like receptor 4 TLR4 cascade main pathway	1.19	0.32	3.72
Reactome interleukin receptor SHC signaling main pathway	0.85	0.24	3.51
Cytokine network pathway	1.04	0.32	3.24
NCI CXCR3 mediated signaling events pathway (cell adhesion)	1.06	0.39	2.76
NCI CXCR3 mediated signaling events pathway (actin polymerization or depolymerization)	1.02	0.40	2.57
NCI LPA receptor mediated events pathway (cAMP biosynthetic process)	0.87	0.35	2.46
Biocarta lck and fyn tyrosine kinases in initiation of tcr activation main pathway	0.98	0.42	2.36
NCI IL2 mediated signaling events pathway (T cell proliferation)	1.00	0.46	2.19
NCI BCR signaling pathway (reentry into mitotic cell cycle)	0.88	0.42	2.12
NCI IL4-mediated signaling events main pathway	0.96	0.46	2.08
KEGG inflammatory bowel disease IBD main pathway	1.01	0.49	2.06
Lower ratio pathways			
Reactome IRAK2 mediated activation of TAK1 complex main pathway	1.17	1.40	0.84
Reactome IRAK2 mediated activation of TAK1 complex upon TLR7 8 or 9 stimulation main pathway	1.17	1.40	0.84
KEGG Fanconi anemia main pathway	0.98	1.17	0.83
Reactome Fanconi anemia main pathway	0.89	1.08	0.82
Reactome CD28 dependent Vav1 main pathway	1.13	1.37	0.83
Reactome thromboxane signaling through TP receptor main pathway	0.94	1.15	0.82
NCI Thromboxane A2 receptor signaling pathway (JNK cascade)	0.95	1.20	0.79
NCI Fc epsilon receptor I signaling in mast cells pathway (regulation of mast cell degranulation)	0.82	1.03	0.79
IL-10 pathway IL-10 responsive genes transcription of BCLXL cyclin-D1 D2 D3 Pim1 c-myc and P19 (INK4D) <i>via</i> STAT3	0.99	1.31	0.76
IL-10 pathway inflammatory cytokine genes expression <i>via</i> STAT3	0.99	1.31	0.76
Reactome membrane binding and targeting of GAG proteins main pathway	0.94	1.25	0.75

pathways based on thorough analysis of many objects including various cell lines and hopefully intact and pathological human tissues.

Finally, given that REs make up >40% of genomic sequence and that >80% of the REs are located outside promoter-neighboring regions, it remains of a great interest to further investigate if this larger subset of REs may have significant role in the evolution of human molecular pathways that can be mediated *via* chromatin remodeling or regulation of non-coding RNAs. This will be a matter of further investigation in our consortium.

AUTHOR CONTRIBUTIONS

DN, DP, and AG analyzed transcription factor binding sites (TFBS) data; MS constructed molecular pathways library; VT mapped TFBS and retrotransposons on human genome; NB, AP, VP, and AB wrote and implemented algorithms for data analysis; and AB and DN wrote the paper.

ACKNOWLEDGMENTS

This article would have been impossible without kindly granted access to the computational server of Institute of Parasitology, Biology Centre, Czech Academy of Sciences, Ceske Budejovice.

REFERENCES

- Eickbush TH, Jamburuthugoda VK. The diversity of retrotransposons and the properties of their reverse transcriptases. *Virus Res* (2008) 134:1–2. doi:10.1016/j.virusres.2007.12.010
- International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* (2004) 431:7011. doi:10.1038/nature03001
- Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, et al. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci U S A* (2004) 101:14. doi:10.1073/pnas.030780010
- Buzdin A. Human-specific endogenous retroviruses. *ScientificWorldJournal* (2007) 7:1848–68. doi:10.1100/tsw.2007.270
- Suntsova M, Garazha A, Ivanova A, Kaminsky D, Zhavoronkov A, Buzdin A. Molecular functions of human endogenous retroviruses in health and disease. *Cell Mol Life Sci* (2015) 72:19. doi:10.1007/s00018-015-1947-6
- Gogvadze E, Buzdin A. Retroelements and their impact on genome evolution and functioning. *Cell Mol Life Sci* (2009) 66:23. doi:10.1007/s00018-009-0107-2
- Sverdlov ED. Retroviruses and primate evolution. *Bioessays* (2000) 22:2. doi:10.1002/(SICI)1521-1878(200002)22:2<161::AID-BIES7>3.0.CO;2-X
- De Souza FSJ, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel Cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* (2013) 30:6. doi:10.1093/molbev/mst045
- Göke J, Ng HH. CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. *EMBO Rep* (2016) 17:8. doi:10.15252/embr.201642743
- Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* (2016) 18:2. doi:10.1038/nrg.2016.139
- Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* (2016) 351:1083–7. doi:10.1126/science.aad5497
- Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E. At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. *J Virol* (2006) 80:21. doi:10.1128/JVI.00871-06

FUNDING

This work was supported by the Russian Science Foundation grant no. 14-50-00060.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/articles/10.3389/fimmu.2018.00030/full#supplementary-material>.

TABLE S1 | GRE score for REs of different classes and evolutionary ages.

TABLE S2 | Number of RE-linked transcription factor binding sites for human genes and 225 analysed transcription factors.

TABLE S3 | GO terms for top and bottom 6% of genes by GRE of all REs.

TABLE S4 | PII score for all REs for human molecular pathways.

TABLE S5 | PII score for all REs for top and bottom 65 human molecular pathways.

TABLE S6 | GTE score for human genes.

TABLE S7 | NGRE score for all and young REs.

TABLE S8 | GO terms for top and bottom 6% of genes by NGRE of all REs.

TABLE S9 | Comparison of NPII for all and young REs.

TABLE S10 | Ratio of all NGRE and young NGRE for human genes.

TABLE S11 | Ratio of all NPII and young NPII for human molecular pathways.

TABLE S12 | List of studied transcription factors.

TABLE S13 | List of data files corresponding to studied transcription factors.

- Chuong EB, Rumi MAK, Soares MJ, Baker JC. Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* (2013) 45:3. doi:10.1038/ng.2553
- Suntsova M, Gogvadze EV, Salozhin S, Gaifullin N, Eroshkin F, Dmitriev SE, et al. Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc Natl Acad Sci U S A* (2013) 110:48. doi:10.1073/pnas.1318172110
- Schumann GG, Gogvadze EV, Osanai-Futahashi M, Kuroki A, Münk C, Fujiwara H, et al. Unique functions of repetitive transcriptomes. *Int Rev Cell Mol Biol* (2010) 285:115–88. doi:10.1016/B978-0-12-381047-2.00003-7
- Young JM, Whiddon JL, Yao Z, Kasinathan B, Snider L, Geng LN, et al. DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet* (2013) 9:11. doi:10.1371/journal.pgen.1003947
- Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A* (2007) 104:47. doi:10.1073/pnas.0703637104
- Ho B, Baker PM, Singh S, Shih S-J, Vaughan AT. Localized DNA cleavage secondary to genotoxic exposure adjacent to an Alu inverted repeat. *Genes Chromosomes Cancer* (2012) 51:5. doi:10.1002/gcc.21938
- Garazha A, Ivanova A, Suntsova M, Malakhova G, Roumiantsev S, Zhavoronkov A, et al. New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome. *Cell Cycle* (2015) 14:9. doi:10.1080/15384101.2015.1022696
- Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* (2009) 41:5. doi:10.1038/ng.368
- Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* (2000) 28:1. doi:10.1093/nar/28.1.27
- Demidenko ZN, McCubrey JA. Recent progress in targeting cancer. *Aging (Albany NY)* (2011) 3:12. doi:10.18632/aging.100421
- Elkon R, Vesterman R, Amit N, Ulitsky I, Zohar I, Weisz M, et al. SPIKE – a database, visualization and analysis tool of cellular signaling pathways. *BMC Bioinformatics* (2008) 9:1. doi:10.1186/1471-2105-9-110
- Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics* (2003) 19:16. doi:10.1093/bioinformatics/btg290

25. UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res* (2011) 39:D214–9. doi:10.1093/nar/gkq1020
26. Mathivanan S, Periaswamy B, Gandhi T, Kandasamy K, Suresh S, Mohmood R, et al. An evaluation of human protein-protein interaction data in the public domain. *BMC Bioinformatics* (2006) 7(Suppl 5):S19. doi:10.1186/1471-2105-7-S5-S19
27. Marshall CJ. Opportunities for pharmacological intervention in the ras pathway. *Ann Oncol* (1995) 6(Suppl 1):63–7.
28. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell* (2000) 100:1. doi:10.1016/S0092-8674(00)81683-9
29. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144:5. doi:10.1016/j.cell.2011.02.013
30. Borisov N, Aksamitiene E, Kiyatkin A, Legewie S, Berkhout J, Maiwald T, et al. Systems-level interactions between insulin-EGF networks amplify mitogenic signaling. *Mol Syst Biol* (2009) 5:256. doi:10.1038/msb.2009.19
31. Larance M, Lamond AI. Multidimensional proteomics for cell biology. *Nat Rev Mol Cell Biol* (2015) 16:5. doi:10.1038/nrm3970
32. Echeverría PC, Bernthaler A, Dupuis P, Mayer B, Picard D. An interaction network predicted from public data as a discovery tool: application to the Hsp90 molecular chaperone machine. *PLoS One* (2011) 6:10. doi:10.1371/journal.pone.0026044
33. Zhang QC, Petrey D, Garzón JL, Deng L, Honig B. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res* (2012) 41:D1. doi:10.1093/nar/gks1231
34. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* (2017) 45:D1. doi:10.1093/nar/gkw937
35. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway knowledgebase. *Nucleic Acids Res* (2016) 44:D1. doi:10.1093/nar/gkv1351
36. Ogishima S, Mizuno S, Kikuchi M, Miyashita A, Kuwano R, Tanaka H, et al. A map of Alzheimer's disease-signaling pathways: a hope for drug target discovery. *Clin Pharmacol Ther* (2013) 93:5. doi:10.1038/clpt.2013.37
37. Demidenko ZN, Blagosklonny MV. The purpose of the HIF-1/PHD feedback loop: to limit mTOR-induced HIF-1 α . *Cell Cycle* (2011) 10:10. doi:10.4161/cc.10.10.15789
38. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* (2008) 4:1. doi:10.1038/nprot.2008.211
39. Buzdin AA, Prassolov V, Zhavoronkov AA, Borisov NM. Bioinformatics meets biomedicine: OncoFinder, a quantitative approach for interrogating molecular pathways using gene expression data. *Methods Mol Biol* (2017) 1613:53–83. doi:10.1007/978-1-4939-7027-8_4
40. Aliper AM, Korzinkin MB, Kuzmina NB, Zenin AA, Venkova LS, Smirnov F, et al. Mathematical justification of expression-based pathway activation scoring (PAS). *Methods Mol Biol* (2017) 1613:31. doi:10.1007/978-1-4939-7027-8_3
41. Borisov N, Suntsova M, Sorokin M, Garazha A, Kovalchuk O, Aliper A, et al. Data aggregation at the level of molecular pathways improves stability of experimental transcriptomic and proteomic data. *Cell Cycle* (2017) 16:19. doi:10.1080/15384101.2017.1361068
42. Buzdin AA, Zhavoronkov AA, Korzinkin MB, Venkova LS, Zenin AA, Smirnov PY, et al. Oncofinder, a new method for the analysis of intracellular signaling pathway activation using transcriptomic data. *Front Genet* (2014) 5:55. doi:10.3389/fgene.2014.00055
43. ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SE, Collins PJ, Davis CA, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489(7414):57–74. doi:10.1038/nature11247
44. Sloan CA, Chan ET, Davidson JM, Malladi VS, Strattan JS, Hitz BC, et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res* (2016) 44:D1. doi:10.1093/nar/gkv1160
45. Yeh C-S, Wang J-Y, Cheng T-L, Juan C-H, Wu C-H, Lin S-R. Fatty acid metabolism pathway play an important role in carcinogenesis of human colorectal cancers by microarray-bioinformatics analysis. *Cancer Lett* (2006) 233:2. doi:10.1016/j.canlet.2005.03.050
46. Alberio T, Lopiano L, Fasano M. Cellular models to investigate biochemical pathways in Parkinson's disease. *FEBS J* (2012) 279:7. doi:10.1111/j.1742-4658.2012.08516.x
47. Mattson MP, Pedersen WA, Duan W, Culmese C, Camandola S. Cellular and molecular mechanisms underlying perturbed energy metabolism and neuronal degeneration in Alzheimer's and Parkinson's diseases. *Ann N Y Acad Sci* (1999) 893:1. doi:10.1111/j.1749-6632.1999.tb07824.x
48. Buzdin AA, Prassolov V, Garazha AV. Friends-enemies: endogenous retroviruses are major transcriptional regulators of human DNA. *Front Chem* (2017) 5:35. doi:10.3389/fchem.2017.00035
49. Sultana T, Zamborlini A, Cristofari G, Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet* (2017) 18:5. doi:10.1038/nrg.2017.7
50. Campos-Sánchez R, Kapusta A, Feschotte C, Chiaromonte F, Makova KD. Genomic landscape of human, bat, and ex vivo DNA transposon integrations. *Mol Biol Evol* (2014) 31:7. doi:10.1093/molbev/msu138
51. Cheng Y, Ma Z, Kim B-H, Wu W, Cayting P, Boyle AP, et al. Principles of regulatory information conservation between mouse and human. *Nature* (2014) 515:7527. doi:10.1038/nature13985
52. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* (2014) 24:12. doi:10.1101/gr.168872.113
53. Lowdon RF, Jang HS, Wang T. Evolution of epigenetic regulation in vertebrate genomes. *Trends Genet* (2016) 32:5. doi:10.1016/j.tig.2016.03.001
54. Giordano J, Ge Y, Gelfand Y, Abrusán G, Benson G, Warburton PE. Evolutionary history of mammalian transposons determined by genome-wide defragmentation. *PLoS Comput Biol* (2007) 3:7. doi:10.1371/journal.pcbi.0030137
55. Koonin EV. Are there laws of genome evolution? *PLoS Comput Biol* (2011) 7:8. doi:10.1371/journal.pcbi.1002173
56. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* (2005) 102:40. doi:10.1073/pnas.0504070102
57. Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. *Mol Biol Evol* (2005) 22:5. doi:10.1093/molbev/msi122
58. Drummond DA, Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* (2008) 134:2. doi:10.1016/j.cell.2008.05.042
59. Capewell P, Cooper A, Clucas C, Weir W, Macleod A. A co-evolutionary arms race: trypanosomes shaping the human genome, humans shaping the trypanosome genome. *Parasitology* (2015) 142:S1. doi:10.1017/S0031182014000602
60. Wilkinson GS, Breden F, Mank JE, Ritchie MG, Higginson AD, Radwan J, et al. The locus of sexual selection: moving sexual selection studies into the post-genomics era. *J Evol Biol* (2015) 28:4. doi:10.1111/jeb.12621
61. Ye K, Gu Z. Recent advances in understanding the role of nutrition in human genome evolution. *Adv Nutr* (2011) 3:10. doi:10.3945/an.111.001024.FIGURE
62. Siepel A, Arbizu L. Cis-regulatory elements and human evolution. *Curr Opin Genet Dev* (2014) 29:1. doi:10.1016/j.gde.2014.08.011
63. Carmona LM, Schatz DG. New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J* (2017) 284:11. doi:10.1111/febs.13990
64. Hoffmann H-H, Schneider WM, Rice CM. Interferons and viruses: an evolutionary arms race of molecular interactions. *Trends Immunol* (2015) 36:3. doi:10.1016/j.it.2015.01.004
65. Stanyon R, Bigoni F. Sexual selection and the evolution of behavior, morphology, neuroanatomy and genes in humans and other primates. *Neurosci Biobehav Rev* (2014) 46:579–90. doi:10.1016/j.neubiorev.2014.10.001
66. Imbeault M, Helleboid P-Y, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* (2017) 543:7646. doi:10.1038/nature21683

Conflict of Interest Statement: 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.

Copyright © 2018 Nikitin, Penzar, Garazha, Sorokin, Tkachev, Borisov, Poltorak, Prassolov and Buzdin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Regular Wounding in a Natural System: Bacteria Associated With Reproductive Organs of Bedbugs and Their Quorum Sensing Abilities

Oliver Otti^{1,2*}, Peter Deines^{1,3}, Katrin Hammerschmidt^{1,4} and Klaus Reinhardt^{1,5}

¹Animal and Plant Sciences, University of Sheffield, Sheffield, United Kingdom, ²Animal Population Ecology, Animal Ecology I, University of Bayreuth, Bayreuth, Germany, ³Zoological Institute, Christian Albrechts University Kiel, Kiel, Germany, ⁴Institute of Microbiology, Christian Albrechts University Kiel, Kiel, Germany, ⁵Applied Zoology, Department of Biology, Technische Universität Dresden, Dresden, Germany

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, United States

Reviewed by:

Jing Yan,
Princeton University,
United States
Sampriti Mukherjee,
Princeton University,
United States

*Correspondence:

Oliver Otti
oliver.otti@uni-bayreuth.de

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 10 August 2017

Accepted: 07 December 2017

Published: 18 December 2017

Citation:

Otti O, Deines P, Hammerschmidt K
and Reinhardt K (2017) Regular
Wounding in a Natural System:
Bacteria Associated With
Reproductive Organs of Bedbugs
and Their Quorum Sensing Abilities.
Front. Immunol. 8:1855.
doi: 10.3389/fimmu.2017.01855

During wounding, tissues are disrupted so that bacteria can easily enter the host and trigger a host response. Both the host response and bacterial communication can occur through quorum sensing (QS) and quorum sensing inhibition (QSI). Here, we characterize the effect of wounding on the host-associated bacterial community of the bed bug. This is a model system where the male is wounding the female during every mating. Whereas several aspects of the microbial involvement during wounding have been previously examined, it is not clear to what extent QS and QSI play a role. We find that the microbiome differs depending on mating and feeding status of female bedbugs and is specific to the location of isolation. Most organs of bedbugs harbor bacteria, which are capable of both QS and QSI signaling. By focusing on the prokaryotic quorum communication system, we provide a baseline for future research in this unique system. We advocate the bedbug system as suitable for studying the effects of bacteria on reproduction and for addressing prokaryote and eukaryote communication during wounding.

Keywords: quorum quenching, interspecific communication, reproductive immunity, genitalia-associated microbes, genital infection

INTRODUCTION

All animals live in intimate associations with bacteria, which can live on host surfaces, reside within or between host cells or be associated with specific organ systems (1–3). The entirety of a host and its associated bacterial community (microbiome) is called the metaorganism (4). Studying the effects of microbiomes on host ecology and evolution has become a major line of research (5–8).

Most interactions of the host with bacteria from the environment happen at the host surfaces. One important way by which environmental bacteria might enter the host organism is wounds (9). This is of particular importance in cases where wounding occurs on a regular basis such as during mating (10–12) and when bacteria are transferred to mating partners *via* contaminated reproductive organs. Bacteria are ubiquitously found on male and female genitalia, including insects, birds, or humans (12–16), and copulatory wounding has been shown to be very widespread in the animal kingdom. In many species, males cause micro- and macro-lesions in the female reproductive tract during mating (11) and, even in humans, 10–52% of copulations result in mucous lesions, abrasions, or lacerations of female genital organs [(11) and references therein]. While males may protect their sperm from bacteria, they transfer to females by transferring antimicrobial substances in their

seminal fluid alongside the sperm (17, 18), it remains largely unknown how females prepare for bacterial invasions after copulation (19) and how the bacterial community residing in the female responds to the foreign intruders. For example, in other metaorganisms, the resident microbiota plays a critical role in maintaining host health by interacting with invading microbes (8, 20–22).

The host-associated microbial community is shaped by the host but also through interactions within the microbial community. Bacterial communication systems, such as quorum sensing (QS) and quorum sensing inhibition (QSI), influence the stability of the microbial community, and thus the integrity of the metaorganism (23, 24). However, little is known about how these quorum communication systems work between resident and invading microbes.

Quorum sensing and QSI occur within and between bacterial species (25). Essentially, QS regulates the gene expression to produce and release chemical signal molecules called autoinducers in response to fluctuations in bacterial cell-population density (26). These responses include adaptation to the availability of nutrients or the defense against other microorganisms, which may compete for the same nutrients (or hosts). Bacteria also coordinate their behavior in infections with QS, e.g., many pathogenic bacteria coordinate their virulence to evading the immune response of the host and establishing a successful infection.

Competing bacteria species have evolved mechanisms to interfere with each other's QS communication by quenching the signal molecules, called quorum sensing inhibition (QSI) (25, 27) or by inhibiting each other's growth (28). As expected, hosts have evolved counteradaptations that interfere with the QS process and limit the spread of information among infecting bacteria, or interfere with bacterial growth to prevent the colonization by bacteria, e.g., through temperature and pH increase (29). Although bacterial communication is currently attracting a lot of interest, not much is known about the distribution of bacteria competent to perform QS, QSI, or growth inhibition in natural bacteria-host systems.

A further important player in the host-microbe interaction has recently been identified. Ismail et al. (30) have shown that the damage of eukaryotic host cells, as occurs during wounding, also releases signals that interfere with bacterial QS systems. This provides, yet another, very fast line of defense once bacteria have bypassed the host's epithelia. While future work will doubtlessly bring more such exciting research results and will eventually lead to identifying the relative significance of pro- and eukaryotic quorum communication, we here present a first step into that direction. We present a unique arthropod model of regular copulatory wounding—the natural traumatic insemination of bedbugs—and characterize the prokaryotic side of the quorum communication by investigating the ability to perform QS or QSI *in vitro* of bacteria isolated from male and female reproductive organs.

Briefly, the male bedbug possesses a stylet-like copulatory organ (called the paramere) with which it wounds the female (breaches their integument) during every copulation. On the paramere, environmental bacteria have been found (17, 31), which can be transported into the female (17). An experimental

overabundance of bacteria on the male's paramere dramatically accelerated female death and has selected for the evolution of a novel female immune organ (32). This immune organ, the mesospermalege, is filled with immune cells, hemocytes, of more or less unknown function, which significantly reduces the negative effect of wounding and bacterial infection (32). Females have little control over whether or not they mate other than by feeding—fully fed females cannot resist copulation, non-fed females partially can (33). Therefore, fully fed females can expect to be mated, and in order to characterize the prokaryotic quorum communication in our model metaorganism, it is necessary to separate the effects of feeding from the effects of wounding.

The objectives of the current research are: (1) to isolate and identify the site-specific, culturable microbiome of the bedbug, (2) to test the effect of wounding and feeding on the microbiome of female bedbugs, and (3) to quantify the potential for quorum communication of the culturable bacteria species of the bedbug microbiome *in vitro*. We sample eight different bedbug populations for their bacteria using a culture-dependent method. We separately screen the bacteria from the bedbug environment, the cuticle of males and females, the male paramere, the female hemolymph, and mesospermalege. We contrast virgin and mated females to isolate the effect of wounding and feeding on the microbiome and conduct QS and QSI assays to establish the competence of the isolated bacterial lineages for QS, QSI, and growth inhibition. The microbiome varied between mated and non-mated individuals, between fed and non-fed ones as well as between organs. Most of the screened reproductive organs harbored bacteria capable of both signaling pathways, QS and QSI. Our findings provide a baseline for future research in bedbugs and promote it as a system for studying the effects of bacteria on reproduction and prokaryote-eukaryote communication during wounding in a natural system.

MATERIALS AND METHODS

Study Animals

All bedbugs were derived from one large stock population (>1,000 individuals) maintained at the University of Sheffield for more than 6 years (34). We conducted two experiments to (i) obtain all culturable bacteria from specific sites in the female and male bedbug and from the bedbug environment and (ii) disentangle the effect of feeding and mating on bacterial diversity in different sites in the female bedbug. For the bacteria in the bedbug environment, we sampled eight different stocks, six field-caught (five from the UK and one from Kenya) and two long-term lab stocks originally obtained from the London School of Hygiene and Tropical Medicine and from Bayer Environmental Science (Monheim, Germany).

Site-Specific, Culturable Microbiome of the Bedbug

To isolate and identify most culturable bacteria from the bedbug microbiome, six different growth media [sterile Grace's Insect medium (GM; G8142, Sigma-Aldrich, Dorset, UK), NB, NBTA, LB, Potato extract, and R2A] were used to prepare 1.5% agar plates. To assess the diversity of bacteria in the female hemolymph and mesospermalege, we analyzed 3-week old females ($N = 17$),

which were either fully mated with randomly picked males from the same stock population ($N = 12$) or left virgin as control ($N = 5$). We allowed the males to copulate for as long as they wanted and removed them immediately after they let go of the female. Thirty minutes after mating, hemolymph samples were taken from all females after which they were dissected to remove the mesospermae. Additionally, we sampled the microbial diversity on the parameres from five males, which was separated from the rest of the body and incubated in GM before plating.

The microbial diversity in the bedbug environment (stock cultures), and thus possible origins of the microbes on the paramere, the hemolymph, and the mesospermae was also assessed. To this end, we incubated filter papers from the pots in which the stock populations are kept in 5 ml sterile GM. To ensure high enough bacterial numbers from the tissue samples for detecting them on the different growth media, we incubated all tissues in 250 μ l GM for 4 h at 26°C, after which 30 μ l were plated using sterile glass beads.

The Effect of Wounding and Feeding on the Microbiome of Female Bedbugs

To test whether feeding or mating has an effect on the type or number of different bacteria found in females, we randomly assigned 3-week old females ($N = 16$) to the following four treatments (each $N = 4$): (1) unfed virgins, (2) fed virgins, (3) unfed mated, and (4) fed mated. Females were mated with randomly picked males from our stock population. We allowed the males to copulate for as long as they wanted, measured the mating duration, and removed them immediately after they let go of the female. Fed females mated almost 20 s longer than unfed females (Table 3). Thirty minutes after mating, female bedbugs were sampled for bacteria. We also removed and sampled the paramere of the males ($N = 7$) by incubating it in GM. The males ($N = 7$) and females ($N = 3$) were also rinsed in GM to obtain cuticular bacteria.

Isolation and Cultivation of Bacteria

Before dissection, females were sterilized using a kimtech tissue dipped in 96% ethanol. Then hemolymph samples were collected by introducing a sterilized glass capillary pulled to a fine point between the second and third abdominal sternite. Subsequently, on average, 0.5 μ l hemolymph were added to 15 μ l of GM in a 0.5 ml Eppendorf tube on ice. Then females were dissected in GM and the mesospermae removed and rinsed in 10 μ l of sterile GM on a glass slide. After rinsing, the mesospermae was put into a 0.5 ml Eppendorf tube containing 15 μ l of sterile GM on ice. Using a sterile plastic pestle (Z359947, Sigma-Aldrich, Dorset, UK) and thorough vortexing, we homogenized the mesospermae samples before spreading them on agar plates.

Each hemolymph and mesospermae sample was split into two, i.e., 7 μ l each, and spread with sterile glass beads (3 mm, Merck) on 1.5% LB agar plates (60 mm). We also ran procedural controls to check for potential contamination. Agar plates were incubated at 26°C until visible colonies were present, at most 48 h. After incubation, we screened for visible colonies and different colony morphotypes. These were picked and re-cultured

to obtain single clonal cultures. From 74 samples, glycerol stocks were prepared and served as a culture collection for future analysis.

DNA Extraction and PCR Amplification of the 16S rRNA Gene

Each glycerol stock was used to grow a liquid culture for extracting DNA (MO BIO, Ultra Clean, Microbial DNA Isolation Kit, Cat. Nr. 12224-250, Cambio Ltd., Cambridge, UK). DNA samples were sent to SourceBioscience Geneservice™ (Nottingham, UK) for PCR amplification and forward sequencing of a fragment of the 16S rRNA gene using the universal primer 27F (5'-AGAGTTTGTATCMTGGCTCAG). The obtained sequences were trimmed at the 5' and 3' end to remove ambiguous parts, i.e., non-identified nucleotides, in order to optimize blast results and sequence alignment. Due to their poor quality, four sequences had to be excluded from the analysis. We used blast2go¹ for the blasting of the sequences and ClustalX² for the sequence alignment. The Maximum Likelihood phylogenetic tree of the bacterial sequences was determined by using the web interface RAXML Black Box³ (Figure 1) and NJplot to draw the phylogenetic tree⁴ (Figure 1).

QS Assay

N-acyl homoserine lactone (AHL) production was measured as a surrogate for QS using the indicator strain *Chromobacterium violaceum* (CV026), which was assayed with all samples from our bacterial culture collection. CV026 does not produce C6-HSL but does produce violacein (purple pigment) in response to the presence of exogenous C6-HSL. This violacein production in strain CV026 is inducible by AHLs with *N*-acyl side chains from C4 to C8 in length. In contrast, AHLs with *N*-acyl side chains from C10 to C14 inhibit violacein production. Therefore, CV026 can be used as an indicator strain to detect a variety of AHLs.

Chromobacterium violaceum was first grown overnight in liquid LB broth. By mixing 50 ml of the overnight culture with 200 μ l of warm 1% agar medium, we produced assay plates (10 ml in 90 mm Petri dishes). Then, we punched holes in the agar using the top end of a glass Pasteur pipette. For each sample from our bacteria stock library, two overnight cultures were produced ($N = 148$ overnight cultures). From each of those, we ran two replicates on two different plates. We added 50 μ l of overnight culture to a hole in an assay plate. As a positive control, we also assayed the indicator strain *C. violaceum* ATTC 31532 that produces AHL. After inoculation with a bacteria sample, the plates were incubated for 48 h at 30°C. Then, each sample was scored for the occurrence of purple coloration around the well, which is the positive test of AHL production. No coloration indicated a negative test. For each positive assay, we measured the zone diameter twice in a perpendicular fashion and subsequently calculated the area of the zone in square millimeters.

¹<http://www.blast2go.com>.

²<http://www.clustal.org/clustal2/>.

³<https://embnet.vital-it.ch/raxml-bb/>

⁴<http://pbil.univ-lyon1.fr/software/njplot.html>.

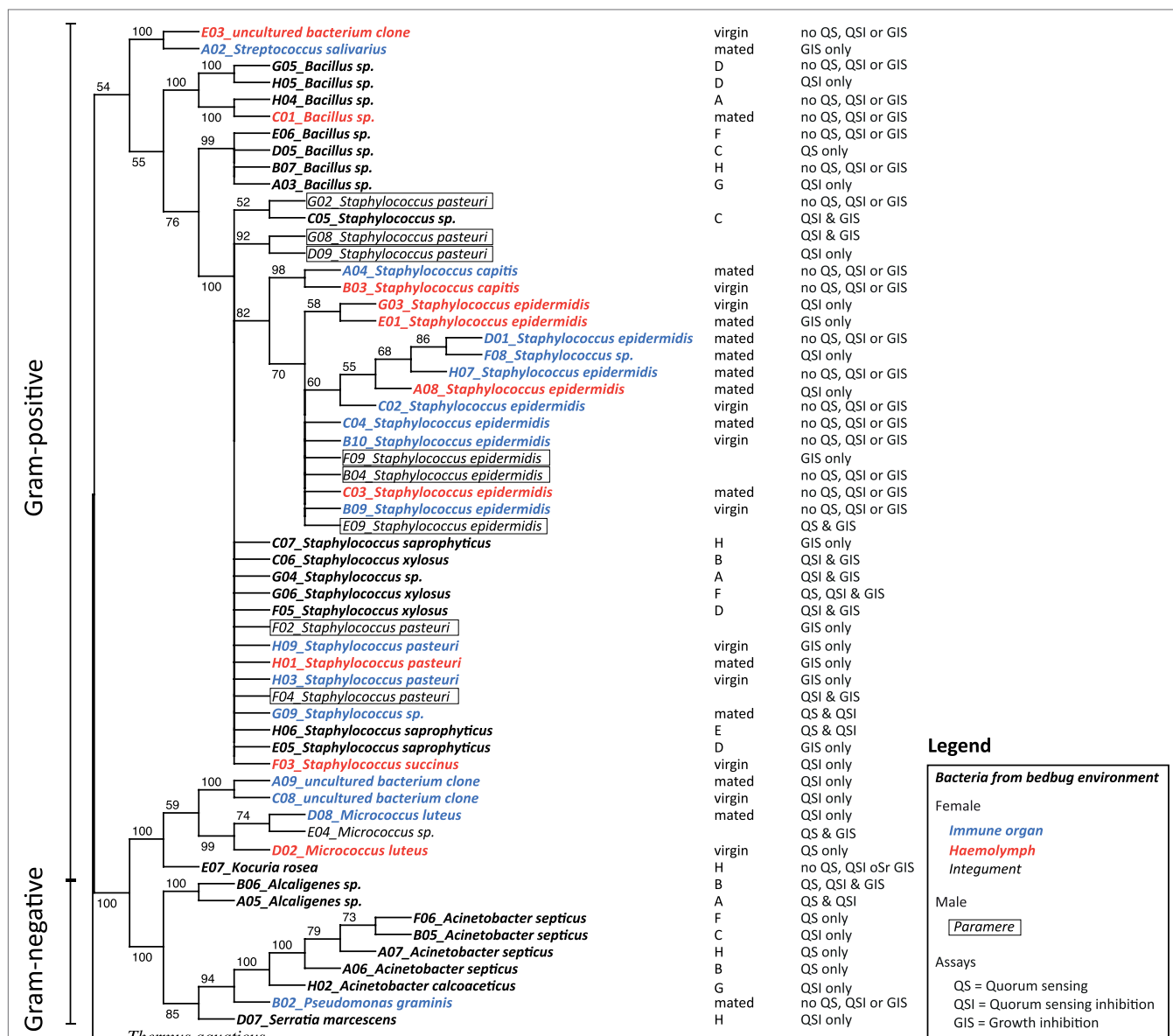


FIGURE 1 | Phylogenetic tree [RAxML rapid bootstrap (35)], reconstructed for the 16S rDNA gene sequences from bacteria in bedbugs and their environment. Bacteria are given in different fonts depending on their location. Bacteria from filter papers on which bedbug stock populations were kept are given in bold italic, bacteria from female hemolymph samples are given in red italic, bacteria from female immune organs are given in blue italic, bacteria from female integuments are given in italic, and bacteria from parameres are given in italic and boxed. In the column next to the tree, the stock population ID and the mating status of females are presented, including the ability of the isolated bacteria to perform quorum sensing (QS), quorum sensing inhibition (QSI), or growth inhibition of the indicator strain (GIS).

QSI Assay

Similar to the QS assay, we used a *C. violaceum* strain (ATTC 12472) as an indicator for QSI. In this strain, *N*-(3-hydroxydecanoyl)-L-homoserine lactone controls violacein production, a purple pigment, by QS (36). As the growth medium did not affect the QS assay, we produced assay plates in the same way as described above only with LB medium. Again, we tested all samples from our culture collection twice and with two overnight cultures. Plates were incubated for 48 h at 30°C. As the indicator strain is

producing the purple pigment constantly when growing, a positive test of QSI is seen as a white, milky zone around the well. A negative test would be no zone and purple coloration right up to the edge of the well. In our case, we also observed clear zones around the well, which we scored as growth inhibition of the indicator strain (GIS). For each positive assay (QSI and GIS), we measured the zone diameter twice in a perpendicular fashion and subsequently calculated the area of the zone in square millimeters.

Statistical Analysis

- (i) *Site-specific, culturable microbiome of the bedbug*: the number of culturable bacteria was similar on the six growth media used (Table 2). We thus pooled the data to identify differences between numbers of species found in different organs and mating status with a Chi-squared test. We compared the number of females from which bacteria could be cultivated versus the number of females from which no bacteria could be cultivated using a Fisher Exact test.
- (ii) *The effect of wounding and feeding on the microbiome of female bedbugs*: due to the low number of different bacteria species found in the differently treated females, we did not perform a statistical analysis (Table 3). As the sample sizes were limited, we fitted Fisher Exact tests to identify differences between numbers of females from which bacteria could be cultivated in comparison to females from which no bacteria could be cultivated with either mating or feeding status as a factor.

Quorum sensing, QSI, and GIS were analyzed in a qualitative manner by giving an account of how many culturable bacteria species were able to perform QS, QSI, or GIS in relation to mating status, bedbug population, and tissue. All statistical analyses were conducted using R 3.4.1 (37).

RESULTS

Overall Diversity of Species and of Quorum Communication

In total, we identified 20 different culturable bacterial species across all our samples (five Gram-negative, thirteen Gram-positive bacteria, and two clones that were not identified) (Table 1; Figure 1). Ten species were cultured from the environment of the

bedbug (four Gram-negative, six Gram-positive bacteria), eleven from female tissues, and two from males (Table 1). Samples from the same bacteria species that were sampled from different collection sites (tissues) clustered together in the phylogenetic analysis (Figure 1).

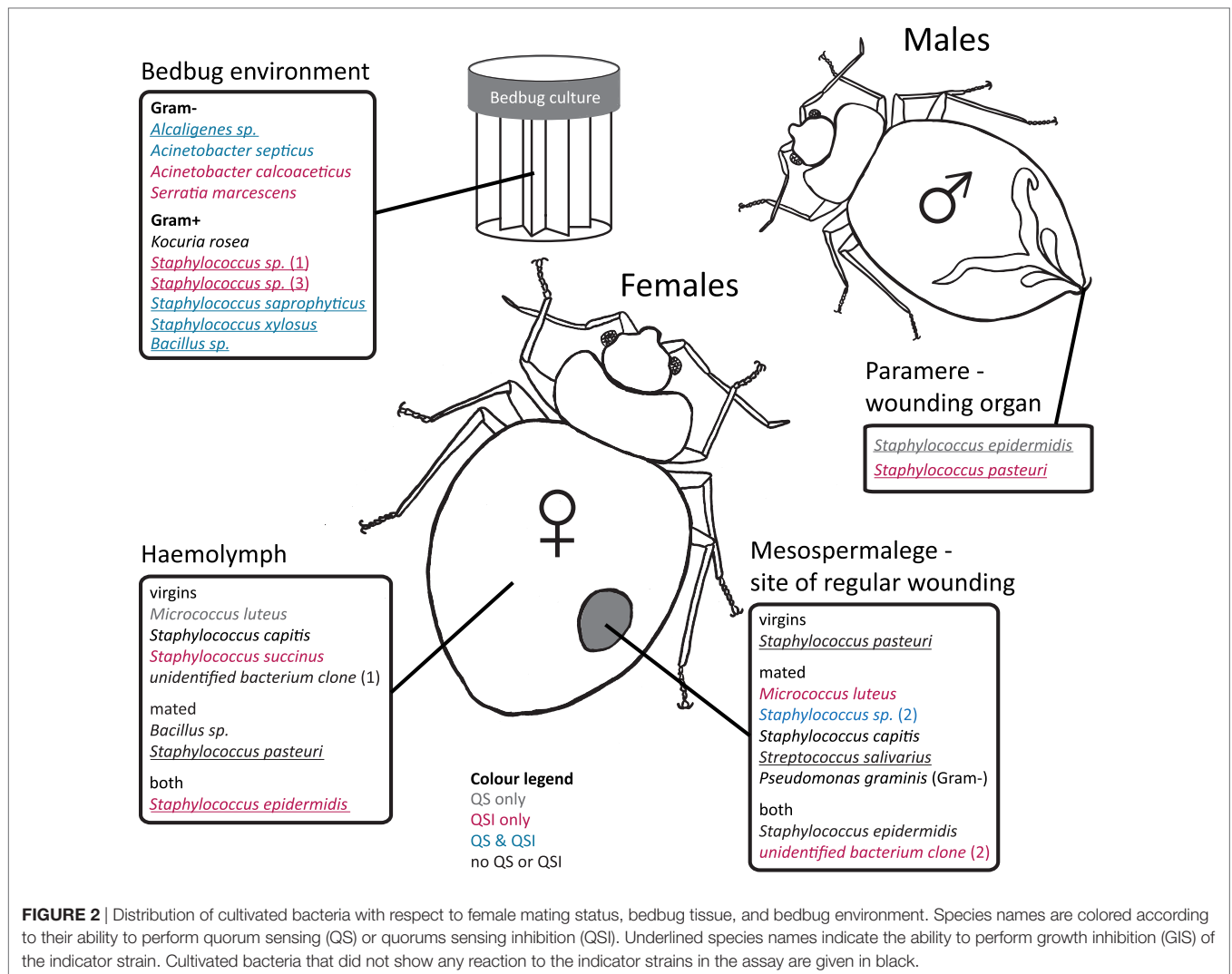
Overall, 56% of the cultivated bacteria isolates showed QS (Gram-negative bacteria: 40%; Gram-positive bacteria: 67%), 72% showed QSI (Gram-negative bacteria: 80%; Gram-positive bacteria: 67%), and 50% showed growth inhibition of the indicator strain (GIS) (Gram-negative bacteria: 20%; Gram-positive bacteria: 67%). Generally, QSI response was strongest showing a mean area of $175.98 \text{ mm} \pm 151.13 \text{ mm}$ (mean \pm SD), GIS zones were on average $135.20 \text{ mm} \pm 138.97 \text{ mm}$, and QS areas $45.50 \text{ mm} \pm 36.30 \text{ mm}$. The expression of QS, QSI, and GIS of the different bacteria isolates was dependent also on geographic origin, mating status and tissue (Figure 2). For example, eight of the nine bacteria species found in the environment showed QS, QSI, or GIS or a combination of those three (Figure 2) but only one of the bacteria species cultured from female tissues performed both QS and QSI. The strength of response in QS, QSI, and GIS varied between species (Figure 3). From five of the six field-caught bedbug populations, environmental bacteria could be cultivated that were able to quorum sense and/or to inhibit the growth of the indicator strain (Figure 1). Of the lab populations, one harbored bacteria that could perform QSI, but not QS. Environmental bacteria grown from the second laboratory population could perform QS, QSI, and GIS. All bedbug populations had cultivable bacteria that were able to quorum quench the signal of the indicator strain (Figure 1).

In total, we recovered 24 isolates from the bedbug environments of the eight stock populations (UK: 14, Kenya: 3, and Lab: 7). In the environment of the UK bedbug stock populations, we found seven, in the Kenyan population three and in the Lab stocks six different bacteria species. Five bacteria species were only found in one population [A: *Staphylococcus* sp. (1); C: *Staphylococcus* sp.

TABLE 1 | Bacterial species found in bedbugs and their environment for all experiments combined.

Bedbug environment	Female							Male	
	N stocks	Integument	N	Hemolymph	N	Immune organ	N	Paramere	N
<i>Alcaligenes</i> sp. ^{G-}	2	<i>Micrococcus</i> sp.	1	<i>Bacillus</i> sp.	0;1	<i>Pseudomonas graminis</i> ^{G-}	0;1	<i>Staphylococcus epidermidis</i>	3
<i>Acinetobacter septicus</i> ^{G-}	4	<i>Staphylococcus pasteurii</i>	1	<i>Micrococcus luteus</i>	1;0	<i>M. luteus</i>	0;1	<i>Staphylococcus pasteurii</i>	5
<i>Acinetobacter calcoaceticus</i> ^{G-}	1			<i>Staphylococcus capitis</i>	1;0	<i>S. capitis</i>	0;1		
<i>Serratia marcescens</i> ^{G-}	1			<i>S. epidermidis</i>	1;3	<i>S. epidermidis</i>	3;3		
<i>Bacillus</i> sp.	6			<i>Staphylococcus pasteurii</i>	0;1	<i>S. pasteurii</i>	2;0		
<i>Kocuria rosea</i>	1			<i>Staphylococcus succinus</i>	1;0	<i>Staphylococcus</i> sp. (2)	0;1		
<i>Staphylococcus</i> sp. (1)	1			Unidentified bacterial clone 1	1;0	<i>Streptococcus salivarius</i>	0;1		
<i>Staphylococcus</i> sp. (3)	1					Unidentified bacterial clone 2	1;1		
<i>Staphylococcus saprophyticus</i>	3								
<i>Staphylococcus xylosus</i>	3								
Number of units/individuals screened	16		3		13;20		13;20		12

For the environment, two samples from eight stock populations were screened. Hemolymph and immune organs were sampled in virgin and mated females from only one stock population. For these, the number of samples with bacteria was given for both mating states (virgin;mated). The bottom row gives the number of units or individuals screened. G⁻, Gram-negative bacterium.

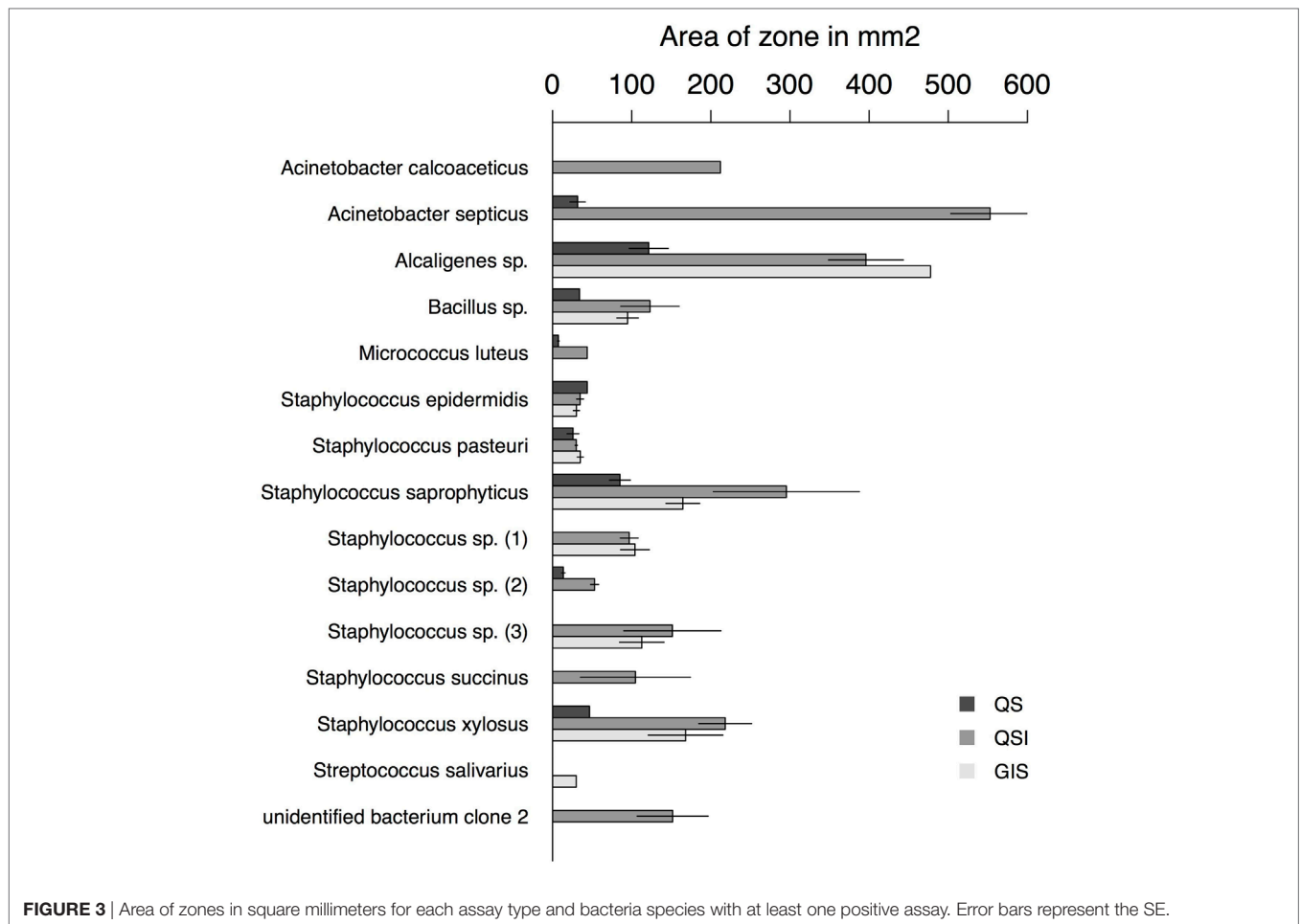


(3); G: *Acinetobacter calcoaceticus*; H: *Kocuria rosea* and *Serratia marcescens*] (Figure 1). In three UK populations (A, B, C), in the Kenyan population and in one Lab stock (G) bacteria showed QS, QSI, and GIS. One UK population contained bacteria that only showed QSI and GIS (D) and another only QS and QSI (E). And the second Lab stock only had bacteria showing QSI (Figure 1).

Site-Specific, Culturable Microbiome of the Bedbug Species

In the hemolymph of a mated female, we found a *Bacillus sp.* From the hemolymph of virgin females, we cultivated *Micrococcus luteus*. The mesospermaleges of a mated and a virgin female yielded an unidentified bacterium clone (2) that clustered with *Micrococcus sp.* (Figure 1). In addition to *M. luteus*, the mesospermaleges of mated females harbored *Streptococcus salivarius* and the Gram-negative *Pseudomonas graminis* (Figure 2). Parameres—the male wounding organs—harbored *Staphylococcus epidermidis* and *Staphylococcus pasteuri* (Figure 2). One female harbored four different types of bacteria (hemolymph: *S. epidermidis* and *S. pasteuri*; mesospermalege: *Staphylococcus sp.* and *S. salivarius*).

Whereas after mating, the number of bacteria species increased in comparison to virgin bedbugs, the proportion of females harboring bacteria decreased after mating (Figure 3; Table 3). From the hemolymph of virgin females, one bacteria species could be grown, but four from the hemolymph of mated females (Table 3). Mating status and tissues did not differ in the number of females from which bacteria could be grown (Fisher's Exact test: $P = 0.19$). Four of twelve mesospermaleges (33%) of mated females contained cultivable bacteria, in contrast to four of five mesospermaleges (80%) of virgin females, a difference that was, however, not significant (Fisher's Exact test: $P = 0.13$) (Table 3). Different growth media did not affect the number (Fisher's Exact test: $P = 1$) or type of bacteria species that could be cultivated (Table 3). Although not significant (Fisher's Exact test: $P = 0.15$), in both experiments, mating reduced the number of culturable bacteria found in the mesospermalege—the site of regular wounding (Figure 4). Overall, female hemolymph showed a lower proportion of cultivable bacteria than the mesospermalege—the site of regular wounding (Figure 4). Mesospermaleges of mated females contained less than half as many cultivable bacteria



as the same organs from virgin females and the hemolymph of showed a similar pattern (**Figure 4**).

Quorum Communication

Bacteria from female tissues seemed to be limited to one form of communication (e.g., only QS in *M. luteus* from hemolymph of virgin female; only QSI in *Staphylococcus* sp. from hemolymph of virgin female; or only GIS in *S. pasteurii* from hemolymph of mated female) (**Figure 2**). Three bacteria species from mesospermaleges of mated females showed QSI (**Figure 2**), one of which was the same in mesospermaleges of virgin females. *S. salivarius* from mated and *S. pasteurii* from virgin female mesospermaleges showed GIS and the hemolymph of mated females contained no bacteria showing either QS or QSI, but two performed GIS. In contrast, the hemolymph of virgin females harbored one cultivable bacteria species with QS, one with QS but none with GIS (**Figure 2**). *M. luteus* from the mesospermalege of mated females perform QSI, whereas the *M. luteus* found in virgin hemolymph were only able to QS. We found a similar contrast for parameres and female tissues. *S. epidermidis* from a paramere performed QS and GIS, *S. epidermidis* found in the hemolymph of females performed QSI and GIS (**Figure 2**). While *S. pasteurii* from a paramere

could quench and inhibit the indicator strain, *S. pasteurii* from females would only inhibit the growth of it.

The Effect of Wounding and Feeding on the Microbiome of Female Bedbugs

Species

In addition to the eight bacteria species found while investigating the site-specific, culturable microbiome of the bedbug, we identified another three when testing for the effect of wounding and feeding on the microbiome of female bedbugs. The proportion of females from which bacteria were cultivated was dependent on the females' mating and feeding status (**Table 3**; Fisher's Exact test: $P = 0.04$).

Mated females harbored a lower number of culturable bacteria species than virgin females (**Table 3**). In addition, fewer mated females harbored bacteria than virgin females (Fisher's Exact test: $P = 0.12$) (**Figure 4**). Mated fed females harbored bacteria whereas we could not grow bacteria from unfed mated females (Fisher's Exact test: $P = 0.14$). From mated fed females, we identified *S. epidermidis*, *S. capitis*, and one *Staphylococcus* sp. From four fed and three unfed virgin females, we could cultivate bacteria (**Table 2**). The proportion of fed and unfed females from which bacteria

could be cultivated did not differ (Fisher's Exact test: $P = 0.12$) (Table 3). Virgin fed females harbored *S. capitis*, *S. succinus*, one *Staphylococcus* sp. and one unidentified bacterium clone (1) similar to *S. salivarius* (Figure 1). The *S. succinus* and the unidentified bacterium clone (1) originated from the same female. From two unfed virgin females, we identified *S. epidermidis* and from one *S. pasteurii*. And from the washed female integuments, we cultivated *S. pasteurii* and a *Micrococcus* sp. Two of the seven screened males harbored two bacteria species, *S. epidermidis* and *S. pasteurii*, on their paramere (Figure 2). These two bacteria species were most frequently found in both experiments (site-specific, culturable microbiome of the bedbug: *S. epidermidis*: 2 parameres, 2 hemolymph samples, 4 mesospermaleges, *S. pasteurii*: 3 parameres, 1 hemolymph sample, 1 mesospermalege; effect of wounding and feeding on the microbiome of female bedbugs: *S. epidermidis*: 1

paramere, 2 hemolymph samples, 2 mesospermaleges, *S. pasteurii*: 2 parameres, and 1 mesospermalege).

Quorum Communication

One bacterium from mated females showed QSI in contrast to three from virgin females. Virgin females also harbored bacteria capable of GIS. Fed females harbored three bacteria species that showed QSI and three that did not signal. From unfed females, we recovered bacteria that were able of QSI and GIS. None of the bacteria found in this experiment showed QS.

DISCUSSION

We advocate the bedbug mating as a suitable system to study the effects of bacteria on reproduction and to address prokaryote and eukaryote communication during wounding in a natural system. We found that most organs of bedbugs harbor bacteria, which are capable of both signaling pathways, QS and QSI. Some of the bacteria were able to stop the growth of an indicator strain indicating potential higher competitiveness. Finally, we show that the microbiome varies between mated and non-mated individuals, between fed and non-fed ones as well as between organs. By focusing on the prokaryotic quorum communication system, we provide a baseline for future research in this unique system.

The Bedbug Microbiome

In 2013, the first in-depth assessment and characterization of the bed bug (*Cimex lectularius*) microbiome was conducted (38). Although variation in diversity and structure was found among geographical locations, the presence of similar bacterial lineages across populations provided evidence for the presence of a *Cimex* core microbiome. To date, several studies including our own have found similar bacterial taxa in bedbug populations from all over the world (31, 38, 39) further supporting the existence of a core microbiome. However, our study also shows considerable variation in the bedbug microbiomes of different tissues, sexes, and reproductive states. In addition, bacteria found in the bedbug environment differed from the ones found in female or on male reproductive organs, hinting at a specific interaction between bacteria and host. Isolates from the same bacterial species that were sampled from different collection sites (tissues) clustered together in the phylogenetic analysis. These bacteria species seem to rather opportunistically colonize tissues and do not select a habitat within hosts. It is, therefore, likely that those bacterial species get transmitted between individuals regularly.

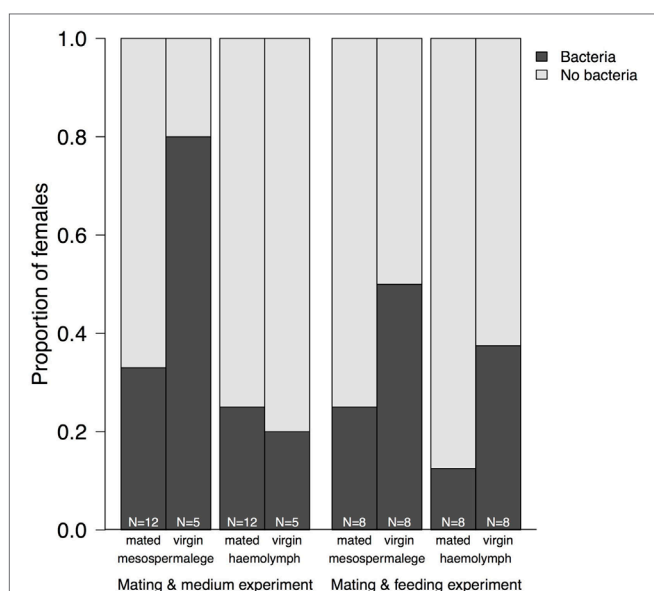


FIGURE 4 | Proportion of mesospermaleges—the site of regular wounding—from mated and virgin females from which bacteria could be cultivated and their hemolymph in contrast to the number of females without bacteria. The sample sizes below indicate the number of females screened for bacteria. “Mating and medium experiment” refers to the screen of the site-specific, culturable microbiome of the bedbug and “Mating and feeding experiment” refers to the test of the effect of wounding and feeding on the microbiome of female bedbugs.

TABLE 2 | Effect of mating and test medium on bacteria presence in female bedbugs.

Treatment	N	Growth medium						Number of different bacteria species	Percent females or samples with bacteria
		LB	GM	NB	NBTA	Potato extract	R-2A		
Mated	12	4	5	4	4	4	4	7	42
Hemolymph		2	2	2	2	2	3	4	25
Mesospermalege		2	3	2	2	2	1	4	33
Virgin	5	3	2	3	2	2	2	3	80
Hemolymph		1	0	0	0	0	0	1	20
Mesospermalege		2	2	3	2	2	2	3	80

Numbers given for growth media are the number of different bacteria species found in each tissue.

TABLE 3 | Effect of mating and feeding on bacteria presence in tissues from female bedbugs.

Treatment	N	Mean mating duration in seconds (\pm SD)	Percent females with bacteria	Number of different bacteria species found
Fed				
Mated	4	99.0 \pm 9.7	75	2
Virgin	4		100	4 ^a
Unfed				
Mated	4	80.5 \pm 22.3	0	0
Virgin	4		75	2

^aTwo different bacteria species from the same individual.

Any bacterium specifically associated with only one organ, i.e., similar to an endosymbiont, might benefit from using QS and QSI to occupy the niche and protect the host organ from intruding bacteria. Most bacteria found in the bedbug environment performed QS, QSI, or GIS in some combination, except for three species. Only one did not perform any QS, QSI, or GI. In contrast, none of the bacteria from male or female bedbugs was performing QS, QSI, or GIS in combination, suggesting that either the host suppresses signaling or no need was present for the bacteria to communicate. Whereas several bacteria species in mated females showed QSI or GIS when isolated from the mesospermalege (the site of regular wounding), only one bacterium showed QS. Whether this indicates a certain degree of specificity of the given bacteria in this tissue remains to be shown.

While organ-specific bacterial communities are well established for humans [e.g., Ref. (1)], whether quorum communication can also be organ-specific seems less clear. Currently, evidence is lacking for such a specificity. But given the intricate interaction and communication between animals and bacteria already described (40), an organ-specific communication would not be unlikely. We found no previous report on quorum communication of bacteria associated with bedbugs. In our study, we not only identified a diverse range of QS and QSI communicating bacterial species but we also found that some aspects appeared to be related to the organ from which the culturables were isolated. For example, we found that *M. luteus* showed QSI when sampled from the mesospermalege but not when sampled from the hemolymph. Similarly, *S. epidermidis* showed QS and GIS on the paramere, QSI and GIS in the hemolymph, and no QS, QSI, or GIS in the mesospermalege. Although we examined QS and QSI *in vitro*, it is unlikely that these differences arose from the *in vitro* conditions, because we treated all culturables the same way. At least some aspect correlated to eliciting a QS or QSI response *in vitro* must have been different between the organs.

Sex Differences and the Effect of Mating on the Microbiome and Its Quorum Communication

The sex differences in the microbiome of bedbugs that we found agree with many other species and are not surprising, given the

large habitat difference the male and female reproductive organs represent to microbes. For example, mosquitoes, mice, and humans show differences in microbiome diversity or abundance of bacteria taxa between the sexes (41–43). Sex differences in the microbial community can even lead to sex-specific hormone regulation in mice (42), suggesting differences in communication between the microbiome and its host. However, habitat differences [e.g., Ref. (15, 44)] and sexual transmission are not the only determinants of sex differences in the microbial community of animal genitalia. For example, Gendrin et al. (45) showed that the deposition of infectious microbes on the genital plate caused a systemic, rather than a localized immune response in males but not in females.

Mated females had fewer bacterial species in the mesospermalege than virgin females, a trend that was also found in the hemolymph. This observation implies that the site of regular wounding—the mesospermalege—has a role in controlling bacteria transmission or bacteria growth. This is consistent with the finding that in bedbug females, the growth inhibition of bacteria and antibacterial activity of the mesospermalege is stronger in mated than virgin females (O. Otti, unpublished data). It is, therefore, possible that part of the observed bacterial reduction in the mesospermalege of mated females might be caused by an upregulation of the growth inhibition factors during and/or the production of constitutive immune agents, such as lysozyme, during or even in anticipation of mating [see Ref. (19) for an example in *Drosophila*].

None of the bacteria recovered from mated females showed QS. Actually, QS was only found in one bacterial clone from the hemolymph of a virgin. Bacteria from the hemolymph of mated females did neither quorum sense nor quench, whereas bacteria from the hemolymph of virgin females were capable of both signaling pathways. Females might control the signaling or growth of bacteria in the hemolymph during reproduction to minimize the risk of a systemic infection. Although the investment into pathogen protection is often reduced *via* immune suppression during reproduction (46), regular wounding combined with a threat of genital infection might select for means to control and localize bacterial growth.

The Effect of (Blood) Feeding on the Microbiome and Its Quorum Communication

Feeding increased the number of species collected from female bedbugs. Unfed mated females were even free from culturable bacteria. Only two species were collected from unfed females. Interestingly, bacteria from fed females only showed QSI or no signaling at all, whereas unfed females were capable of QSI and GIS. The reason for these differences will have to be investigated further.

CONCLUSION

We characterize the culturable microbiome for a system with natural regular sexual wounding and show that the host's sex as well as feeding and mating, are associated with striking

differences in the microbiome and the quorum communication system. Despite the uniqueness of the system, the frequency of copulatory wounding suggests that similar differences might be worth studying in other eukaryotic hosts and may represent an important part of the metaorganism.

AUTHOR CONTRIBUTIONS

OO, PD, KH, and KR conceived the idea and designed the experiment; performed the statistical analysis; and interpreted the results and wrote the manuscript. OO and PD carried out the experiment. All authors read and approved of the final manuscript.

REFERENCES

- Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* (2012) 486:207–14. doi:10.1038/nature11234
- Kostic AD, Howitt MR, Garrett WS. Exploring host-microbiota interactions in animal models and humans. *Genes Dev* (2013) 27:701–18. doi:10.1101/gad.212522.112
- Goodrich JK, Davenport ER, Waters JL, Clark AG, Ley RE. Cross-species comparisons of host genetic associations with the microbiome. *Science* (2016) 352:532–5. doi:10.1126/science.aad9379
- Bosch TCG, McFall-Ngai MJ. Metaorganisms as the new frontier. *Zoology* (2011) 114:485–190. doi:10.1016/j.zool.2011.04.001
- Dong Y, Manfredini F, Dimopoulos G. Implication of the mosquito midgut microbiota in the defense against malaria parasites. *PLoS Pathog* (2009) 5:e1000423. doi:10.1371/journal.ppat.1000423
- Feldhaar H. Bacterial symbionts as mediators of ecologically important traits of insect hosts. *Ecol Entomol* (2011) 36:533–43. doi:10.1111/j.1365-2311.2011.01318.x
- Moran NA, Wernegreen JJ. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* (2000) 15:321–6. doi:10.1016/S0169-5347(00)01902-9
- Rilling MC, Antonovics J, Caruso T, Lehmann A, Powell JR, Veresoglou SD, et al. Interchange of entire communities: microbial community coalescence. *Trends Ecol Evol* (2015) 30:470–6. doi:10.1016/j.tree.2015.06.004
- Theopold U, Schmidt O, Söderhäll K, Dushay MS. Coagulation in arthropods: defence, wound closure and healing. *Trends Immunol* (2004) 25:289–94. doi:10.1016/j.it.2004.03.004
- Siva-Jothy MT. Reproductive immunity. In: Rolff J, Reynolds S, editors. *Insect Infection and Immunity: Evolution, Ecology, and Mechanisms*. Oxford, UK: Oxford University Press (2009). p. 241–8.
- Reinhardt K, Anthes N, Lange R. Copulatory wounding and traumatic insemination. *Cold Spring Harb Perspect Biol* (2015) 7:a017582. doi:10.1101/cshperspect.a017582
- Otti O. Genitalia-associated microbes. *Insect Sci* (2015) 22:325–39. doi:10.1111/1744-7917.12183
- Hirsh DC. The genital tract as a microbial habitat. In: Hirsh DC, Zee YC, editors. *Veterinary Microbiology*. Oxford, UK: Wiley-Blackwell (1999). p. 190–1.
- Knell RJ, Webberley KM. Sexually transmitted diseases of insects: distribution, evolution, ecology and host behaviour. *Biol Rev Camb Philos Soc* (2004) 79:557–81. doi:10.1017/S1464793103006365
- Reinhardt K. Natural selection and genital variation: a role for the environment, parasites and sperm ageing? *Genetica* (2010) 138:119–27. doi:10.1007/s10709-009-9360-4
- Virecoulon F, Wallet F, Fruchart-Flamenbaum A, Rigot JM, Peers MC, Mitchell V, et al. Bacterial flora of the low male genital tract in patients consulting for infertility. *Andrologia* (2005) 37:160–5. doi:10.1111/j.1439-0272.2005.00673.x
- Otti O, McTighe AP, Reinhardt K. *In vitro* antimicrobial sperm protection by an ejaculate-like substance. *Funct Ecol* (2013) 27:219–26. doi:10.1111/1365-2435.12025

ACKNOWLEDGMENTS

We thank Sara Bellinva for comments on the manuscript. PD and KH received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 655914 and 657096. KH was also supported by the Wellcome Trust. KR was supported by an advanced postdoctoral fellowship by the VolkswagenStiftung, OO by a fellowship from the Swiss National Science Foundation (PA00P3_124167/1). This publication was funded by the German Research Foundation (DFG) and the University of Bayreuth in the funding programme Open Access Publishing.

- Zheng H, Zhou L, Yang X, Wang D, Liu J. Cloning and characterization of a male-specific defensin-like antimicrobial peptide from the tick *Haemaphysalis longicornis*. *Dev Comp Immunol* (2012) 37:207–11. doi:10.1016/j.dci.2011.10.004
- Zhong W, McClure CD, Evans CR, Mlynski DT, Immonen E, Ritchie MG, et al. Immune anticipation of mating in *Drosophila*: *Turandot M* promotes immunity against sexually transmitted fungal infections. *Proc Biol Sci* (2013) 280:20132018. doi:10.1098/rspb.2013.2018
- Dethlefsen L, McFall-Ngai M, Relman DA. An ecological and evolutionary perspective on human-microbe mutualism and disease. *Nature* (2007) 449:811–8. doi:10.1038/nature06245
- Kaltenpoth M, Engl T. Defensive microbial symbionts in hymenoptera. *Funct Ecol* (2014) 28:315–27. doi:10.1111/1365-2435.12089
- Weiss B, Aksoy S. Microbiome influences on insect host vector competence. *Trends Parasitol* (2011) 27:514–22. doi:10.1016/j.pt.2011.05.001
- Smith D, Subramanian S, Lamont J, Bywater-Ekegärd M. Signaling in the phytomicrobiome: breadth and potential. *Front Plant Sci* (2015) 6:709. doi:10.3389/fpls.2015.00709
- Gao B, Bian X, Liang C, Pengcheng T, Hongyu R, Kun L. Organophosphate diazinon altered quorum sensing, cell motility, stress response, and carbohydrate metabolism of gut microbiome. *Toxicol Sci* (2017) 157:354–64. doi:10.1093/toxsci/kfx053
- Waters CM, Bassler BL. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* (2005) 21:319–46. doi:10.1146/annurev.cellbio.21.012704.131001
- Miller MB, Bassler BL. Quorum sensing in bacteria. *Annu Rev Microbiol* (2001) 55:165–99. doi:10.1146/annurev.micro.55.1.165
- Bauer WD, Robinson JB. Disruption of bacterial quorum sensing by other organisms. *Curr Opin Biotechnol* (2002) 13:234–7. doi:10.1016/S0958-1669(02)00310-5
- Rendueles O, Ghigo J-M. Mechanisms of competition in biofilm communities. In: Ghannoum M, Parsek M, Whitely M, Mukherjee PK, editors. *Microbial Biofilms*. Washington, USA: ASM Press (2015). p. 319–42.
- Grandclement C, Tannières M, Moréra S, Dessaux Y, Faure D. Quorum quenching: role in nature and applied developments. *FEMS Microbiol Rev* (2016) 40:86–116. doi:10.1093/femsre/fuv038
- Ismail AS, Valastyan JS, Bassler BL. A host-produced autoinducer-2 mimic activates bacterial quorum sensing. *Cell Host Microbe* (2016) 19:470–80. doi:10.1016/j.chom.2016.02.020
- Reinhardt K, Naylor RA, Siva-Jothy MT. Potential sexual transmission of environmental microbes in a traumatically inseminating insect. *Ecol Entomol* (2005) 30:607–11. doi:10.1111/j.0307-6946.2005.00730.x
- Reinhardt K, Naylor R, Siva-Jothy MT. Reducing a cost of traumatic insemination: female bedbugs evolve unique organ. *Proc Biol Sci* (2003) 270:2371–5. doi:10.1098/rspb.2003.2515
- Reinhardt K, Naylor RA, Siva-Jothy MT. Situation exploitation: higher male mating success when female resistance is reduced by feeding. *Evolution* (2009) 63:29–39. doi:10.1111/j.1558-5646.2008.00502.x
- Stutt AD, Siva-Jothy MT. Traumatic insemination and sexual conflict in the bed bug *Cimex lectularius*. *Proc Natl Acad Sci U S A* (2001) 98:5683–7. doi:10.1073/pnas.101440698

35. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* (2008) 57:758–71. doi:10.1080/10635150802429642
36. Morohoshi T, Kato M, Fukamachi K, Kato N, Ikeda T. N-acylhomoserine lactone regulates violacein production in *Chromobacterium violaceum* type strain ATCC12472. *FEMS Microbiol Lett* (2008) 279:124–30. doi:10.1111/j.1574-6968.2007.01016.x
37. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (2017). Available from: <https://www.R-project.org/>
38. Meriweather M, Matthews S, Rio R, Baucom RS. A 454 survey reveals the community composition and core microbiome of the common bed bug (*Cimex lectularius*) across an urban landscape. *PLoS One* (2013) 8:e61465. doi:10.1371/journal.pone.0061465
39. Cockburn C, Amoroso M, Carpenter M, Johnson B, McNeive R, Miller A, et al. Gram-positive bacteria isolated from the common bed bug, *Cimex lectularius*. *Entomol Am* (2013) 119:23–9. doi:10.1664/12-RA-042R.1
40. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* (2013) 110:3229–36. doi:10.1073/pnas.1218525110
41. Haro C, Rangel-Zúñiga OA, Alcalá-Díaz JF, Gómez-Delgado F, Pérez-Martínez P, Delgado-Lista J, et al. Intestinal microbiota is influenced by gender and body mass index. *PLoS One* (2016) 11:e0154090. doi:10.1371/journal.pone.0154090
42. Markle JGM, Frank DN, Mortin-Toth S, Robertson CE, Feazel LM, Rolfe-Kampczyk U, et al. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* (2016) 339:1084–8. doi:10.1126/science.1233521
43. Valiente Moro C, Tran FH, Nantenaina Raharimalala F, Ravelonandro P, Mavingui P. Diversity of culturable bacteria including *Pantoea* in wild mosquito *Aedes albopictus*. *BMC Microbiol* (2013) 13:70. doi:10.1186/1471-2180-13-70
44. Rawls JF, Mahowald MA, Ley RE, Gordon JL. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* (2006) 127:423–33. doi:10.1016/j.cell.2006.08.043
45. Gendrin M, Welchman DP, Poidevin M, Hervé M, Lemaitre B. Long-range activation of systemic immunity through peptidoglycan diffusion in *Drosophila*. *PLoS Pathog* (2009) 5:e1000694. doi:10.1371/journal.ppat.1000694
46. Lawnczak MKN, Barnes AI, Linklater JR, Boone JM, Wigby S, Chapman T. Mating and immunity in invertebrates. *Trends Ecol Evol* (2007) 22:48–55. doi:10.1016/j.tree.2006.09.012

Conflict of Interest Statement: 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.

Copyright © 2017 Otti, Deines, Hammerschmidt and Reinhardt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



F-Type Lectins: A Highly Diversified Family of Fucose-Binding Proteins with a Unique Sequence Motif and Structural Fold, Involved in Self/Non-Self-Recognition

Gerardo R. Vasta^{1*}, L. Mario Amzel², Mario A. Bianchet^{2,3}, Matteo Cammarata⁴, Chiguang Feng¹ and Keiko Saito⁵

¹ Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, University of Maryland, Baltimore, Baltimore, MD, United States, ² Department of Biophysics and Biophysical Chemistry, School of Medicine, Johns Hopkins University, Baltimore, MD, United States, ³ Department of Neurology, School of Medicine, Johns Hopkins University, Baltimore, MD, United States, ⁴ Department of Earth and Marine Sciences, University of Palermo, Palermo, Italy, ⁵ Department of Marine Biotechnology, Institute of Marine and Environmental Technology, University of Maryland Baltimore County, Baltimore, MD, United States

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, United States

Reviewed by:

Miki Nakao,
Kyushu University, Japan
Klaus Ley,
La Jolla Institute for Allergy and
Immunology (LJI), United States
Ulrich Theopold,
Stockholm University, Sweden

*Correspondence:

Gerardo R. Vasta
gvasta@som.umaryland.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 31 August 2017

Accepted: 10 November 2017

Published: 29 November 2017

Citation:

Vasta GR, Amzel LM, Bianchet MA, Cammarata M, Feng C and Saito K (2017) F-Type Lectins: A Highly Diversified Family of Fucose-Binding Proteins with a Unique Sequence Motif and Structural Fold, Involved in Self/Non-Self-Recognition. *Front. Immunol.* 8:1648. doi: 10.3389/fimmu.2017.01648

The F-type lectin (FTL) family is one of the most recent to be identified and structurally characterized. Members of the FTL family are characterized by a fucose recognition domain [F-type lectin domain (FTLD)] that displays a novel jellyroll fold (“F-type” fold) and unique carbohydrate- and calcium-binding sequence motifs. This novel lectin family comprises widely distributed proteins exhibiting single, double, or greater multiples of the FTLD, either tandemly arrayed or combined with other structurally and functionally distinct domains, yielding lectin subunits of pleiotropic properties even within a single species. Furthermore, the extraordinary variability of FTL sequences (isoforms) that are expressed in a single individual has revealed genetic mechanisms of diversification in ligand recognition that are unique to FTLs. Functions of FTLs in self/non-self-recognition include innate immunity, fertilization, microbial adhesion, and pathogenesis, among others. In addition, although the F-type fold is distinctive for FTLs, a structure-based search revealed apparently unrelated proteins with minor sequence similarity to FTLs that displayed the FTLD fold. In general, the phylogenetic analysis of FTLD sequences from viruses to mammals reveals clades that are consistent with the currently accepted taxonomy of extant species. However, the surprisingly discontinuous distribution of FTLDs within each taxonomic category suggests not only an extensive structural/functional diversification of the FTLs along evolutionary lineages but also that this intriguing lectin family has been subject to frequent gene duplication, secondary loss, lateral transfer, and functional co-option.

Keywords: F-type lectins, fuclectins, structural modeling, glycan recognition, fucose-binding, self/non-self-recognition, innate immunity

INTRODUCTION

Recognition of glycans exposed on the surface of microbial pathogens and parasites by the host’s cell-associated and soluble lectins is considered the initial key step in the innate immune response of both invertebrates and vertebrates (1–5). Members of several lectin families characterized by unique sequence motifs and structural folds such as C-type lectins (CTLs) (6), peptidoglycan

binding proteins (7), ficolins (8), pentraxins (PXNs) (9), galectins (10), and most recently, F-type lectins (FTLs) (11–14) have been implicated in immune surveillance and homeostasis. However, the participation of these and other lectin families in multiple intra- and extracellular functions including folding, sorting, and secretion of glycoproteins, cell–cell interactions, and signaling and transport in early development, tissue repair, and general cell functions, as well as host colonization by microbial pathogens and parasites have also been firmly established (5).

F-type lectins are fucose-binding proteins of wide taxonomic distribution from viruses to vertebrates and constitute the most recently identified lectin family (11–14). They are characterized by a fucose recognition domain [F-type lectin domain (FTLD)] that displays a novel fold (the “F-type” fold) consisting of a β -barrel with jellyroll topology and unique fucose- and calcium-binding sequence motifs (13). Although FTLs can display a single FTLD, sometimes associated with one or more structurally and functionally distinct domains in a single polypeptide, the presence of a variable number of tandemly arrayed FTLDs is also a common occurrence in members of this lectin family. Some FTLs mediate immune recognition (13–16), whereas others are involved in microbial pathogenesis (17–23), fertilization (24–26), and other diverse functions.

The identification of the FTL family was a fortuitous discovery that resulted from the search for fucose-binding CTLs in serum and liver extracts from the striped bass (*Morone saxatilis*) (11, 12). Affinity chromatography on L-fucose-Sepharose yielded a 32 kDa protein (MsaFBP32) that did not require calcium or other divalent cations for binding to cells. Partial Edman sequencing of the protein enabled cDNA and genomic cloning and revealed the presence of two 140-amino acid tandemly arrayed domains. Analysis of the deduced polypeptide sequence of MsaFBP32 failed to identify the signature motif of the CTLs or any of the known lectin families described at the time and suggested that MsaFBP32 represented a novel lectin type. Although no matches to known lectins were initially identified, the search of sequence databases revealed a stretch of N-terminus sequence from a single protein named PXN1-XENLA (27) that shared significant similarity to the MsaFBP32 lectin motif. Surprisingly, PXN1-XENLA, which is described as a PXN-fusion protein cloned from the liver of the African clawed frog (*Xenopus laevis*), consists of an MsaFBP32-like domain linked to a PXN domain that also exhibits lectin activity (28). Furthermore, a search of *X. laevis* and *Xenopus tropicalis* EST databases revealed additional FBPLs different from PXN1-XENLA, with multiple FBPLs. The information obtained enabled the cloning of similar lectins in several fish species and later the *in silico* identification of FBPLs in the growing number of EST and genomic databases for multiple invertebrate and vertebrate species, mostly fish and amphibians (11, 12). Surprisingly, three FBPL tandemly arrayed sequences were identified in the SP2159 ORF from the genome of the capsulated and virulent strain (TIGR4) of *Streptococcus pneumoniae* (11, 12). As a whole, this experimental and *in silico* effort led to the identification of the novel lectin family (FTL family) characterized by proteins present in both prokaryotes and eukaryotes, which displayed the newly identified lectin domain (FTLD), either tandemly

arrayed or in mosaic combinations with other structurally and functionally distinct domains (11, 12).

Structural studies were initiated with the simplest FTL family member carrying a single FTLD, the European eel agglutinin [*Anguilla anguilla* agglutinin (AAA)] (13). These were followed by the FTL from the striped bass (*M. saxatilis*; MsaFBP32) that carries two tandemly arrayed FTLDs (14). The resolution of structures for the AAA and MsaFBP32 complexed with fucose enabled the identification of a novel structural fold (the F-type fold) and the amino acid residues in the carbohydrate recognition domain (CRD) that interact with the fucose ligand, as well as with the subterminal sugar units in fucose-containing oligosaccharides. Furthermore, a fold-based search [Dali database (29)] revealed several proteins display the F-type fold, but that only share negligible sequence homology with FTLs, including discoidins, clotting factors, and fungal and bacterial glycoenzymes (13). This information enabled not only the formulation of proposals about the possible evolutionary origin of the FTLD but also about its functional co-option along vertebrate lineages.

In later functional studies, the biological characterization of FTLs from teleost fish revealed their capacity for pathogen recognition and their roles as opsonins in innate immunity, the characterization of the gene products from the identified *Streptococcus* spp. FTL sequences as virulence factors (lectinolysins) (20–22), and the identification of sperm acrosomal proteins (bindins) from the oyster *Crassostrea gigas* as extremely diversified FTLs with role(s) in fertilization (24–26). In recent years, the exponentially growing number of sequenced genomes from multiple species, ranging from viruses to pro- and eukaryotes has enabled the identification of FTLs in additional taxa, thereby greatly expanding our knowledge about the distribution of the FTLD in nature. In this regard, a rigorous and exhaustive computational study has recently provided significant insight into the taxonomic prevalence of the FTLD (30). Finally, functional studies aimed at elucidating the role(s) of FTLs in innate immunity using the invaluable resources available for the genetically tractable zebrafish model system are ongoing. In the following sections, the most relevant structural and functional aspects of the FTL family are discussed.

STRUCTURAL ASPECTS

The sequence alignment of the *M. saxatilis* FTL (MsaFBP32) and *X. laevis* PXN-fusion protein (PXN1-XENLA) led to the identification of an approximately 140-amino acid long lectin domain and a tentative amino acid sequence motif common to a number of lectins, as well as selected domains present in sequences that had been described in other contexts such as the *Drosophila* furrowed gene and the *Streptococcus* fucose regulon. In turn, this resulted in the identification of a novel fucose-binding lectin family (FTL family) that included both prokaryotes (*S. pneumoniae* TIGR4) and eukaryotes (*Drosophila*, fish, amphibians, and others) (11, 12) (**Figure 1**). The resolution of the structure of the AAA–fucose complex revealed a new lectin fold (FTL fold) and identified the amino acid residues that interact with the non-reducing terminal fucose and coordinate the divalent cation and model those that are established with the subterminal sugar

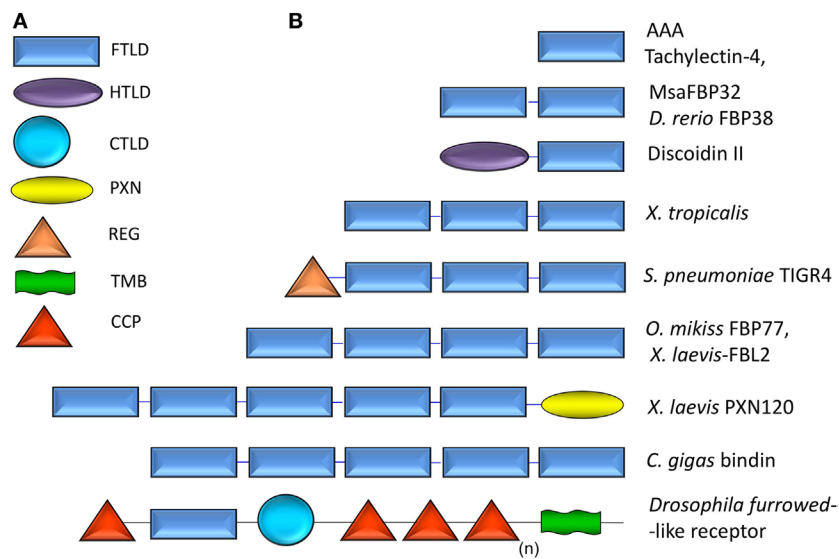


FIGURE 1 | Domain organization of F-type lectins (FTLs) from prokaryotes, invertebrates, and vertebrates: **(A)** schematic illustration of selected domain types found in FTLs: FTLD, F-type lectin domain; HTLD, H-type lectin domain; CTLD, C-type lectin domain; PXN, pentraxin; REG, regulatory; TMB, transmembrane domain; and CCP, complement control protein domain. **(B)** Schematic illustration of the domain organization in selected examples of FTLs (1–5 FTLDs and chimeric molecular species) described in prokaryote, invertebrate, and vertebrate species. The subscript “n” indicates the extended number of CCP domains present in furrowed [adapted from Ref. (11, 12)].

units of an oligosaccharide ligand (13). In turn, this structural information led to the rigorous identification of the FTL fucose- and calcium-binding sequence motifs (13).

FTL Fold

The FTL fold, initially described in the AAA/ α -L-fucose (α -Fuc) complex (Figure 2), consists of a β -barrel with jelly roll topology comprising two β -sheets of three ($\beta 5$, $\beta 8$, and $\beta 11$) and five ($\beta 2$, $\beta 3$, $\beta 10$, $\beta 6$, and $\beta 7$) antiparallel β -strands, respectively, placed against each other (Figure 2A). Two short antiparallel strands ($\beta 4$ and $\beta 9$) close the “bottom” of the barrel, from which the N- and C-termini protrude to form an antiparallel two-strand β -sheet (13). On the “top” face of the barrel, the connecting β -strands from the opposite sheets form five loops (CDR1–5) that surround the heavily positively charged pocket that binds the α -Fuc (Figure 2A). CDR1 is the most protruding loop, and at its exposed apex Glu²⁶ is placed over the aromatic ring of His²⁷, and both over the central hollow. At the side of the barrel, a substructure containing three 3_{10} helices (h2, h3, and h4) tightly coordinates a cation (tentatively identified as calcium) *via* seven oxygen atoms of six residues [Asn³⁵ (O), Asp³⁸ (Od1), Asn⁴⁰ (O), Ser⁴⁹ (O, Og1), Cys¹⁴⁶ (O), and Glu¹⁴⁷ (Oe1)] both from the peptide backbone and side chains in a pentagonal bipyramidal geometry. The distance between the cation binding site and the sugar binding pocket indicates that the divalent cation does not directly interact with the carbohydrate as in CTLs, but that together with two disulfide bridges (Cys⁵⁰-Cys¹⁴⁶ and Cys¹⁰⁸-Cys¹²⁴) and two salt bridges (Arg⁴¹-Glu¹⁴⁹ and Asp⁶⁴-Arg¹³¹) that clamp the structure together, rather stabilizes the fold and shape the key CDR1 and CDR2 loops (13). The AAA subunits can form chloride-induced trimers that contain one cation (Ca²⁺) per domain and several Cl[−] placed

on the three-fold axis, and two trimers can form hexamers with opposing carbohydrate-binding surfaces (Figure 2B).

Primary Fucose-Binding Site

The AAA/ α -Fuc structure revealed that the protein binds to α -Fuc through hydrogen bonds established between the side chains of three basic amino acid residues (N ϵ of His⁵² and the guanidinium groups of Arg⁷⁹ and Arg⁸⁶) situated in a shallow cleft and the axial 4-OH of the sugar. Interactions are also established between this basic triad and the ring O5 and equatorial 3-OH of the sugar (13) (Figure 2C). A unique disulfide bridge formed by contiguous cysteines (Cys⁸² and Cys⁸³) establishes a van der Waals contact with the bond between C1 and C2 of the α -Fuc ring, and the C6, which fits into a hydrophobic pocket formed by His²⁷ and Phe⁴⁵, together with Leu²³ and Tyr⁴⁶ (13). As AAA can also recognize 3-O-methyl-D-galactose and 3-O-methyl-D-fucose, sugars that display similar key configurational features of α -Fuc (i.e., axial hydroxyl and hydrophobic moiety), it becomes clear that as for most animal lectins, the specificity of AAA for α -Fuc is nominal rather than absolute.

Extended Carbohydrate-Binding Site

The AAA/ α -Fuc structure also enabled the modeling of potential interactions between the protein and α -Fuc-containing oligosaccharides such as H and Lewis moieties that are specifically recognized through interactions with amino acid residues located in the so-called “extended binding site” (13) (Figure 2D). AAA recognizes blood group H type 1 (Fuc α 1-2 Gal β 1-3GlcNAc β 1-3Gal β 1-4Glc) and Le^a (Gal β 1-3[Fuc α 1-4]GlcNAc β 1-3Gal β 1-4Glc) oligosaccharides *via* additional interactions established between amino acid residues in CDRs 1–5 that encircle the

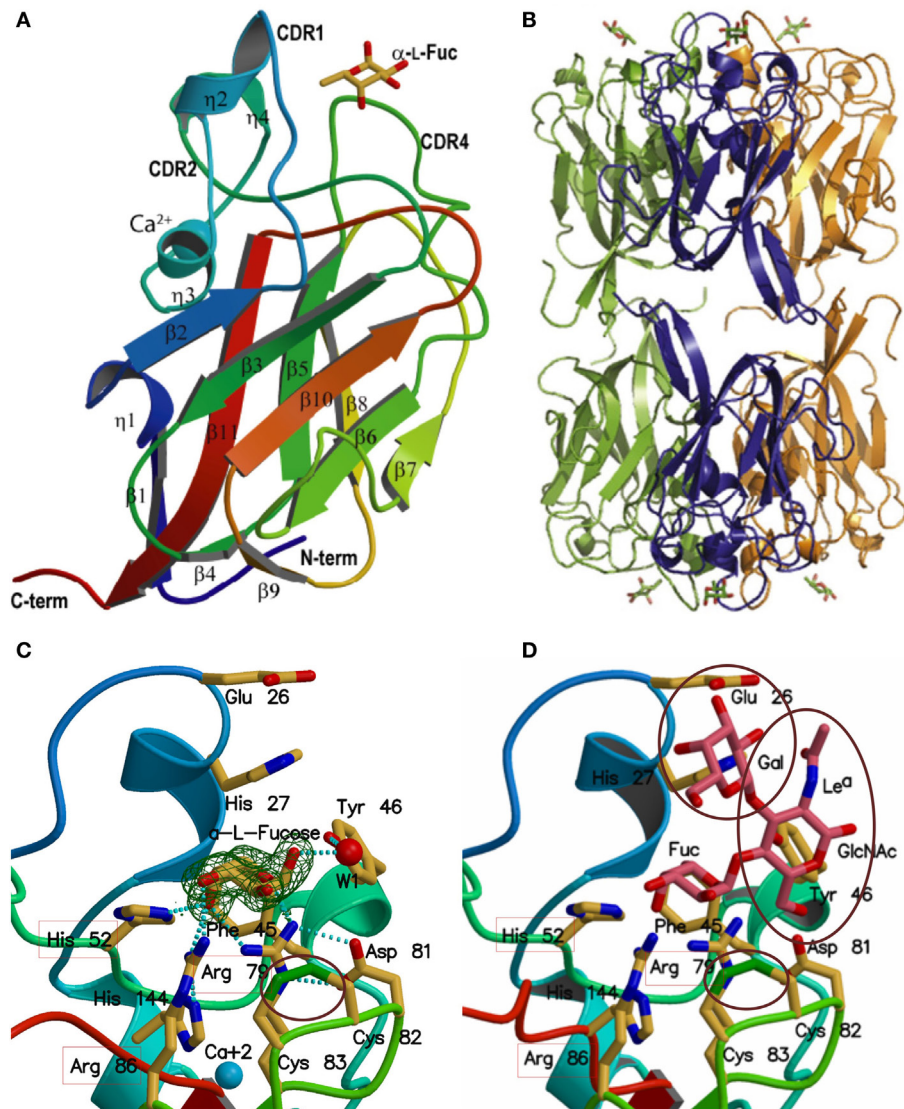


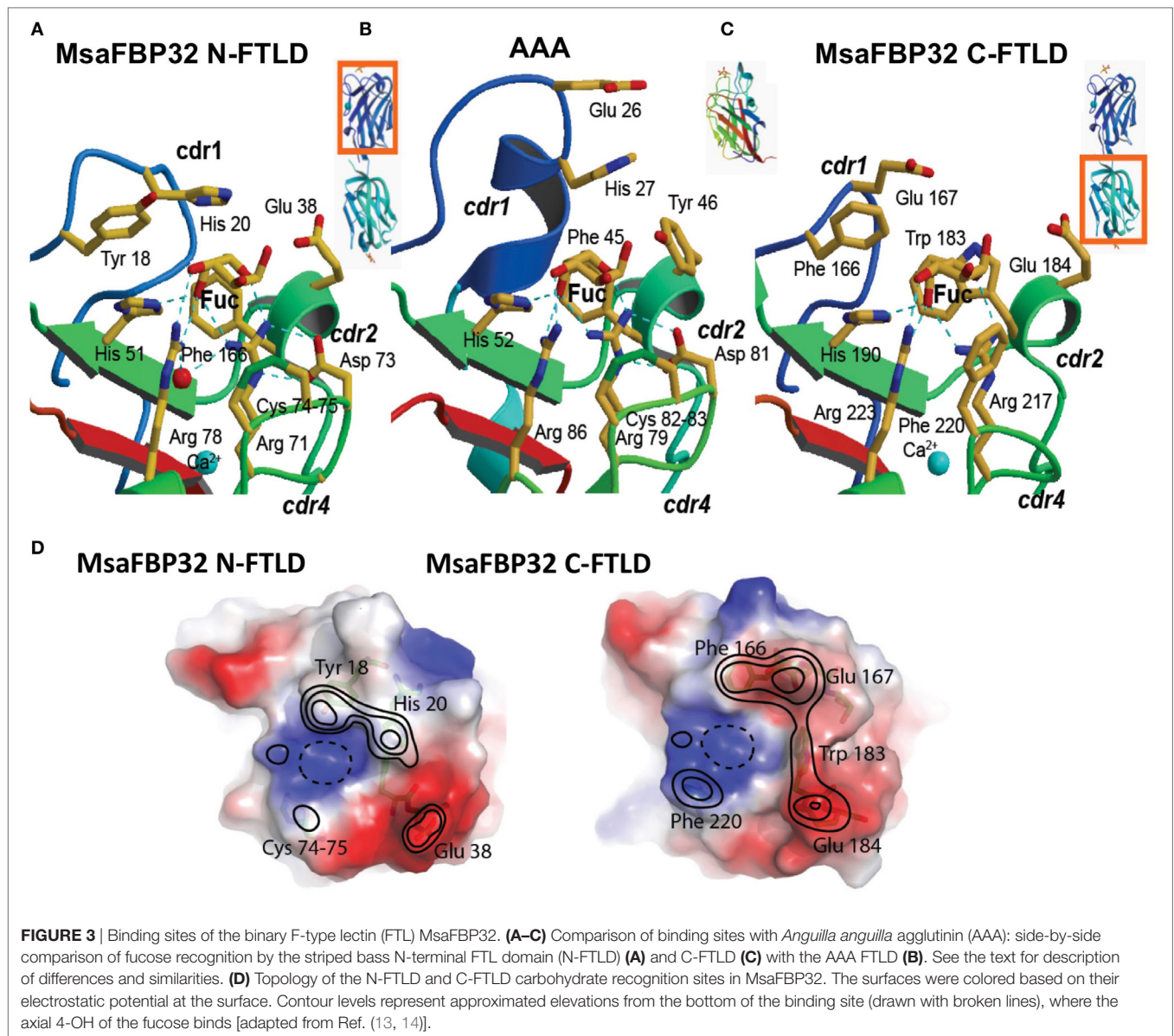
FIGURE 2 | Structure of *Anguilla anguilla* agglutinin (AAA) and quaternary structure of AAA oligomers. **(A)** Ribbon diagram of AAA showing the two β -sheets, the loops (CDRs) encircling the binding site, and 3_{10} helices. Bound α -L-fucose is shown as a stick model above the lectin in yellow. Calcium is shown as a blue sphere. **(B)** Quaternary structure of the AAA hexamer: in each trimer a single chlorine ion coordinated by Lys¹⁶ from each subunit determines the three-fold axis of rotation, and the hexamer is formed by two stacked trimers with opposing carbohydrate-binding faces. **(C)** Primary binding site: interactions of the AAA binding site with L-fucose: the three basic amino acid residues that interact with the axial OH on C4 are indicated with the red boxes. The interaction of the disulfide bond (Cys⁸²-Cys⁸³) with the C1-C2 bond of the L-fucose is indicated with a circle. **(D)** Extended binding site: model of the interactions between AAA and a terminally fucosylated Le^a trisaccharide. Interactions on the protein with the L-Fuc are indicated as in **(C)** above. Subterminal GlcNAc and Gal are indicated by purple circles, and the interacting amino acid residues are labeled. See text for details [adapted from Ref. (13, 14)].

binding cleft, with the subterminal units of the H1 and Le^a trisaccharides. Specifically, Glu²⁶ and His²⁷ on CDR1 can interact with hydroxyls (3-OH and 2-OH) in Gal and oxygen of the GlcNAc 2-N-Acetyl group in Le^a, or GlcNAc 6-OH and 4-OH groups in H. The OH group of Tyr⁴⁶ in CDR2 can interact with the glycosidic bond oxygen between Gal and GlcNAc moieties. Furthermore, Asp⁸¹ and Arg⁷⁹ in CDR4 interact with the GlcNAc 6-OH group in Le^a, and a water molecule can bridge the Gal 4-OH group with Asp⁸¹ in H1. The rigidity of the CDR1 loop prevents recognition of Le^x, in which the 2-N-acetyl is pointed toward the Fuc side

of the oligosaccharide (13). In contrast, MsaFBP32, an FTL that displays a shorter CDR1 loop (**Figure 3A**), would have a broader specificity for Le oligosaccharides (14).

Carbohydrate and Cation Binding Sequence Motifs

From both the initial sequence alignment of MsaFBP32 (11, 12) and analysis of the binding site structure of AAA (13) described above, highly conserved sequence motifs for carbohydrate and



calcium binding were identified in most FTL sequences available at that time. The fucose-binding sequence motif was defined as: His followed 24 residues downstream by a segment of sequence that starts with an arginine followed one residue apart by a negatively charged residue, which salt-bridges the preceding arginine and ends with a basic residue [HX₂₄RXDX₄ (R or K), where X indicates any amino acid residue]. Loops participating in the hydrophobic pocket for the fucose methyl group, such as CDR2, are also conserved in both hydrophobic profile and length. The cation binding sequence motif is h₂DGx, where h indicates a small hydrophobic amino acid residue (i.e., V, A, or I) and x stands for a small hydrophilic residue (i.e., N, D, or S). Three of the seven oxygens that bind the cation are contributed by this motif, which in AAA is located just after the 3₁₀-helix h3 (13).

Some FTLs, however, deviate from the fucose-binding sequence motif, and the changes may suggest either different

specificities or loss of sugar-recognition activity. In *Drosophila* CG9095, two amino acid residues of the basic triad are replaced by aliphatic residues, which are unlikely to establish the hydrogen bonds typical of the canonical FTLD (11, 12). Furthermore, most duplicate tandem FTLs such as MsaFBP32 possess a unique combination of sugar-binding motif in which although the triad of basic residues that interact with the sugar's axial hydroxyl on C4 is conserved in both FTLs, one domain has lost the disulfide bond from the contiguous cysteines (Cys⁸² and Cys⁸³ in AAA). Similarly, replacements of metal-coordinating residues are frequent, and those that occur at Ser⁴⁹ are of special interest, since as a bidentate ligand, it is central to the coordination geometry. In most cases, this position is substituted by residues that are able to form similar coordination bonds, such as Asp, Gln, Glu, Thr, and Tyr, but in some sequences this is not the case. In the latter, a water molecule may substitute in cation

coordination, or it is possible that the coordination geometry is modified (11, 12).

In general, sequence insertions or deletions (indels) are permissible as long as any potentially disruptive effect on the core fold is minimal, as in the FTL CDRs, where most indels are present. Interestingly, the CDR1 loop, which interacts with subterminal sugar units, shows considerable divergence suggesting that it might determine the fine specificity for a wide diversity of glycoconjugates. Coincidentally, in the MsaFBP32 gene, the exons coding the two FTLs are split by introns localized at the lower side of the barrel close to a turn that is also variable in length (in AAA: Glu¹²³-Cys¹²⁴) and would not be subject to junctional diversity during splicing (11, 12).

Tandemly Arrayed FTLDs Are Similar but Not Identical

Alignment of the amino acid sequences of the N-terminal FTL (N-FTLD) and C-terminal FTL (C-FTLD) of MsaFBP32 revealed that they are similar but not identical (11, 12). Sequence of N-FTLD is closer to AAA than the C-FTLD, suggesting that they display different carbohydrate specificity (11, 12). The structure of the MsaFBP32/L-Fuc complex revealed that the overall structure of the N-FTLD is similar to that of the C-FTLD and that recognition of L-Fuc by each FTLD is mediated by a repertoire of polar and apolar interactions similar to those observed in AAA (Figure 3). However, in both N- and C-FTLDs of MsaFBP32, the pocket for the C6 is more solvent accessible than that in AAA due to the shorter CRD1 (14). In addition, significant differences were observed between the binding sites of the MsaFBP32 N- and C-FTLDs. The C6 pocket in the C-FTLD binding site is less open than in the N-FTLD binding site due to the replacement of Phe³⁷ by the bulkier Trp¹⁸³ and the replacement of apolar contact of the S-S bridge with L-Fuc observed in AAA and N-FTLD by a bulkier Phe²²⁰ that partially displaces the sugar from the shallow binding pocket (14) (Figures 3A–C). Furthermore, an examination of the topology and surface potential of the primary and extended binding sites reveals

significant differences in the N- and C-FTLDs, specifically in the extended binding site (14) (Figure 3D), suggesting that the N-FTLD binding site recognizes more complex fucosylated oligosaccharides, with a relatively higher avidity than the C-FTLD. For example, in the N-FTLD, a methyl group of a second fucose may dock on top of Phe³⁷, but in the C-FTLD, Trp¹⁸³ closes the pocket with its indole ring, thereby interfering with the second fucose unit in Lewis tetrasaccharides (14).

FTL Isoforms and Diversity in Ligand Recognition

The presence in single individuals of multiple FTL isolectins, which display sequence replacements at positions that are critical for sugar recognition, strongly suggests diversity in carbohydrate specificity, a feature that is key not only for proteins involved in innate immunity, such as in the eel FTLs (15), but also for those that recognize heterogeneous “self” glycan ligands, as proposed for the Pacific oyster *C. gigas* bindins (24–26). It should also be kept in mind, however, that our knowledge about regulation of expression of FTLs in both immune and developmental processes is very limited at this time.

FTL Isoforms in AAA and the Japanese Eel

Although the structural analysis of the predominant sequence in the AAA and MsaFBP32 crystals revealed that the number of carbohydrate moieties specifically recognized by these lectins is limited, the expression of multiple isoforms with amino acid substitutions at key positions for sugar binding significantly broadens the range of recognized ligands (13–15). For example, variability in key sequence positions in the binding cleft and the surrounding loops in the multiple FTL isoforms expressed in the Japanese eel (*Anguilla japonica*) (Figure 4) may expand the range of glycan ligands recognized by the lectin isoform repertoire by the establishment of alternative interactions with terminal and subterminal sugar units of the oligosaccharides (15). The AAA sequence predominant in the crystal shows sequence identities with the seven FTLs from *A. japonica* ranging from 68% to 78%.

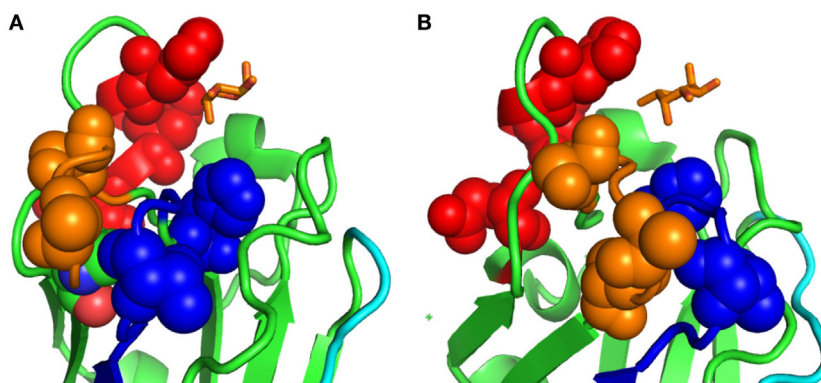


FIGURE 4 | Sequence variability in the loops surrounding the binding cleft of the *Anguilla japonica* F-type lectin (FTL) isoforms: **(A,B)** two side views of the residues that show high variability among the eFL-1-7 FTL isoforms in the Japanese eel *A. japonica*. The residues shown correspond to those present at those positions in the *Anguilla anguilla* agglutinin (AAA) structure. Sequence alignment of the AAA with eFL-1-7 isoforms shows high variability in the loops CDR1 and CDR2, with the amino acid replacements at several positions suggesting a broader specificity for some FTL isoforms.

All FTL sequences from both *A. anguilla* (AAA) and *A. japonica* conserve the basic amino acid triad that interacts with the C4 hydroxyl in fucose, showing strict conservation of His⁵² and the CDR4 sequence (13, 15). CDR1 and CDR2 conserve their size, although they present interesting sequence variations in residues associated with the hydrophobic pocket for the fucose 5-Me and oligosaccharide binding, with CDR1 showing the greatest variability. Most of the isoforms, however, conserve polar residues at the CDR1 apex, probably for interaction with the third moiety of putative oligosaccharide ligands and the two aromatic CDR2 residues in the N-terminus of h4 [like Phe⁴⁵ and Tyr⁴⁶ in AAA (13)] that form the 5-Me pocket. In the isoforms eFL-1 and eFL-5, however, the CDR1 is thinner and more flexible due to smaller residues in the apex of CDR1, thereby the 5-Me hydrophobic pocket more solvent accessible. In eFL-5, sequence replacement by smaller residues in this pocket is maximized, perhaps leading to broader specificity. Furthermore, Ser substitutions of Leu²³ and Phe⁴⁵ may result in recognition of galactose-containing oligosaccharides, by providing additional polar interactions with the 6-OH (13).

FTL Isoforms in the Pacific Oyster

In FTLs (“bindins”) of the Pacific oyster *C. gigas*, the genetic mechanisms for generating diversity in ligand recognition by lectin isoforms have been characterized in detail (24–26). Oyster bindins are gamete recognition proteins present in sperm acrosomes that bond sperm to the egg vitelline envelope during fertilization. Oyster bindins can display from one to five tandemly arrayed FTLDs. Although oyster bindins are encoded by a small number of distinct single copy genes, it appears that oysters have evolved multiple genetic mechanisms to enhance FTLD variability in sperm bindin (24). First, the FTLD repeats have diversified by positive selection at eight sites clustered on the FTLs fucose-binding pocket, similarly to the *A. japonica* isoforms (Figures 5A,B). It is noteworthy that some *C. gigas* FTL isoforms conserve the triad of basic residues (His-Arg-Arg) that in the AAA structure interact with the hydroxyl on C4 of

L-Fuc (Figure 5A), while in other isoforms these residues by other combinations that would be unable to bind fucose at the recognition site (Figure 5B). Second, increased diversity is generated by recombination in an intron that is highly variable in size and sequence located in the middle of each FTLD, to yield many different lectin domain sequences. Finally, alternative splicing in bindin cDNAs can determine the number of repeats (between one and five) per bindin mRNA (24). Interestingly, a retroposon with high homology to reverse transcriptase was identified in a three FTLD gene immediately upstream of the first FTLD repeat, suggesting that retroposition is one mechanism by which F-lectin repeats are duplicated (25, 26). In addition, the identification of a GA microsatellite in each intron, immediately upstream of the start of each FTLD exon and a downstream CT microsatellite, suggests that loopout strand hybridization can occur and that lectin repeats may replicate and transpose within the gene. It is noteworthy that neither the retrotransposon nor the CT microsatellite is present in the single FTLD containing gene (25, 26). In summary, positive selection, alternative splicing, and recombination can generate the most extraordinary intraspecific polymorphism for any known lectin, with potentially thousands of bindin variants with different numbers of FTLDs and distinct carbohydrate specificity. However, male oysters only translate one or two isoforms into protein, yielding sperm cells with potentially bindin preference for selected egg's vitelline envelopes (24–26).

Domain Organization of FTLs

The identification of a large number of proteins exhibiting the FTL sequence motif as multiple tandemly arrayed FTLDs enabled the establishment of the FTLs as a novel lectin family (11, 12) (Figure 1). In this regard, the variety of sequences identified as encoding for multiple FTLDs also illustrates the predominance of domain duplication and domain shuffling within the FTL family. Furthermore, identification in both prokaryotes and eukaryotes of mosaic FTLs displaying the FTLD in various combinations with other structurally and functionally domains

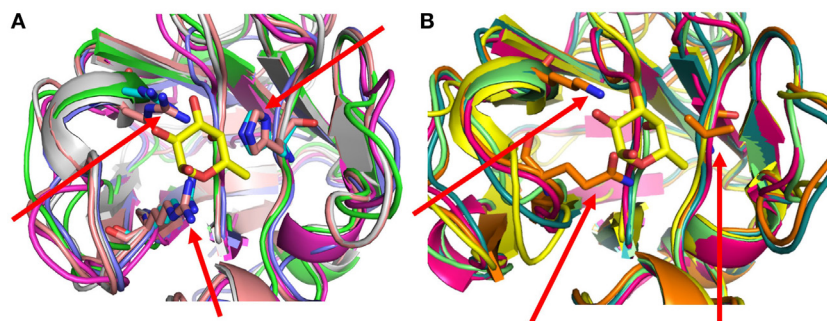


FIGURE 5 | Sequence variability in the loops surrounding the binding cleft of the Pacific oyster *Crassostrea gigas* F-type lectin (FTL) isoforms: **(A)** five *C. gigas* FTL isoforms (#1, 2, 4, 5, and 6) conserve the triad of basic residues (His-Arg-Arg) that in the *Anguilla anguilla* agglutinin structure (His⁵², Arg⁷⁹, and Arg⁸⁶) were shown to interact with the hydroxyl on C4 of L-fuc. Bound α-L-fucose is shown as a stick model in yellow, at the center of the recognition cleft. All five sequences are modeled but for clarity, only the His, Arg, and Arg side chains of two of the models are shown (red arrows). **(B)** Five *C. gigas* FTL isoforms (#3, 7, 8, 9, and 10) lack the triad of basic residues, and these are replaced by different combinations of residues. All five models are shown but only the side chains of the residues of *C. gigas* isoform 7 (Val 39, Gln 66, and Lys 69) are shown (red arrows).

suggests its extensive functional diversification in the evolution of the FTL family (11, 12). In general, taxonomically consistent domain organization of FTLs can be observed among closely related organisms, although multiple exceptions of unusual domain associations occur, which illustrate the evolutionary and ecological adaptability of this lectin family and potentially frequent lateral transfer along viral, prokaryotic, and eukaryotic lineages (11, 12, 30). It is noteworthy that while prokaryotic FTLs usually display single FTLDs in combination with diverse domains, in eukaryotes the FTLDs occur more frequently in multiple repeats, sometimes also in tandem with other domains (11, 12, 30). Among these distinct domains, carbohydrate-binding domains from other lectin families (CTLs and PXNs), complement control modules (CCP), transmembrane domains, and FA58C domains are the frequently co-occurring domains present in eukaryotic FTLs (11, 12, 30).

From the FTL sequences examined, those from *Drosophila* (CG9095 and *furrowed*) (31), sea urchin (SpCRL) (32), *S. pneumoniae* TIGR4 (11, 12), *Streptococcus mitis* (20, 21), and the amphioxus *Branchiostoma floridae* (30) represent interesting examples of polypeptides that display diverse domains in combination with FTLDs. In *Drosophila*, these domains include complement control domains (CCP), a CTLD, and a predicted transmembrane domain (12, 31). It is noteworthy that in CG9095 the CTLD is unlikely to bind carbohydrate because the canonical residues of the CRD are missing (11, 12) (Figure 1). For the sea urchin SpCRL, domains associated with the FTLD include CCP, S/T/P domain, and factor I-membrane attack complex domain (11, 12). In the *X. laevis* Xla-PXN-FBPL, another mosaic protein, a PXN domain is joined to multiple FTLDs (11, 12, 27, 28). Most interestingly, a hypothetical protein of *Microbulbifer degradans*, a microorganism capable of degrading diverse polysaccharides, has an FTLD that adjoins the structurally analogous F5/8 discoidin domain [FA58C (33)] of coagulation factors. The association of these two analogous domains is intriguing from an evolutionary perspective because they share the same fold (13) despite showing weak sequence homology. It is possible that these domains perform roles analogous to the so-called carbohydrate-binding modules present in microorganisms (34) for which similarities have already emerged (35). The considerable diversity evident from these topologies, in which the binding site motif is strictly conserved, suggests a diverse spectrum of functions fulfilled by specific recognition of L-Fuc in various environments (11, 12, 30).

Oligomeric Organization of FTL Polypeptides

Oligomerization of lectin subunits results in multivalency, a property that enables ligand cross-linking and cell agglutination and confers higher lectin avidity for clustered glycans (36). For those lectins such as FTLs that carry multiple CRDs in each polypeptide, these properties are further enhanced by the association of lectin subunits into oligomeric species (13, 14). The physiological structures of AAA are homotrimers and hexamers, which enable cooperative binding to multivalent glycans (13). Like the MBL, the three-fold cyclic symmetry of the AAA trimer would optimize the orientation and spacing of the individual FTLD binding sites for optimal binding to glycan

ligand presentation on microbial surfaces. Thus, even if the AAA and MBL recognize the same monosaccharide (in addition to mannose, MBL also binds fucose), the microbial surface glycan architecture recognized by the AAA and MBL trimers is different, as the distances between CRDs in AAA (26 Å) is almost half of that in MBL (45 Å) (13). Therefore, by recognizing different microorganisms, FTLs and CTLs would considerably expand the lectin-mediated recognition capacity in species that are endowed with both lectin types.

As described above, MsaFBP32 consists of two tandemly arrayed FTLDs, and in the native oligomer three MsaFBP32 subunits are arranged in a “tail-to-tail” manner (14) (Figure 6A). The resulting MsaFBP32 trimer of approximately 81 Å long and 60 Å wide displays two opposing globular structures, one with the three N-FTLDs and the other with the three C-FTLDs, connected by the linker peptides (14). At the opposite ends of the cylindrical trimer, the 3-CRD binding surfaces resemble the typical “bouquet” displays observed in collectins and can potentially cross-link different humoral or cell surface glycans. Although the N- and C-FTLDs are structurally similar, important differences between their binding sites suggest that the N-FTLD recognizes fucosylated oligosaccharides of higher complexity, with a relatively higher avidity than the C-FTLD (14).

Other Proteins That Display the FTL Fold

Although the novel FTL fold is distinctive of FTLs in viruses, prokaryotes, and eukaryotes, a structure-based search [DALI database (29)] identified three proteins with no significant sequence similarity to FTLs (2–14% sequence identity with AAA), but shared the jellyroll FTL fold with AAA (13). These sequences correspond to the C1 and C2 repeats of human blood coagulation factor V (37) (FVa-C1 and -C2), the C-terminal domain of a bacterial sialidase (CSlase) (38), and the NH₂-terminal domain of a fungal galactose oxidase (NGOase) (39, 40). In addition, other proteins sharing the FTL fold, but with even lower sequence similarities were identified: the human APC10/DOC1 ubiquitin ligase (PDB 1XNA) (41), the N-terminal domain of the XRCC1 single-strand DNA repair complex (PDB 1JHJ) (42), and a yeast allantoicase (PDB 1SG3) (43). An alignment of the CSlase, NGOase, FVa-Ca, and AAA sequences showed that residues equivalent to the Asp⁶⁴, Pro¹⁰⁶, and Arg¹³¹ are strictly conserved. In the four structures, the core and the bottom of the β-barrel are very similar, with the loops at the top varying in length and conformation. In CSlase and NGOase, two members (His and Arg) of the triad of basic residues that interact with the axial hydroxyl of fucose in AAA are conserved. CSlase, the galactose-binding domain of the bacterial sialidase, has been shown to bind carbohydrate (38). Furthermore, in NGOase also two residues (His⁴⁰ and Arg⁷³) homologous to those involved in carbohydrate recognition by AAA (His⁵² and Arg⁷⁹) and in CSlase (His⁵³⁹ and Arg⁵⁷²) are conserved, suggesting that it may bind carbohydrate. In FVa-C2, all residues of the basic triad related to carbohydrate binding are absent, making this pocket the most hydrophobic and the deepest. Interestingly, FVa-C2 has affinity for phospholipids instead of carbohydrates (37). Thus, these observations provide potentially useful clues either about the evolutionary history of FTLs as emerging from

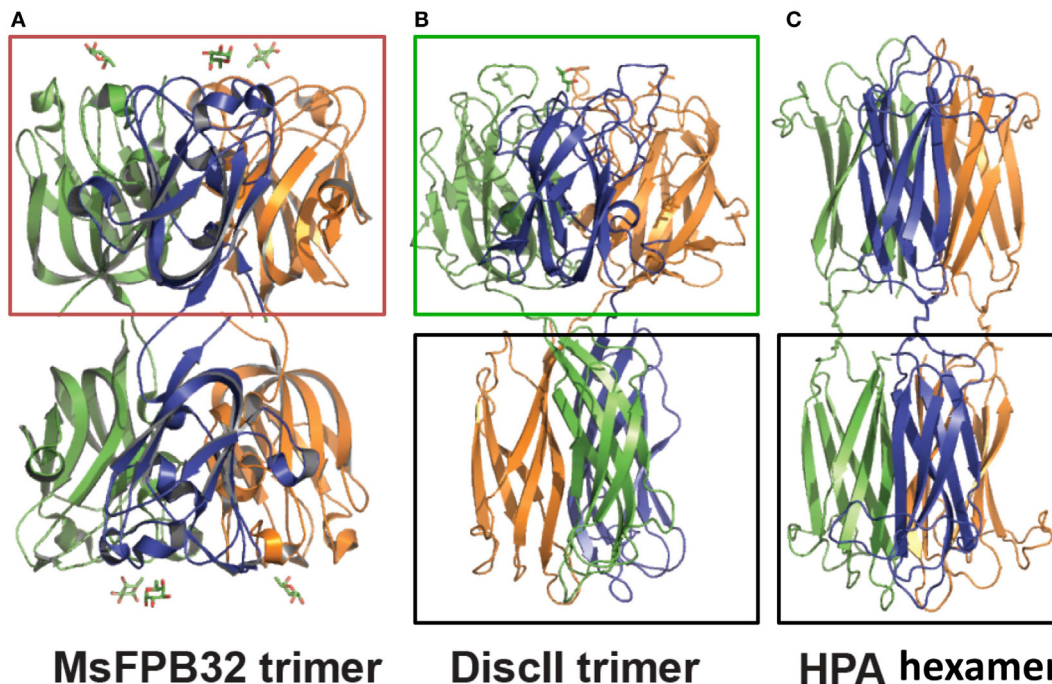


FIGURE 6 | Quaternary structure of tandemly arrayed F-type lectin domains (FTLDs) in MsaFPB32, discoidin II (Discll), and *Helix pomatia* agglutinin (HPA): Side-by-side comparison between (A) MsaFPB32 trimer, (B) Discll trimer, and (C) HPA hexamer. FTLD and discoidin domains are boxed on top (brown box in MsaFPB32 and green box in Discll); H-type lectin domains are boxed in the bottom (black box for Discll and HPA) [adapted from Ref. (14, 47)].

carbohydrate-binding domains in glycoenzymes or suggest that the recognition properties of the FTLs have been drastically modified or coopted to bind membrane phospholipids (37–40).

TAXONOMIC DISTRIBUTION AND EVOLUTIONARY ASPECTS OF THE FTLD

The initial recognition of FTLs as a novel lectin family resulted from the identification and characterization of the FTLD sequence motif in taxa ranging from prokaryotes to amphibians (11, 12) and the identification of the F-type structural fold (13). These studies identified the FTLD sequence motif in lophotrochozoan (mollusks and planaria) and ecdysozoan protostomes (horseshoe crabs and insects), deuterostome invertebrates (sea urchin), elasmobranchs (skate), lobe- and ray-finned teleost fish, and amphibians (*Xenopus* spp. and salamander) (11, 12). However, intriguing observations in these earlier studies such as the discontinuous taxonomic distribution, and diversified domain architecture of the FTL family members, frequently in combination with other structurally distinct domains, pointed to a functionally plastic FTLD, which had been specifically tailored in each lineage, subjected to lateral transfer, and that either enhanced or lost its fitness value in some taxa (11, 12). The absence of the FTL sequence motif in archaea, protozoa, urochordates, and higher vertebrates suggested that it may have been selectively lost even in relatively closely related lineages (11, 12, 30).

The advent of innovative sequencing technologies during the last decade has enabled comprehensive genomic and transcriptomic studies on a large variety of organisms and significantly expanded our knowledge about the taxonomic distribution of the FTLD from viruses to prokaryotes and eukaryotes. In this regard, a rigorous and exhaustive computational study on publicly available databases by Bishnoi et al. has recently provided significant insight and greatly expanded the range of taxa in which the FTLD is found (30). Using a three-pronged database mining approach, Bishnoi et al. identified FTLDs for the first time in viruses, fungi, reptiles, birds, and prototherian mammals (30). Furthermore, their study confirmed the diversity observed in mollusks (24–26) and revealed a substantial expansion in both FTLD occurrence and domain organization diversity in hemichordates and cephalochordates. Consistently with the aforementioned earlier reports (11, 12, 30), however, the study revealed that FTLDs appear to be absent in archaea, protozoans, urochordates, and eutherian mammals. Furthermore, no FTLDs were identified in higher plants (30).

From over 400 FTLD sequence clusters (at 80% sequence identity) tentatively identified in available databases by Bishnoi et al. (30), six FTLD sequence clusters from dsDNA viruses isolated from unicellular algae were confirmed, five from the chlorophytes *Ostreococcus* sp., *O. tauri*, and *O. lucimarinus* [*O. sp.* virus OsV5, *O. tauri* virus 1 (two distinct sequences), *O. lucimarinus* OIV1, and *O. lucimarinus* OIV6], and one from the coccolithophore *Emiliania huxleyi* (*E. huxleyi* virus 203), which are microalgal species abundant in photosynthetic phytoplankton.

Except for a viral FTLD joined by a PTX domain found in the *E. huxleyi* virus 203, all other viral FTLD sequences are single. It is noteworthy that although *E. huxleyi* and *Ostreococcus* spp. also display FTLDs, some with high similarity to the viral FTLDs, the microalgal host's FTLDs are associated with other distinct non-FTLD domains (30). The structural models of the viral FTLDs threaded on the AAA structure (13) revealed interesting features (**Figures 7A–D**). First, all viral FTLDs display the triad of basic residues (His, Arg, and Arg) that interact with the hydroxyl on C4 of L-Fuc, with the exception of *E. huxleyi* virus 203 that has only Arg-Arg. Furthermore, they all display phenylalanine instead of the disulfide bond between contiguous cysteines (Cys⁸² and Cys⁸³ in AAA) that in AAA interacts with the bond between ring atoms C1 and C2 of α -Fuc. Second, two strands of the AAA fold (AAA residues 126–136 and residues 145–155), of which the former strand (indicated by the green

arrow in **Figure 7B**) is structurally very important, are missing in the viral proteins (**Figure 7A**). It is not clear whether the FTLD structure without this strand would be stable, and it is possible that in the expressed protein the sequence corresponding to this strand might be inserted by a splicing event that was not detected in the DNA sequencing. Additionally, in the model for *E. huxleyi* virus 203 FTLD, a strand that forms the floor of the cavity of the binding site tightly overlaps with the equivalent strand in the AAA structure (**Figure 7C**). This strand, which in AAA connects the last two strands of the β -barrel, is also missing in all other viral FTLDs (**Figure 7D**). Interestingly, the viral FTLDs cluster with those from several other microalgal species (*Volvox* sp. and *Chlorella* sp.) and with several oyster (*Crassostrea* spp.) and mussel (*Mytilus* sp.) species (30). The fact that oysters and mussels are filter feeders that actively uptake phytoplankton together with any associated viruses supports the possibility

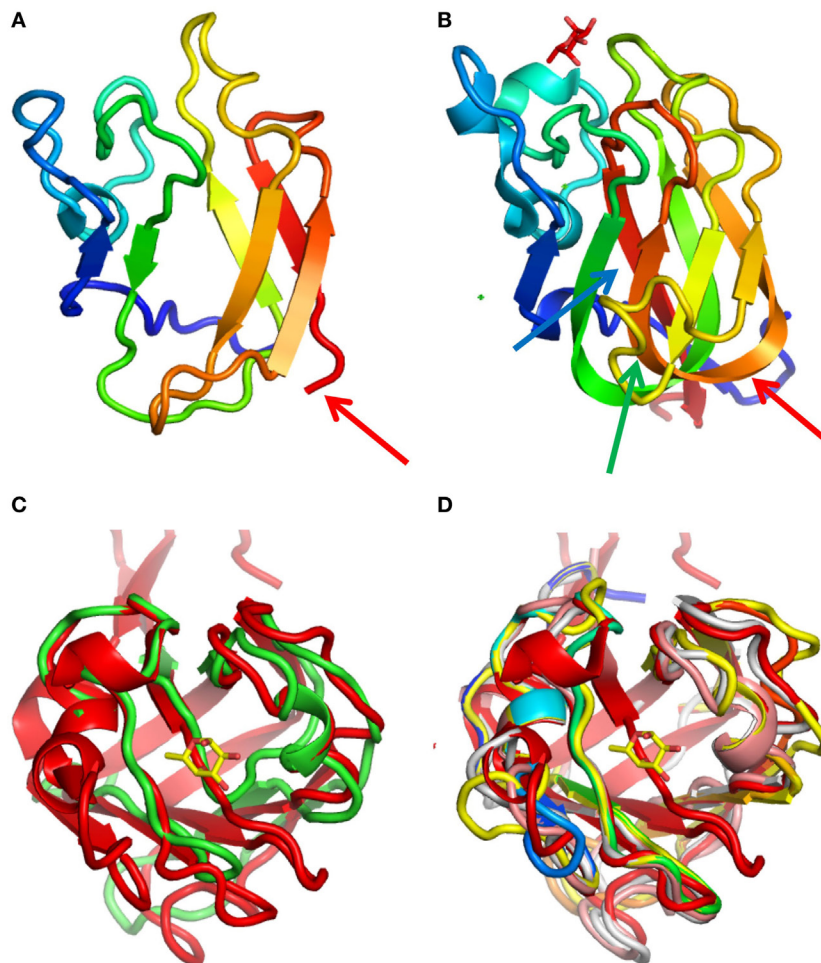


FIGURE 7 | Structural models of viral F-type lectin domains (FTLDs): (**A,B**) lateral views of the model of a viral FTLD (*Ostreococcus lucimarinus* OIV1) (**A**) and the structure of *Anguilla anguilla* agglutinin (AAA) (**B**). The traces are colored from the blue N-terminal, and the red arrows show the C-terminal of the viral protein and the position of the equivalent residue in the AAA structure. The green (AAA residues 126–136) and blue (AAA residues 145–155) arrows point to two strands of the AAA fold that are missing in the viral protein, of which the former strand (indicated by the green arrow) is structurally very important. (**C,D**) Superposition of the model of the viral FTLD from *Emiliania huxleyi* virus 203 (green) with the AAA structure (red). (**C**) *E. huxleyi* virus 203 contains a strand (indicated in green) that forms the floor of the cavity of the binding site and that overlaps with the AAA structure (red). (**D**) This strand is missing in all other viral (from *Ostreococcus* spp.) FTLDs.

of horizontal transfer between bivalves, microalgae, and their viruses. Diatoms, cryptomonads, brown algae, green algae, and fungi (*Phytophthora* spp.) also possess singly or tandemly arrayed FTLDs, mostly associated with other structurally and functionally distinct domains, but no FTLDs were identified in higher plants (30).

In prokaryotes, FTLDs were initially identified in a hypothetical protein (GenBank™ accession number ZP_00065873) from *M. degradans*, a Gram-negative bacterium with broad polysaccharide substrate degrading capability (33) and a gene (GenBank™ accession number AE007504) that is part of the L-Fuc catabolic regulon (44) in the Gram-positive *S. pneumoniae*. In the latter, three tandemly arrayed FTLDs were identified. Later studies characterized the *Streptococcus* spp. FTLDs as carbohydrate-binding domains of cholesterol-dependent cytolysins (CDCs), a large

family of pore-forming and platelet-aggregating toxins (20–23). Comparison of structural models of *S. pneumoniae* (GA41301 1.2 and 1.1) and *S. mitis* FTLDs with the structure of AAA revealed the conserved triad of basic residues (His-Arg-Arg) that in the AAA structure (His⁵², Arg⁷⁹, and Arg⁸⁶) were shown to interact with the hydroxyl on C4 to provide α -Fuc specificity. The other residues in the primary and extended carbohydrate-binding sites of AAA are not conserved among the three streptococcal FTLDs and could reflect recognition of different fucose-containing oligosaccharides (Figure 8A). However, the structure of *S. pneumoniae* FTL determined in complex with the blood group H-trisaccharide shows almost no additional interactions besides those with the α -Fuc. Furthermore, the superposition of the models shows that most of the variability resides in the loops (CDRs) that encircle the binding cleft (Figures 8B,C). This is supported

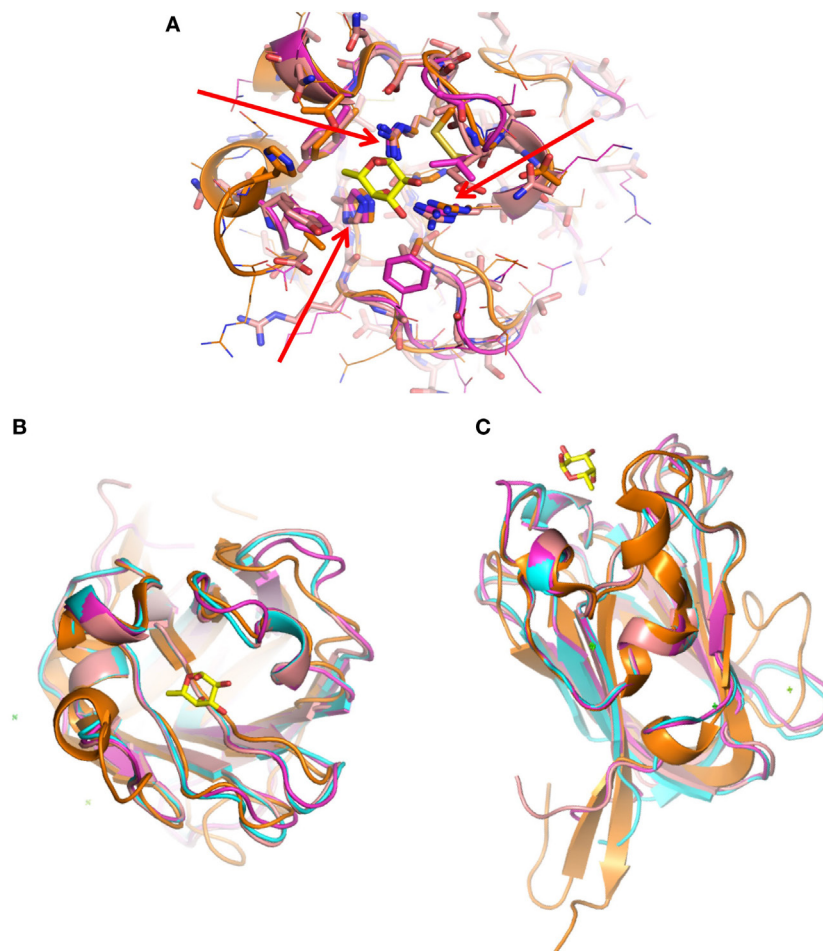


FIGURE 8 | Structural models of F-type lectin domains (FTLDs) from *Streptococcus* spp. **(A)** Superposition of structural models *Streptococcus pneumoniae* (GA41301 1.2 and 1.1) and *Streptococcus mitis* FTLDs with the structure of *Anguilla anguilla* agglutinin (AAA): the red arrows show the conserved triad of basic residues (His-Arg-Arg) in the three streptococcal FTLDs, which in the AAA structure were shown to interact with the hydroxyl on C4 of L-fuc (orange: His⁵², Arg⁷⁹, and Arg⁸⁶) to provide α -L-fucose (α -Fuc) specificity. Bound α -Fuc is shown as a stick model in yellow. The other residues in the primary and extended carbohydrate-binding sites of AAA are not conserved among the three streptococcal FTLDs and could reflect recognition of different fucose-containing oligosaccharides. However, the structure of *S. pneumoniae* FTL determined in complex with the blood group H-trisaccharide shows almost no additional interactions besides those with the α -Fuc. **(B,C)** Variability of the CDRs in FTLDs from *Streptococcus* spp.: top and lateral views of the FTLD fucose-binding site: the superposition of the models shows that most of the variability resides in the loops (CDRs) that encircle the binding cleft. AAA (orange), *S. pneumoniae* 1.2 (light blue), *S. pneumoniae* 1.1 (violet), and *S. mitis* (light purple/orchid). Bound α -Fuc is shown as a stick model in yellow.

by the electrostatic potential of the FTLD surfaces that show that the positively charged binding cleft for the α -Fuc ligand is highly conserved in all three streptococcal FTLDs as compared to AAA, but the charge characteristics of the surrounding residues in the CDRs are highly variable (**Figures 9A–D**).

Bishnoi et al. (30) recently identified FTLDs in several additional prokaryotic taxa (i.e., Actinobacteria, Bacteroidetes, Planctomycetes, Firmicutes, Proteobacteria, Cyanobacteria, Verrucomicrobia, and others) both as single or replicate FTLDs, in most cases associated with distinct sequences that included domains from other carbohydrate-binding proteins, as well as glycoenzymes, lipases, methyltransferases, and other enzymes (30). This observation is suggestive of environmental adaptations of prokaryotes for the catalytic modification of glycosylated substrates. Furthermore, the intermittent distribution of FTLDs in prokaryotic taxa suggests either their acquisition from metazoans through horizontal transfer, or less likely, that many prokaryote lineages or taxa suffered a secondary loss of the FTLD (11, 12).

The discoidins I and II (DiscI and DiscII) from the slime mold (*Dictyostelium discoideum*) are trimers of protein subunits that carry two distinct lectin domains: an N-terminal “discoidin” domain that displays the FTL fold (**Figures 6A,B**) and a C-terminal lectin domain structurally similar to the snail *Helix pomatia* lectin (HPA) H-type domains (**Figures 6B,C**) (45–47). The oligomeric organization of discoidins and *H. pomatia* lectin

strongly resembles the trimeric structure of the MsaFBP32 (47) (**Figure 6A**). Although, discoidins are reported to bind GalNAc, as expected from the presence of the H-type domain, their potential binding to fucosylated ligands has been recently analyzed in a glycan array (47).

Among the invertebrate taxa, FTLDs have also been identified in those species for which abundant genomic or transcriptomic information is available, either due to their long-standing evolutionary, ecological, or commercial interest, their use as effective model systems, or their biomedical relevance. In cnidarians such as the freshwater hydra, *Hydra vulgaris*, the FTLD is associated with a CTLD, while in the marine anemone *Nematostella vectensis*, it forms part of a complex protein that carries CCP, EGF-like, and other distinct domains (30). FTLDs were also identified in worms, including the nematode *Caenorhabditis elegans* and the annelid *Capitella teleta* (30). In arthropods, FTLDs have been identified either associated with multiple non-FTL domains or as standalone domains. For example, among chelicerates, FTLDs are present in the tachylectin from the horseshoe crab *Tachyleus tridentatus* (11, 12, 16, 30), and in the tick *Ixodes scapularis* (30), while in crustaceans FTLs were found in the prawn *Macrobrachium rosenbergii* (30). As discussed above, FTLDs were initially identified in insects as the *furrowed* gene from the fruit fly *Drosophila*, and in the mosquito, *Anopheles gambiae* (GenBank accession #

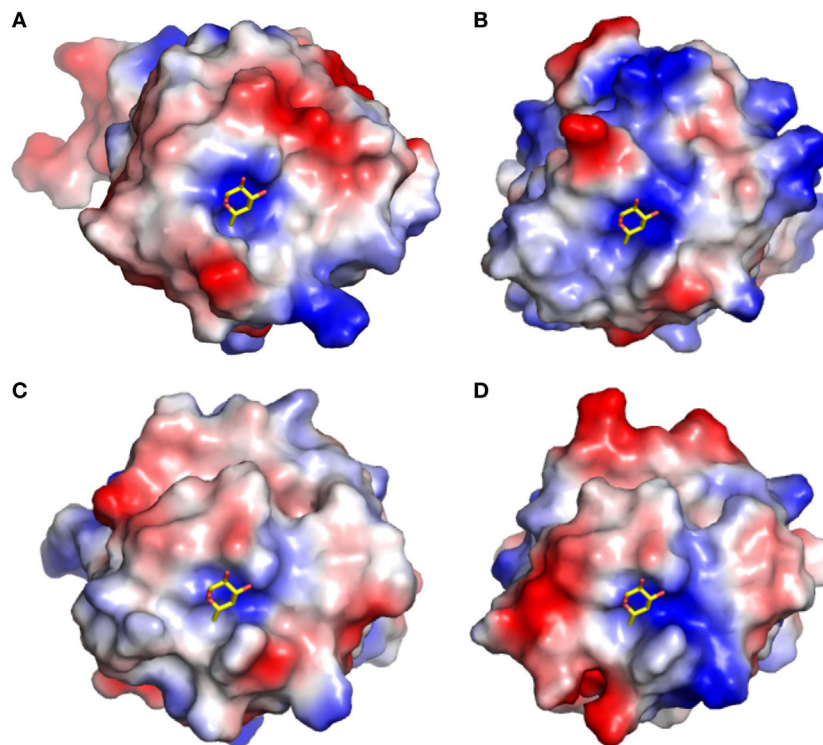


FIGURE 9 | Surface electrostatic potential surfaces of F-type lectin domains (FTLDs) from *Streptococcus* spp.: **(A)** *Anguilla anguilla* agglutinin (AAA) structure, **(B)** *Streptococcus mitis*, **(C)** *Streptococcus pneumoniae* 1.1, and **(D)** *S. pneumoniae* 1.2. The electrostatic surfaces show that the positively charged binding cleft for the α -L-fucose ligand is highly conserved in all three streptococcal FTLDs as compared to AAA, but the charge characteristics of the surrounding residues in the CDRs are highly variable.

AAAB01008846 and AAAB01008811) (11, 12). In mollusks, FTLDs have been well characterized in the highly diversified oyster bindings discussed above (24–26), as well as other oyster species, mussels, and clams (30, 48–52). Similarly, an extraordinary expansion in prevalence and organizational complexity of the FTLD were noted among protochordates, specifically in the hemichordates (acorn worms) and cephalochordates (over 70 different FTLD sequence clusters in the amphioxus *B. floridae*) (30), but surprisingly, they appear to be absent in urochordates (i.e., ascidians and salps) (11, 12, 30).

The initial studies revealed that the substantially diverse FTLD organizational topologies in cold-blooded vertebrates, such as fish and amphibians, appear to be in some cases lineage-related (11, 12). As the F-type fold displays joined N- and C-terminals, this structural feature promotes the assembly of multiple CRD topologies that are consistent with phylogenetic clustering. In this regard, the binary FTLs have diversified through lineage-dependent gene duplications that are unique to teleosts and amphibians (11, 12). For example, most teleost FTLs contain either two or four tandemly arrayed FTLDs, whereas in *Xenopus* spp. FTLs are organized from single FTLDs to combinations of two, three, or four FTLDs and as chimeric proteins containing five tandemly arrayed FTLDs adjacent to a PXN domain (11, 12). The study by Bishnoi et al. (30) identified additional CTLD and the clotting factor FA58C domain associated with FTLDs in teleost fish, including the coelacanth *Latimeria* sp., and for the first time identified FTLDs in reptiles, birds, and mammals, in the latter associated with PXN domains (30). Interestingly, FTLDs were only identified in prototherian mammals, including the monotremes, such as platypus, and didelphid marsupials, such as the opossum (30), but appear to be absent in eutherian (placental) mammals (11, 12, 30).

FUNCTIONAL ASPECTS

In spite of the broad range of taxonomic distribution of FTLDs, their functional properties have only been experimentally demonstrated in a limited number of examples. In most cases, their biological roles have been rather inferred from their gene expression levels and cell- or tissue-specific localization upon experimental immune challenge or environmental stressors, together with their structural features, biochemical properties, including their binding selectivity for endogenous and microbial glycosylated ligands. In those few examples in which FTLDs have been studied in genetically tractable model systems, such as the streptococcal lectinolysins, their roles have been rigorously established not only by genetic approaches but also by significant contributions of the rigorous analysis of their structures (20–23). In *Drosophila*, however, although the role of the *furrowed* gene in cell adhesion was clearly established, the specific function of the FTLD in this process remains to be elucidated (11, 12, 31). In the slime mold *D. discoideum*, a widely recognized genetically tractable model system for developmental and cell biology studies, the role(s) of DiscI and DiscII remains to be rigorously established (47). Although initially both discoidins were reported as secreted lectins involved in cell–substratum adhesion and spore coat formation (53, 54),

later studies questioned these results as no evidence of their secretion could be found. A recent study, however, concluded that DiscI is implicated in cell–substratum adhesion and plays a role in streaming (55) although the mechanistic aspects have not been elucidated yet.

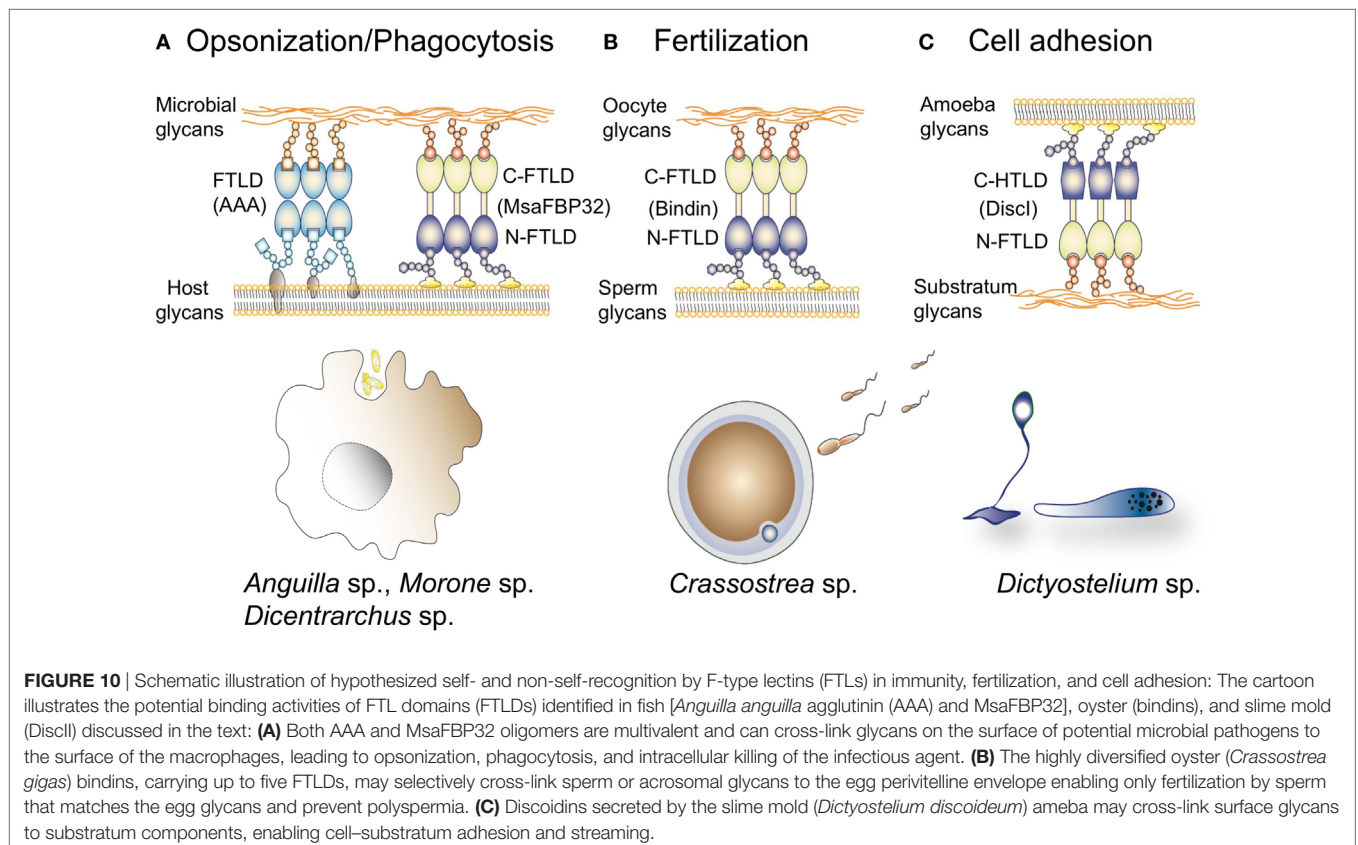
Initially identified and characterized in teleost fish, the multivalent FTLD display and their distinct carbohydrate specificity revealed the clear potential of oligomeric FTLs for binding to microbial surface glycans (13, 14). For example, both the trimeric arrangement of the AAA FTLDs and the opposite orientation of the distinct N- and C-terminal binding surfaces of the trimeric MsaFBP32 strongly suggest that in circulation these lectins can cross-link fucosylated glycoconjugates displayed on different cells (13, 14). Modeling of the MsaFBP32 recognition of fucosylated oligosaccharides from prokaryotes and eukaryotes supports the observation that FTLs with binary tandem CRDs can function as opsonins (14). Opsonization of potential pathogens would take place by FTL-mediated cross-linking exposed carbohydrate moieties on microbial pathogens with surface glycans on the host's phagocytic cells (14). By recognizing Le^a-containing glycans on the phagocytic cell surface *via* the N-CRD, MsaFBP32 would cross-link the infectious agent *via* the C-CRD, which recognizes glycans α -linked L-Fuc, 2-acetoamido L-Fuc, 3-deoxy-L-fucose (colitose) or L-Rha (6-deoxy-L-mannose, present in *Escherichia coli* glycans) as non-reducing terminal residues on the microbial surface (14). The tissue expression of fish FTLs that primarily takes place not only in liver (11, 12, 56–60), the typical source of acute phase reactants, but also in gills (15, 56–58) and intestine (56–58), which are organs continuously exposed to infectious challenge, is highly suggestive of their role(s) in innate immune defense. The opsonic properties of FTLs were experimentally demonstrated with the binary tandem FTLs from sea bass (DIFBL; *Dicentrarchus labrax*) and gilt head bream (SauFBL; *Sparus aurata*) (58, 60). Pre-exposure of *E. coli* to DIFBL or SauFBL significantly increases their uptake by peritoneal macrophages as compared to the unexposed bacteria (58, 60) supporting the concept that F-lectins with multivalent FTLs such as AAA, DIFBL, SauFBL, and MsaFBP32 can function as opsonins that promote phagocytosis of microbial pathogens. By transfecting the EPC cell line with an FTL (RbFTL-3) that is highly expressed in the intestine of rock bream (*Oplegnathus fasciatus*), followed by viral (viral hemorrhagic septicemia virus) challenge, Cho et al. (61) recently showed that RbFTL-3 controls viral budding and increases the viability of VHSV infected cells, suggesting that the lectin limits hemorrhage in fish tissues.

Upregulation of FTL expression by immune challenge as it would be expected by analogy to liver expression of acute phase reactants in innate immune responses, however, has not been the general rule for the species examined. For MsaFBP32, an inflammatory challenge only increased the liver transcript levels in about three-fold over the relatively high basal expression levels (11, 12), whereas for DIFBL protein levels were modestly enhanced by *Vibrio alginolyticus* infectious challenge (58). In the Japanese sea perch (*Lateolabrax japonicus*), the FTL JspFL was only upregulated in spleen, while it was also constitutively expressed in liver and gills (62). In contrast, LPS challenge significantly upregulated

expression and increased secretion of FTLs in liver and gill tissue from *A. japonica* (15).

As FTLs have not only been identified in the eukaryotic hosts but also in viral, prokaryotic, and multicellular pathogens and parasites, the intriguing possibility that FTLs may play key roles in microbial virulence has only been examined in detail in bacterial lectinolysins (17–23, 63). It is widely recognized that opportunistic bacteria recognize and attach to host cell glycans *via* carbohydrate-binding domains in their surface proteins (64, 65). However, Gram-positive bacteria (18) such as *Streptococcus* spp. (*S. pneumoniae*, *S. mitis*, and *S. intermedius*) and *Garnderella vaginalis*, among others, produce CDCs (lectinolysin, pneumolysin, intermedilysin, and vaginolysin) that bind to and disrupt the host cell membrane (17–23). The *S. mitis* lectinolysin, also described as a platelet aggregation factor (17), carries an FTLD that recognizes the host's fucosylated moieties to significantly enhance their virulent pore-forming properties in at least one order of magnitude. Upon binding to the host surface glycans, monomeric CDCs spontaneously self-assemble to form large β barrel pores that lead to cell lysis (63). The FTLD of the CDC specifically recognizes difucosylated glycans [Lewis γ (Le^{γ}) and Lewis b (Le^b) moieties], and it has been controversial whether the fucose-binding site remains masked in the CDC monomer and is only exposed following contact with the cell surface (21), or if it is fully accessible to the environment and ready for interaction with host cell glycoreceptors (22).

In contrast with the innate immune host defense and the bacterial virulence functions of the FTLDs described above, the sperm “bindins” from the Pacific oyster (*C. gigas*), discussed in a previous section, are highly polymorphic proteins stored in the acrosomal rings of sperm cells that bind to the surface of the egg perivitelline envelope during fertilization (24). By mechanisms of positive selection, recombination, and alternative splicing, a single copy bindin gene can produce transcripts that are highly diversified both in sequence and domain organization within and among individuals in this oyster species. Interestingly, each individual male oyster will translate only one or two polymorphic bindins carrying between one and five tandemly arrayed F-lectin domains are translated in Ref. (25). The unusual high intraspecific diversity of the oyster bindin F-lectins has been proposed to represent coevolution of sperm gamete recognition mechanisms to “catch-up” with the high diversification of egg receptors aimed at avoiding polyspermy (26). It should be noted, however, that FTLs have also been reported as defense molecules in oysters (48–52) and several other invertebrate species. Among these, an FTL (PmF-lectin) from pearl oyster (*Pinctada martensii*) is highly expressed in hemocytes and gill and significantly upregulated (13-fold) by infectious challenge (*V. alginolyticus*), suggesting that PmF-lectin is involved in the innate immune response (48). The highly diversified FTL repertoire identified in the common periwinkle (*Littorina littorea*) has been hypothesized as an immune defense system (52), whereas in the blunt-gaper clam *Mya truncata*, FTLs have been identified in both the shell



matrix and mantle tissue proteins, suggesting that during the shell biomineralization process, immune defense functions may be carried out by proteins secreted by the mantle, which are later incorporated into the shell matrix (51).

CONCLUSION

The structural and functional analyses of the FTLD, together with its distribution in extant viral, prokaryotic, and eukaryotic species reveal an intriguing evolutionary history of this lectin domain with key adaptations to a diverse array of functions carried out by the FTLD itself, either as single units or as tandemly arrayed domains. This functional diversity is further expanded for FTLDs associated with structurally and functionally distinct associated domains, either belonging to other lectin families (CTLs and PXNs), enzymes, or other proteins. Thus, FTLs are essentially pleiotropic and can orchestrate a vast array of functions based on “self” and “non-self”-recognition that encompass not only innate immunity but also fertilization, cell adhesion, and microbial virulence, among others yet to be unraveled (**Figure 10**). In recent years, a substantial body of evidence has supported the proposal that along their evolution, selected FTLs were co-opted to carry out different functions that may not rely on active carbohydrate-binding sites, and therefore, this property, which is inherent to their definition as lectins, may have been lost in the process. The paucity in the taxonomic distribution of viral and prokaryotic FTLDs suggests the eukaryotic origin of the domain, followed by extensive duplication and mutation, lateral transfer, and secondary loss or cooption (11, 12). This is supported by the observation that a phylogenetic analysis revealed that although in general the clustering of FTLDs is consistent with the taxonomical categories, bacterial FTLDs are interspersed with several eukaryotic FTLDs (30). Furthermore, the viral FTLDs cluster with those from several other microalgal species (*Volvox* sp. and *Chlorella* sp.) and with several oyster (*Crassostrea* spp.) and mussel (*Mytilus* sp.) species (30). In this regard, it is important to note that oysters and mussels are filter-feeder bivalves that actively uptake microalgae (together with their associated viruses) from the suspended phytoplankton, thereby providing clues about the origins and potential lateral transfer of the viral, microalgal, and mollusk FTLDs.

On the other hand, despite the FTL diversity evident in amphibians, reptiles, birds, and prototherian mammals, no *bona fide* FTL homologs are detectable in genomes of eutherian mammals (11, 12, 30). Therefore, above the level of the prototherian mammals this lectin family may have been lost as such, either by becoming truly extinct or by being co-opted into other functions as proposed for the C-1 and C-2 domains of the clotting factors V and VIII (11–13). While lacking carbohydrate binding capacity due to the loss of the triad of basic residues that interact with the axial hydroxyl of the sugar ligand, the aforementioned C-1 and C-2 domains still display the F-type fold and are highly prevalent not only in taxa ranging from fish to birds but also widespread in eutherian mammals (11–13). It is possible that the loss of fucose recognition activity has been driven by the need to avoid self-reactivity to fucosylated moieties exposed on the cell

surface, such as the blood group H and Lewis oligosaccharides that arose along the eutherian mammal lineages.

With regards to their roles in immune recognition, as described above FTLs can display in a single polypeptide monomer single or tandemly arrayed CRDs of similar but distinct specificity. Therefore, cross-linking of “self” and “non-self” carbohydrate moieties can be easily rationalized by: (a) the different specificity of their binding sites, (b) the distinct architecture of the presentation and multivalency of the carbohydrate ligands on the microbial cell surface or the host, and (c) the biophysical properties of the microenvironment where the interactions occur (11–14). Bishnoi et al. identified a substantial expansion in both FTLD occurrence and domain organization diversity in the mollusks, hemichordates, and cephalochordates that was attributed to enhanced emphasis on innate immunity in these taxa (30). Consistently with earlier studies (11, 12), however, the study revealed that FTLDs are absent in urochordates (ascidians and salps) (30). First of all, the FTL diversification observed in mollusks is most likely due at least in part to their expanded functions as gamete recognition molecules in fertilization processes (“bindins,” described above) (24–26). Second, it is well established that urochordates, like hemichordates and cephalochordates, also lack *bona fide* adaptive immune systems such as the variable lymphocyte receptors (VLRs) and immunoglobulin- and B/T cell-mediated immune responses and solely rely on innate immunity for defense against infection. Thus, the increased FTLD diversification in hemichordates and cephalochordates, together with the lack of FTLDs in urochordates could be rather attributed to compensatory effects among multiple lectin families, depending on selective advantage(s) that each can provide to any given taxa as more or less effective pattern recognition receptors in innate immunity. In support of this view, it is noteworthy that the urochordata ascidian *Clavelina picta*, which lacks FTLs, expresses a highly diversified repertoire of fucose-binding CTLs, suggesting that the expansion of the CTL repertoire probably reflects the selective advantage that fucose-binding CTLs provides over FTLs to the ascidian’s innate immune responses (66, 67). In addition to functions carried out by FTLs, such as pathogen recognition, immobilization, and opsonization, CTLs can also initiate complement activation, an ancient enzyme-driven mechanism that can rapidly amplify opsonization and effect direct killing of the potential pathogen *via* the membrane attack complex (66, 67). Therefore, it is possible that these and other functional advantages offered by CTLs led to their expansion as innate immune defense mechanisms in higher mammals, simultaneously with the contraction, cooption, or loss altogether of the FTL family members.

The rapidly expanding genomic databases and their increasing availability for numerous animal species have provided further insight into the structural and functional diversification of lectin repertoires from prokaryotes, invertebrates, protochordates, and vertebrates. In this context, the recent identification of novel lectin families such as the FTLs (11–14), underscores the need of more research in non-mammalian model organisms. This will provide greater insight into the structural, functional, and evolutionary aspects of lectin families that may not be as obvious in the traditional mammalian model systems. In this regard, the

structural analysis of multiple FTL isoforms in eels and oysters (15, 24–26) has revealed substantial diversity in oligosaccharide recognition and has provided conceptually transformative insight into the processes through which lectins can generate an extraordinary structural and, most likely, functional diversity for self/non-self-recognition that resembles those mechanisms operative in adaptive immunity of higher vertebrates. The current exponential increase in the genome, transcriptome, and proteome information on additional non-mammalian model organisms, coupled with structural studies and innovative forward and reverse genetic approaches for functional analyses has the potential to uncover novel structural, functional, and evolutionary features in various lectin families, from viruses and prokaryotes to mammals. Furthermore, homology modeling of novel FTLs on related crystal structures will contribute to rapidly expand our knowledge about their interactions with potential glycosylated ligands (68). Due to its substantial advantages over mammalian models, namely external fertilization, transparent embryos, a continuously expanding collection of mutations and a rapidly growing toolbox for manipulation of gene expression, the zebrafish may constitute an ideal model for

the elucidation of the biological roles of FTLs in innate immunity of vertebrates. Finally, given the prevalence of fucosylated moieties on the surface of neoplastic cells, it is possible that FTLs may become useful reagents for both diagnostics and therapeutic applications in cancer (69, 70).

AUTHOR CONTRIBUTIONS

GV designed, drafted, and edited the final manuscript; LA and MB developed and analyzed structural models, evaluated and edited the draft manuscript; and MC, CF, and KS evaluated and edited the draft manuscript.

FUNDING

The author's research reviewed herein was supported by Grants IOS 1050518, IOB-0618409, MCB 0077928, and IOS-0822257 from the National Science Foundation, Grant R01GM070589 from the National Institutes of Health (GV); grant ARRA-1RO1NS061827 from the NIH (LA); and NIGMS pre-doctoral fellowship GM14903-04 from the NIH to Eric W. Odom.

REFERENCES

- McGreal EP, Martinez-Pomares L, Gordon S. Divergent roles for C-type lectins expressed by cells of the innate immune system. *Mol Immunol* (2004) 41:1109–21. doi:10.1016/j.molimm.2004.06.013
- Holmskov U, Thiel S, Jensenius JC. Collectins and ficolins: humoral lectins of the innate immune defense. *Annu Rev Immunol* (2003) 21:547–78. doi:10.1146/annurev.immunol.21.120601.140954
- Vasta GR, Ahmed H, Odom EW. Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. *Curr Opin Struct Biol* (2004) 14:617–30. doi:10.1016/j.sbi.2004.09.008
- Fujita T, Matsushita M, Endo Y. The lectin-complement pathway – its role in innate immunity and evolution. *Immunol Rev* (2004) 198:185–202. doi:10.1111/j.0105-2896.2004.0123.x
- Vasta GR, Ahmed H. *Animal Lectins: A Functional View*, Taylor & Francis. Boca Raton: CRC Press (2008). 558 p.
- Weis WI, Taylor ME, Drickamer K. The C-type lectin superfamily in the immune system. *Immunol Rev* (1998) 163:19–34. doi:10.1111/j.1600-065X.1998.tb01185.x
- Kim MS, Byun M, Oh BH. Crystal structure of peptidoglycan recognition protein LB from *Drosophila melanogaster*. *Nat Immun* (2003) 4:787–93. doi:10.1038/ni952
- Endo Y, Liu Y, Fujita T. Structure and function of ficolins. *Adv Exp Med Biol* (2006) 586:265–79. doi:10.1007/0-387-34134-X_18
- Garlanda C, Hirsch E, Bozza S, Salustri A, De Acetis M, Nota R, et al. Non-redundant role of the long pentraxin PTX3 in anti-fungal innate immune response. *Nature* (2002) 420:182–6. doi:10.1038/nature01195
- Vasta GR. Roles of galectins in infection. *Nat Rev Microbiol* (2009) 7:424–38. doi:10.1038/nrmicro2146
- Odom EW. *F-Type Lectins: Biochemical, Genetic and Structural Characterization of a Novel Lectin Family in Lower Vertebrates* [Ph.D. Thesis]. Baltimore, MD: University of Maryland, MEES Program (2004).
- Odom EW, Vasta GR. Characterization of a binary tandem domain F-type lectin from striped bass (*Morone saxatilis*). *J Biol Chem* (2006) 281:1698–713. doi:10.1074/jbc.M507652200
- Bianchet MA, Odom EW, Vasta GR, Amzel LM. A novel fucose recognition fold involved in innate immunity. *Nat Struct Biol* (2002) 9:628–34. doi:10.1038/nsb817
- Bianchet MA, Odom EW, Vasta GR, Amzel LM. Structure and specificity of a binary tandem domain F-lectin from striped bass (*Morone saxatilis*). *J Mol Biol* (2010) 401:239–52. doi:10.1016/j.jmb.2010.06.018
- Honda S, Kashiwagi M, Miyamoto K, Takei Y, Hirose S. Multiplicity, structures, and endocrine and exocrine natures of eel fucose-binding lectins. *J Biol Chem* (2000) 275:33151–7. doi:10.1074/jbc.M002337200
- Saito T, Hatada M, Iwanaga S, Kawabata S. A newly identified horseshoe crab lectin with binding specificity to O-antigen of bacterial lipopolysaccharides. *J Biol Chem* (1997) 272:30703–8. doi:10.1074/jbc.272.49.30703
- Ohkuni H, Todome Y, Okibayashi F, Watanabe Y, Ohtani N, Ishikawa T, et al. Purification and partial characterization of a novel human platelet aggregation factor in the extracellular products of *Streptococcus mitis*, strain Nm-65. *FEMS Immunol Med Microbiol* (1997) 17:121–9. doi:10.1111/j.1574-695X.1997.tb01004.x
- Tweten RK. The cholesterol-dependent cytolysins: a family of versatile pore-forming toxins. *Infect Immun* (2005) 73:6199–209. doi:10.1128/IAI.73.10.6199-6209.2005
- Tilley SJ, Orlova EV, Gilbert RJC, Andrew PW, Saibil HR. Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* (2005) 121:247–56. doi:10.1016/j.cell.2005.02.033
- Boraston AB, Wang D, Burke RD. Blood group antigen recognition by a *Streptococcus pneumoniae* virulence factor. *J Biol Chem* (2006) 281:35263–71. doi:10.1074/jbc.M607620200
- Farrand S, Hotze E, Friese P, Hollingshead SK, Smith DF, Cummings RD, et al. Characterization of a streptococcal cholesterol dependent cytolysin with a Lewis y and b specific lectin domain. *Biochemistry* (2008) 47:7097–107. doi:10.1021/bi8005835
- Feil SC, Lawrence S, Mulhern TD, Holien JK, Hotze EM, Farrand S, et al. Structure of the lectin regulatory domain of the cholesterol-dependent cytolysin lectinolysin reveals the basis for its lewis antigen specificity. *Structure* (2012) 20(2):248–58. doi:10.1016/j.str.2011.11.017
- Giddings KS, Zhao J, Sims PJ, Tweten RK. Human CD59 is a receptor for the cholesterol-dependent cytolysin intermedilysin. *Nat Struct Mol Biol* (2004) 12:1173–8. doi:10.1038/nsmb862
- Moy GW, Springer SA, Adams SL, Swanson WJ, Vacquier VD. Extraordinary intraspecific diversity in oyster sperm bindin. *Proc Natl Acad Sci U S A* (2008) 105:1993–8. doi:10.1073/pnas.0711862105
- Moy GW, Vacquier VD. Bindin genes of the Pacific oyster *Crassostrea gigas*. *Gene* (2008) 423:215–20. doi:10.1016/j.gene.2008.07.005
- Springer SA, Moy GW, Friend DS, Swanson WJ, Vacquier VD. Oyster sperm bindin is a combinatorial fucose lectin with remarkable intra-species diversity. *Int J Dev Biol* (2008) 52:759–68. doi:10.1387/ijdb.082581ss

27. Seery LT, Schoenberg DR, Barbaux S, Sharp PM, Whitehead AS. Identification of a novel member of the pentraxin family in *Xenopus laevis*. *Proc Biol Sci* (1993) 253(1338):263–70. doi:10.1098/rspb.1993.0112
28. Tennent GA, Pepys MB. Glycobiology of the pentraxins. *Biochem Soc Trans* (1994) 22(1):74–9. doi:10.1042/bst0220074
29. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* (1993) 233:123–38. doi:10.1006/jmbi.1993.1489
30. Bishnoi R, Khatri I, Subramanian S, Ramya TNC. Prevalence of the F-type lectin domain. *Glycobiology* (2015) 25(8):888–901. doi:10.1093/glycob/cwv029
31. Leshko-Lindsay L, Corces VG. The role of selectins in *Drosophila* eye and bristle development. *Development* (1997) 124:169–80.
32. Multerer KA, Smith LC. Two cDNAs from the purple sea urchin, *Strongylocentrotus purpuratus*, encoding mosaic proteins with domains found in factor H, factor I, and complement components C6 and C7. *Immunogenetics* (2004) 56(2):89–106. doi:10.1007/s00251-004-0665-2
33. Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci* (1998) 7:1626–31. doi:10.1002/pro.5560070717
34. Warren RAJ. Microbial hydrolysis of polysaccharides. *Annu Rev Microbiol* (1996) 50:183–212. doi:10.1146/annurev.micro.50.1.183
35. Boraston AB, Notenboom V, Warren RAJ, Kilburn DG, Rose DR, Davies G. Structure and ligand binding of carbohydrate-binding module CsCBM6-3 reveals similarities with fucose-specific lectins and “galactose-binding” domains. *J Mol Biol* (2003) 327:659–69. doi:10.1016/S0022-2836(03)00152-9
36. Dam TK, Brewer CF. Effects of clustered epitopes in multivalent ligand receptor interactions. *Biochemistry* (2008) 47:8470–6. doi:10.1021/bi801208b
37. Macedo-Ribeiro S, Bode W, Huber R, Quinn-Allen MA, Kim SW, Ortel TL, et al. Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature* (1999) 402(6760):434–9. doi:10.1038/46594
38. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* (1995) 3:1197–205. doi:10.1016/S0969-2126(01)00255-6
39. Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen JN, et al. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* (1991) 350:87–90. doi:10.1038/350087a0
40. Firbank SJ, Rogers MS, Wilmot CM, Dooley DM, Halcrow MA, Knowles PF, et al. From the cover: crystal structure of the precursor of galactose oxidase: an unusual self-processing enzyme. *Proc Natl Acad Sci U S A* (2001) 98:12932–7. doi:10.1073/pnas.231463798
41. Wendt KS, Vodermaier HC, Jacob U, Gieffers C, Gmachl M, Peters JM, et al. Crystal structure of the APC10/DOC1 subunit of the human anaphase-promoting complex. *Nat Struct Biol* (2001) 8:784–8. doi:10.1038/nsb0901-784
42. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Solution structure of the single-strand break repair protein xrc1 n-terminal domain. *Nat Struct Biol* (1999) 6:884–93. doi:10.1038/12347
43. Leulliot N, Quevillon-Cheruel S, Sorel I, Graille M, Meyer P, Liger D, et al. Crystal structure of yeast allantoicase reveals a repeated jelly-roll motif. *J Biol Chem* (2004) 279:23447–52. doi:10.1074/jbc.M401336200
44. Chan PF, O'Dwyer KM, Palmer LM, Ambrad JD, Ingraham KA, So C, et al. Characterization of a novel fucose-regulated promoter (PfcK) suitable for gene essentiality and antibacterial mode-of-action studies in *Streptococcus pneumoniae*. *J Bacteriol* (2003) 185:2051–8. doi:10.1128/JB.185.6.2051-2058.2003
45. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, et al. Biochemical and structural analysis of *Helix pomatia* agglutinin (HPA): a hexameric lectin with a novel fold. *J Biol Chem* (2006) 281:20171–80. doi:10.1074/jbc.M603452200
46. Aragão KS, Satre M, Imbert A, Varrot A. Structure determination of discoidin II from *Dictyostelium discoideum* and carbohydrate binding properties of the lectin domain. *Proteins* (2008) 73(1):43–52. doi:10.1002/prot.22038
47. Mathieu SV, Aragão KS, Imbert A, Varrot A. Discoidin I from *Dictyostelium discoideum* and interactions with oligosaccharides: specificity, affinity, crystal structures, and comparison with discoidin II. *J Mol Biol* (2010) 400(3):540–54. doi:10.1016/j.jmb.2010.05.042
48. Chen J, Xiao S, Yu Z. F-type lectin involved in defense against bacterial infection in the pearl oyster (*Pinctada martensii*). *Fish Shellfish Immunol* (2011) 30(2):750–4. doi:10.1016/j.fsi.2010.12.025
49. Wu Q, Li L, Zhang G. *Crassostrea angulata* bindin gene and the divergence of fucose-binding lectin repeats among three species of *Crassostrea*. *Mar Biotechnol* (NY) (2011) 13(2):327–35. doi:10.1007/s10126-010-9304-0
50. Anju A, Jeswin J, Thomas PC, Vijayan KK. Molecular cloning, characterization and expression analysis of F-type lectin from pearl oyster *Pinctada fucata*. *Fish Shellfish Immunol* (2013) 35(1):170–4. doi:10.1016/j.fsi.2013.03.359
51. Arivalagan J, Marie B, Sleight VA, Clark MS, Berland S, Marie A. Shell matrix proteins of the clam, *Mya truncata*: roles beyond shell formation through proteomic study. *Mar Genomics* (2016) 27:69–74. doi:10.1016/j.margen.2016.03.005
52. Gorbushin AM, Borisova EA. Lectin-like molecules in transcriptome of *Littorina littorea* hemocytes. *Dev Comp Immunol* (2015) 48(1):210–20. doi:10.1016/j.dci.2014.10.007
53. Barondes SH, Cooper DN, Haywood-Reid PL. Discoidin I and discoidin II are localized differently in developing *Dictyostelium discoideum*. *J Cell Biol* (1983) 96:291–6. doi:10.1083/jcb.96.1.291
54. Barondes SH, Haywood-Reid PL, Cooper DN. Discoidin I, an endogenous lectin, is externalized from *Dictyostelium discoideum* in multilamellar bodies. *J Cell Biol* (1985) 100:1825–33. doi:10.1083/jcb.100.6.1825
55. Bastounis E, Álvarez-González B, del Álamo JC, Lasheras JC, Firtel RA. Cooperative cell motility during tandem locomotion of amoeboid cells. *Mol Biol Cell* (2016) 27(8):1262–71. doi:10.1091/mbc.E15-12-0836
56. Cammarata M, Vazzana M, Chinnici C, Parrinello N. A serum fucoselectin isolated and characterized from sea bass *Dicentrarchus labrax*. *Biochim Biophys Acta* (2001) 1528:196–202. doi:10.1016/S0304-4165(01)00193-3
57. Cammarata M, Benenati G, Odom EW, Salerno G, Vizzini A, Vasta GR, et al. Isolation and characterization of a fish F-type lectin from gilt head bream (*Sparus aurata*) serum. *Biochim Biophys Acta* (2007) 1770:150–5. doi:10.1016/j.bbagen.2006.09.015
58. Salerno G, Parisi MG, Parrinello D, Benenati G, Vizzini A, Vazzana M, et al. F-type lectin from the sea bass (*Dicentrarchus labrax*): purification, cDNA cloning, tissue expression and localization, and opsonic activity. *Fish Shellfish Immunol* (2009) 27:143–53. doi:10.1016/j.fsi.2009.01.004
59. Parisi MG, Cammarata M, Benenati G, Salerno G, Mangano V, Vizzini A, et al. A serum fucose-binding lectin (DIFBL) from adult *Dicentrarchus labrax* is expressed in larva and juvenile tissues and contained in eggs. *Cell Tissue Res* (2010) 341:279–88. doi:10.1007/s00441-010-1004-6
60. Cammarata M, Salerno G, Parisi MG, Benenati G, Vizzini A, Vasta GR, et al. Primary structure and opsonic activity of an F-lectin from serum of the gilt head bream *Sparus aurata* (Pisces, Sparidae). *Ital J Zool* (2011) 79(1):34–43. doi:10.1080/11250003.2011.596167
61. Cho SY, Kwon J, Vaidya B, Kim JO, Lee S, Jeong EH, et al. Modulation of proteome expression by F-type lectin during viral hemorrhagic septicemia virus infection in fathead minnow cells. *Fish Shellfish Immunol* (2014) 39(2):464–74. doi:10.1016/j.fsi.2014.05.042
62. Qiu L, Lin L, Yang K, Zhang H, Li J, Zou F, et al. Molecular cloning and expression analysis of a F-type lectin gene from Japanese sea perch (*Lateolabrax japonicus*). *Mol Biol Rep* (2011) 38(6):3751–6. doi:10.1007/s10333-010-0490-7
63. Bouyain S, Geisbrecht BV. Host glycan recognition by a pore forming toxin. *Structure* (2012) 20(2):197–8. doi:10.1016/j.str.2012.01.013
64. Imbert A, Wimmerová M, Mitchell EP, Gilboa-Garber N. Structures of the lectins from *Pseudomonas aeruginosa*: insight into the molecular basis for host glycan recognition. *Microbes Infect* (2004) 6(2):221–8. doi:10.1016/j.micinf.2003.10.016
65. Imbert A, Varrot A. Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* (2008) 18(5):567–76. doi:10.1016/j.sbi.2008.08.001
66. Quesenberry MS, O'leary N, Ahmed H, Bianchet M, Amzel M, Marsh A, et al. The protochordate *Clavelina picta* has key components of innate immunity in mammals MBP-like, MASP-like, and complement-like molecules. *FASEB J* (1998) 12(8):A1351, I124.

67. Quesenberry MS, Ahmed H, Elola MT, O'Leary N, Vasta GR. Diverse lectin repertoires in tunicates mediate broad recognition and effector innate immune responses. *Integr Comp Biol* (2003) 43(2):323–30. doi:10.1093/icb/43.2.323
68. Majumder S, Roy A, Mandal C. Prediction of 3-D structures of fucose-binding proteins and structural analysis of their interaction with ligands. *Glycoconj J* (2004) 20(9):545–50. doi:10.1023/B:GLYC.0000043291.42999.98
69. Wu L, Yang X, Duan X, Cui L, Li G. Exogenous expression of marine lectins DIFBL and SpRBL induces cancer cell apoptosis possibly through PRMT5-E2F-1 pathway. *Sci Rep* (2014) 4:4505. doi:10.1038/srep04505
70. Li G, Gao Y, Cui L, Wu L, Yang X, Chen J. *Anguilla japonica* lectin 1 delivery through adenovirus vector induces apoptotic cancer cell death through

interaction with PRMT5. *J Gene Med* (2016) 18(4–6):65–74. doi:10.1002/jgm.2878

Conflict of Interest Statement: 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.

Copyright © 2017 Vasta, Amzel, Bianchet, Cammarata, Feng and Saito. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Long Non-Coding RNAs: Emerging and Versatile Regulators in Host–Virus Interactions

Xing-Yu Meng¹, Yuzi Luo¹, Muhammad Naveed Anwar¹, Yuan Sun¹, Yao Gao¹,
Huawei Zhang¹, Muhammad Munir² and Hua-Ji Qiu^{1*}

¹ State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, China, ² The Pirbright Institute, Woking, United Kingdom

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Junji Xing,
Houston Methodist Research
Institute, United States
Leticia A. Carneiro,
Universidade Federal do Rio de
Janeiro, Brazil

*Correspondence:

Hua-Ji Qiu
huajiqui@hvri.ac.cn,
qiu_huaji@163.com

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 20 July 2017

Accepted: 13 November 2017

Published: 28 November 2017

Citation:

Meng X-Y, Luo Y, Anwar MN, Sun Y,
Gao Y, Zhang H, Munir M and
Qiu H-J (2017) Long Non-Coding
RNAs: Emerging and Versatile
Regulators in Host–Virus Interactions.
Front. Immunol. 8:1663.
doi: 10.3389/fimmu.2017.01663

Long non-coding RNAs (lncRNAs) are a class of non-protein-coding RNA molecules, which are involved in various biological processes, including chromatin modification, cell differentiation, pre-mRNA transcription and splicing, protein translation, etc. During the last decade, increasing evidence has suggested the involvement of lncRNAs in both immune and antiviral responses as positive or negative regulators. The immunity-associated lncRNAs modulate diverse and multilayered immune checkpoints, including activation or repression of innate immune signaling components, such as interleukin (IL)-8, IL-10, retinoic acid inducible gene I, toll-like receptors 1, 3, and 8, and interferon (IFN) regulatory factor 7, transcriptional regulation of various IFN-stimulated genes, and initiation of the cell apoptosis pathways. Additionally, some virus-encoded lncRNAs facilitate viral replication through individually or synergistically inhibiting the host antiviral responses or regulating multiple steps of the virus life cycle. Moreover, some viruses are reported to hijack host-encoded lncRNAs to establish persistent infections. Based on these amazing discoveries, lncRNAs are an emerging hotspot in host–virus interactions. In this review, we summarized the current findings of the host–virus-encoded lncRNAs and the underlying mechanisms, discussed their impacts on immune responses and viral replication, and highlighted their critical roles in host–virus interactions.

Keywords: long non-coding RNAs, viral replication, antiviral response, virus–host interactions, regulatory mechanisms

INTRODUCTION

With the rapid development of DNA sequencing technologies, the whole genomes of several species have been mapped and annotated. The first transcriptome analysis performed a decade ago came to a surprising conclusion that only about 2% of the genomic DNA harbors protein-coding genes (1). In the beginning of the 21st century, Okazaki et al. have analyzed the mouse transcriptome based on a cDNA library and identified a mass of non-coding RNAs (ncRNAs), which are defined as a class of RNA molecules without protein-coding capacity (2). In addition, the Encyclopedia of DNA Elements (ENCODE) project has widely been applied to identify the functional DNA elements in the human genome, and showed that approximately 62% of the transcriptome is ncRNAs (3, 4), indicating ncRNAs as major components of the transcriptome (5). In comparison with mRNAs, less is known about the functions and underlying mechanisms of ncRNAs in different biological processes. Based on the sequence length, ncRNAs are usually

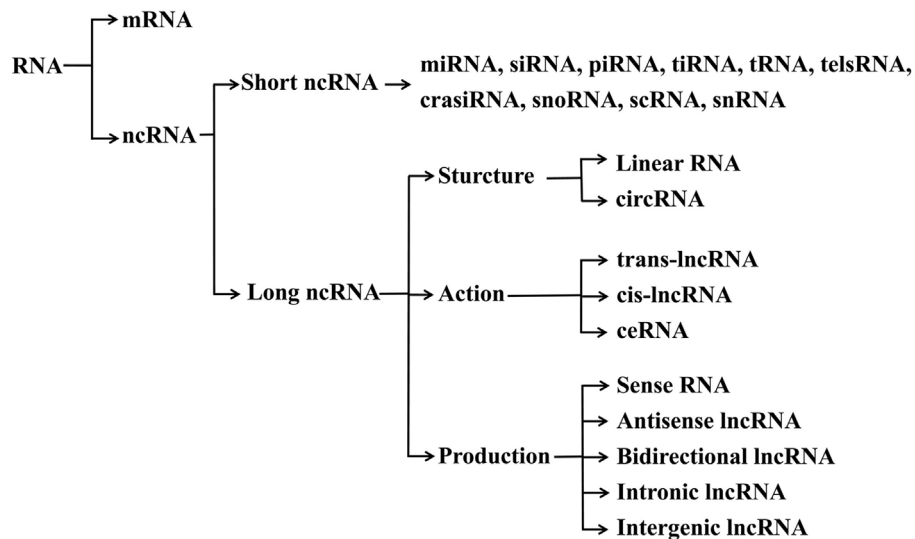


FIGURE 1 | Classification of ncRNAs. mRNA, messenger RNA; ncRNA, non-coding RNA; miRNA, microRNA; siRNA, small interfering RNA; piRNA, piwi RNA; tiRNA, transcription initiation RNA; tRNA, transfer RNA; telsRNA, telomere-specific small RNA; crasiRNA, centromere repeat-associated small interacting RNA; snoRNA, small nucleolar RNA; scRNA, small cytoplasmic RNA; snRNA, small nuclear RNA; trans-lncRNA, *trans*-acting long non-coding RNA; cis-lncRNA, *cis*-acting long non-coding RNA; ceRNA, competing endogenous RNA.

divided into long ncRNAs (lncRNAs, more than 200 nt) and short ncRNAs (sncRNAs, less than 200 nt) (6) (**Figure 1**).

In recent years, lncRNAs have been found to be critical regulators in various biological processes such as cell differentiation, chromatin modification, pre-mRNA transcription and splicing, and protein translation and translocation (7–9). Under a natural physiological state, lncRNAs usually function through enhancing or inhibiting the expression of neighboring protein-encoding genes (10). However, the investigation of potential roles of lncRNAs in virus–host interactions is still in the infancy stage. As a wide range of immunity-related lncRNAs has been identified based on differential expression analysis in response to viral infections, the host lncRNAs have been shown to act as regulators in the innate or adaptive immune signaling pathways (11, 12). Furthermore, emerging evidence demonstrates that viral genomes can transcribe their own lncRNAs by using the host transcription machinery, and these lncRNAs may be involved in the virus life cycle to regulate host or viral gene expression. Meanwhile, viruses can also regulate the expression of host lncRNAs to establish and maintain persistent infections.

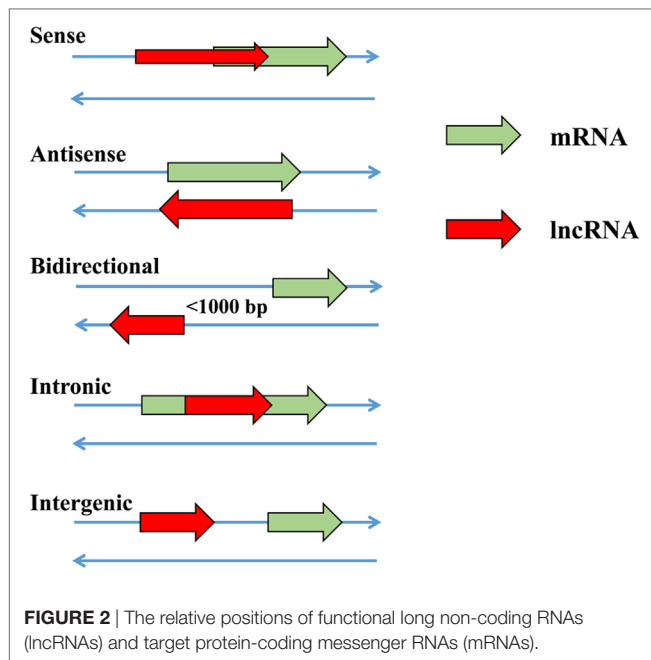
For decades, studies on virus-related host immune responses have been focused mainly on genes or proteins. However, recent studies have shown that lncRNAs may also participate in these biological processes. This review will focus on the lncRNAs involved in host–virus interactions and underlying regulatory mechanisms.

SOURCES AND FUNCTIONS OF lncRNAs

Most of eukaryotic lncRNAs are transcribed by RNA polymerase II, whereas a limited number of lncRNAs are transcribed by

cellular RNA polymerase III (13). After transcription and modification processes, some mature lncRNAs have a similar structure to that of mRNA, including methylguanosine at 5'-terminus and a polyadenylated [poly(A)] tail at the 3'-terminus (13, 14). Indeed, broader analysis has suggested that 39% of lncRNAs transcripts contain one or more of the six most common poly(A) motifs, compared with 51% observed for coding transcripts (13). These properties indicate that there are few particular structural features that allow differentiation of lncRNAs from mRNAs. Nevertheless, compared with mRNAs, lncRNAs are more specific in spatial expression and poorly conserved (15, 16). To date, five possible sources of lncRNAs have been verified: (1) DNA fragments can be assembled and transformed into a functional lncRNA; (2) due to chromosomal rearrangement, two or more mutually independent sequences link together to generate a lncRNA; (3) due to retrotransposition, duplication of non-coding genes can generate functional or non-functional lncRNAs; (4) duplication events from two neighboring tandems give rise to a sequence repeat lncRNA; (5) insertion of a transposable element in a gene generates a lncRNA (17).

In recent years, lncRNAs have been confirmed as a novel group of regulatory molecules in a wide range of biological or cellular processes (17–21). In the nucleus, lncRNAs participate in regulating the expression of nearby and overlapping genes in either RNA-independent or transcription-initiation manner after epigenetic modification (22). The lncRNA HOTAIR has been proved to repress gene expression by recruiting the histone protein (20). lncRNAs may function as enhancers to promote the expression of nearby genes (23–25). At the promoter regions, lncRNAs overlap with DNA sequence and assist the gene to maintain the transcriptional condition, which may be a common function *in cis*-regulation (26). In addition, lncRNAs can competitively bind to



miRNAs to prevent the degradation or repression of target mRNA (27, 28). Based on transcriptional directions and relative positions with target mRNAs, lncRNAs are usually classified into five major categories, i.e., sense lncRNAs, antisense lncRNAs, bidirectional lncRNAs, intrinsic lncRNAs and intragenic lncRNAs (29) (**Figure 2**). The antisense lncRNAs comprise a significant proportion (almost 20%) of the total lncRNAs in mammalian genomes and 75% antisense lncRNAs are able to upregulate the expression of adjacent genes (30). In addition, more than 50% of protein-coding genes carry a complementary lncRNAs in mammals (31).

ANTIVIRAL ACTIVITIES OF HOST lncRNAs

lncRNAs Are Involved in the Innate Immune Responses against Viral Infections

As mentioned above, diverse biological processes in eukaryotic cells are regulated by lncRNAs. However, it is noteworthy that viral infections may lead to the differential expression of host lncRNAs and this change seems to exist as a common pathological phenomenon (32–36). Some differentially expressed host lncRNAs may exert antiviral actions involved in different immune signaling pathways. Guttman et al. reported the modulation of lncRNAs transcription by regulatory proteins for the first time and uncovered over 100 lncRNAs with potential functions in four mouse cell types, i.e., mouse embryonic stem cells, mouse embryonic fibroblasts, mouse lung fibroblasts, and neural precursor cells, by using chromatin immunoprecipitation and massive parallel sequencing (37). Furthermore, it has also been confirmed that the transcription of lncRNAs is associated with immunity-related factors, such as nuclear factor κ B (NF- κ B) (39 lncRNAs), sex-determining region of Y chromosome-related high-mobility-group box 2 (Sox2) (20 lncRNAs), and p53 (118

lncRNAs). With the widespread applications of microarray and RNA sequencing technologies, differentially expressed lncRNAs have been identified to be involved in innate immune responses (32, 38–43).

lncRNAs Regulate the Interferon (IFN) Pathway of the Innate Immune Response

The lncRNA nuclear enriched abundant transcript 1 (NEAT1) is a well-defined positive regulatory component in interleukin (IL)-8 signaling pathway, which can activate the antiviral response. Influenza virus, human immunodeficiency virus (HIV), and other viral infections induce the expression of NEAT1, leading to the formation of nuclear body paraspeckles (44, 45). Splicing factor proline/glutamine-rich (SFPQ) is a negative regulatory factor of IL-8. NEAT1 mediates the relocation of SFPQ from the IL-8 promoter region to paraspeckles and activates the transcription of IL-8 (46). Although the exact antiviral mechanism of IL-8 is not clear, the concentration of IL-8 is proportional to the resistance against HIV infection in a macaque model (47). Moreover, NEAT1, as a binding scaffold, maintains integrity of paraspeckles and prevents the export of spliced pre-mRNA to the cytoplasm for translation. During HIV infection, the upregulated NEAT1 sequesters HIV mRNAs within the nucleus and inhibits viral replication (34). Another study shows that NEAT1 is significantly upregulated postinfection with Hantaan virus (HTNV), whereas inhibiting the expression of NEAT1 delays host innate immune responses and promotes viral replication (48). Further investigations indicate that NEAT1 removes and relocates SFPQ to paraspeckles, inducing the expressions of retinoic acid inducible gene I (RIG-I) and DEXDH box helicase (DDX60). Increased expression of DDX60 and RIG-I enhances IFN- β production and subsequently suppresses HTNV infection.

The lncRNA Cox2, located at 50 kb downstream of the Cox2 protein coding gene, regulates the activation and repression of hundreds of genes (36). It has been revealed that 787 genes are repressed by the lncRNA Cox2 in non-stimulated bone marrow-derived macrophages and 713 genes are expressed following exposure to toll-like receptor (TLR) 1/2 agonist palmitoyl-3-cysteinyl-seryl-(lysyl)₄ (Pam3CSK₄) (41). The subsequent gene ontology (GO) analysis has revealed that the differentially expressed genes are involved in the regulation of immune responses. The whole transcriptome profiling has proven that Cox2 is in charge of activating and inducing interferon regulatory factor 7 (IRF7) and IL-10 and repressing TLR1, 3, and 8, which regulates the expression of various genes in both positive and negative regulatory manners (41). Although the exact regulatory mechanisms remain unknown, researchers speculated that the inhibitory actions of Cox2 could be mediated through binding to heterogeneous nuclear ribonucleoprotein (hnRNP)-A/B and hnRNP-A2/B1. Collectively, lncRNA Cox2 is a key regulatory factor of the circuit adjusting the TLR signaling pathway.

lncRNAs Mediate Other Pathways of the Innate Immune Response

Tumor necrosis factor- α (TNF- α) is a significant activator of host immune responses to viral infections (49–51). Recently, it has been shown that TNF- α is regulated by a lncRNA, TNF- α

and hnRNPL- immunoregulatory lncRNA (THRIL) (38). The THRIL is located downstream of BRI3-binding protein (BRI3BP) and partially overlapped with the 3'-terminus of BRI3BP. This lncRNA THRIL is an essential factor for the induction of TNF- α gene expression by forming a complex with hnRNPL at the promoter/enhancer region of TNF, resulting in the activation of immune response genes (38). On the other hand, THRIL can also be downregulated by the activated TNF through a negative feedback mechanism. These findings highlight a wider spectrum of lncRNA roles in several cellular processes and warrant future investigations.

lncRNAs Participate in the Regulation of the Expression of Interferon-Stimulated Genes (ISGs)

ISGs are induced through the IFN signaling pathway and critical for antagonizing viral infections (52). To date, new antiviral ISGs are discovered as antiviral effectors in the innate antiviral responses (53, 54). In addition, ISGs have been confirmed to have numerous antiviral functions, such as interfering with and inhibiting viral infections, and limiting viral replication within the cells (52). However, molecular mechanisms of regulation of the ISGs expression are complicated (53, 55). Currently, several studies demonstrate that lncRNAs are the key regulators of ISGs.

Some viruses can induce the expression of the lncRNA BISPR (BST2 IFN-stimulated positive regulator) through the JAK-STAT pathway, such as influenza virus, vesicular stomatitis virus or hepatitis C virus (HCV) (56–59). BISPR is located head-to-head with the ISG BST2 gene, the BST2 protein can attach viruses to the cells and inhibit viral release (60, 61). Knockdown or overexpression of BISPR results in a decrease or increase of BST2 expression, respectively, suggesting that BISPR is critically responsible for the transcription of BST2. BISPR exists mainly in the nucleus and possibly facilitates the transcription initiation of protein-coding genes. As mentioned above, some lncRNAs regulate the chromatin state through recruiting and binding to various chromatin-modifying factors. Likewise, BISPR performs its regulatory function by counteracting the repressive action of polycomb repression complex 2 (PRC2) at the promoter of BST2, and the methyltransferase component of EZH2 is also involved in this mechanism (56). In addition, BISPR overlaps with an enhancer region, indicating that BISPR acts as enhancer-associated RNAs (eRNAs) to promote the formation of enhancer-promoter complex.

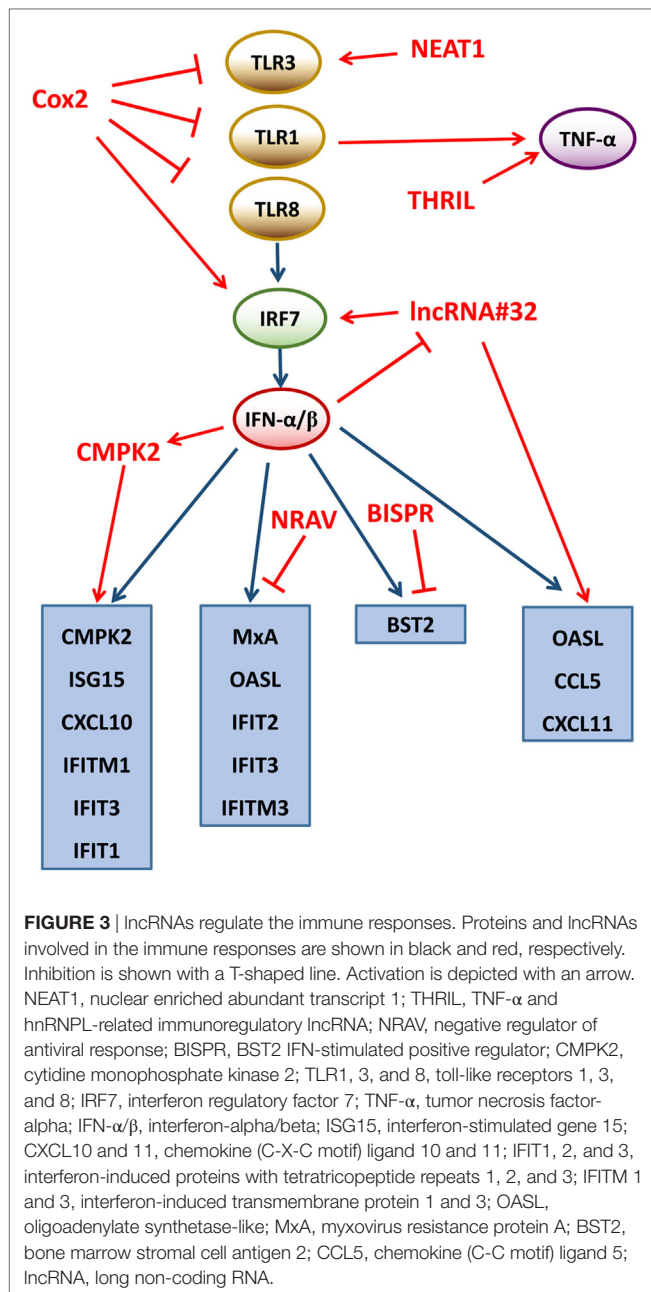
A functional lncRNA, called negative regulator of antiviral response (NRAV), is downregulated dramatically during influenza A virus (IAV) infection (62). Overexpression of NRAV in human cells or transgenic mice significantly increases IAV replication and virulence, whereas knockdown of NRAV suppresses IAV replication, indicating that NRAV is involved in antiviral immune responses. A cDNA microarray analysis reveals that many ISGs are downregulated in NRAV-overexpressing cells, such as IFIT2, IFIT3, IFITM3, OASL, and MxA, and these ISGs exert antiviral effects through multiple mechanisms (63–66). A subsequent study indicates that NRAV negatively regulates the initial transcription rates of IFITM3 and MxA through altering histone modifications (active H3K4me3 and repressive

H3K27me3) on the promoters, and the spatial structure of NRAV is necessary for its regulatory function (62).

The lncRNA CMPK2 is located proximally to the ISGs CMPK2, which is mapped to chr2p25.2 (chr2:6,968,644–6,980,595). The lncRNA CMPK2 can be upregulated significantly by IFN- α or IFN- γ (25, 67). Knockdown of lncRNA CMPK2 in hepatocytes results in remarkable reduction in HCV replication and increases expression of some antiviral ISGs, suggesting that the lncRNA CMPK2 is a critical repressor of ISGs and a lncRNA-mediated negatively regulatory mechanism may exist. In addition, the level of the lncRNA CMPK2 is dramatically higher in the liver of HCV-infected patients compared with healthy donors, indicating that the lncRNA CMPK2 also plays a regulatory role in viral infections *in vivo* (25), whereas overexpression of the lncRNA CMPK2 inhibits the transcription of ISGs, such as CMPK2 and viperin. Interestingly, some ISGs located far from the lncRNA CMPK2 in the genome can also be repressed, such as ISG15, IFIT1, IFIT3, CXCL10, MxA, and IFITM1. Nevertheless, a few of ISGs seem to inhibit the transcription of the lncRNA CMPK2, including IFIT1 and Mx1. However, the impact of silencing of the lncRNA CMPK2 on ISG levels is not consistent with other IFN-stimulated negative regulatory factors, such as activating signal cointegrator 1 complex subunit 3. Thus, it is considered that the regulatory mechanism of lncRNA CMPK2 may be similar to other lncRNAs, such as NRAV. Similarly, lncRNA CMPK2 interacts with transcription factors or chromatin to form complexes to regulate the gene expression.

The lncRNA#32 is located on human chromosome 7p13 and overlaps the 3'-terminus of the HECT, C3, and WW domain containing E3 ubiquitin protein ligase 1 (HECW1) (68). Silencing lncRNA#32 significantly reduces the expression level of some ISGs and chemokines, including IRF7, chemokine (C-C motif) ligand 5 (CCL5), CXCL11, OASL, RSAD2, and IP-10, resulting in susceptibility to encephalomyocarditis virus (EMCV) infection. In contrast, the overexpression of lncRNA#32 dramatically suppressed EMCV replication, indicating that lncRNA#32 positively regulates the host antiviral response (68). The expression of OASL is induced by IFN- β , whereas the expression of lncRNA#32 is repressed by IFN- β in a dose-dependent manner. lncRNA#32 positively regulates the expression of ISGs through its interaction with activating transcription factor 2 (ATF2). The ATF2-binding region deletion mutant of lncRNA#32 does not induce IP-10 expression. The research also finds that heterogeneous nuclear ribonucleoprotein U (hnRNPU) maintains the expression of these ISGs by binding to and stabilizing lncRNA#32. These findings highlight the possibility that the hnRNPU-lncRNA#32 complex may target promoters of ISGs to promote the transcription (Figure 3).

Taken together, current understandings propose the nature and breadth of lncRNAs in the regulation of ISGs, which define the first line of defense against pathogens. While a significant baseline has been made, extensive future studies are required to underpin this important aspect of host-pathogen interactions along with their impacts on virus biology and host responses.



lncRNAs Are Involved in the Adaptive Immune Response

Although the existence of lncRNAs in T cells has been known for years, such as growth-arrest-specific transcript 5 (Gas5) and non-coding transcript in CD4⁺ T cells, the lncRNA screening has recently been conducted in CD8⁺ T cells (69). A total of 1,524 lncRNAs were identified from 42 mouse T cell subsets using a microarray assay and some of them were lymphoid-specific lncRNAs, which were increased during CD8⁺ T cell activation and differentiation into effector T cells (70). At the differentiation state of CD4⁺ T cells to T_H1 or T_H2 subsets, T_H1-related transcription factors, such as STAT4 and T-box transcription factor, can

induce the expression of some T_H1-specific lncRNAs. Likewise, T_H2 transcription factor STAT6 regulates T_H2-specific lncRNAs expression. In addition, lncRNA Gas5 represses T cell proliferation. Overexpression of Gas5 inhibits cell-cycle progression and initiates the cell apoptosis signaling pathways (71). Limited studies have been conducted to investigate the roles of lncRNAs in adaptive immune responses; however, current evidences propose crucial roles of lncRNAs in regulation of adaptive immunity and thus warrant future investigations.

HOST lncRNAs ARE HIJACKED BY VARIOUS VIRUSES

Host lncRNAs have been confirmed as positive or negative antiviral regulators in the immune response; surprisingly, a few of host lncRNAs can be induced and hijacked by certain viruses to establish persistent infections. This is likely due to the mutual adaptability of hosts and viruses for millions of years.

The lncRNA NeST, also known as Tmevpg1 or IfngAS1, is located adjacent to the IFN- γ gene in both humans and mice that can positively regulate the expression of IFN- γ (72). NeST can bind to WD repeat-containing protein 5 (WDR5), a component of histone H3 lysine 4 (H3K4) methyltransferase complex, and alter histone 3 methylation at the IFN- γ locus, resulting in the IFN- γ expression (72). In addition, the transcription of both mouse and human NeST gene is dependent on NF- κ B and transcription factors STAT4 and T-bet (73, 74). An earlier study has shown that NeST is specially expressed in T_H1 CD4⁺ T cells and is considered to be associated with immune response (73). Another similar study has indicated that NeST facilitates Theiler's virus infection (75), which is verified using B10.S and SJL/L mouse models. The SJL/L mice with NeST gene show increased IFN- γ expression in activated CD8⁺ T cells, leading to persistent infection of Theiler's virus, and the NeST gene-knockout B10.S mice can clear the virus by its own immune system. The transgenic B10.S mice carrying the allele of NeST are unable to resist the viral infection either. Thus, Theiler's virus establishes persistent infections by hijacking the host lncRNA NeST.

The lncRNA NRON is required to regulate the activity of nuclear factor of activated T cells (NFAT) by forming a ribonucleoprotein complex with NFAT kinases and expression of this lncRNA is significantly altered following HIV-1 infection (45, 76–79). The regulation of NRON expression during the HIV-1 life cycle is complex. The level of NRON is reduced by the HIV-1 early accessory protein Nef and the dephosphorylated NFAT can be translocated to the nucleus and activates the expression of several genes of HIV (78). Knockdown of NRON enhances virus replication through increasing the activity of NFAT. However, high-level expression of NRON is induced by the HIV-1 accessory protein Vpu at the late stages of HIV infection, resulting in viral release and apoptosis. It has been demonstrated that the expression level of NRON is modulated by the HIV-1 Nef and Vpu proteins at different times postinfection to fit the virus life cycle. This finding explains how HIV regulates the host lncRNA NRON to facilitate viral infection.

lncRNA-ACOD1, located near the ACOD1 protein-coding gene, can be induced by various viruses, including Sendai virus

(SeV), vesicular stomatitis virus (VSV), herpes simplex virus (HSV), and vaccinia virus (VACV) (80). In addition, lncRNA-ACOD1 is an IFN- α -independent lncRNA, of which expression is regardless of IFN- α receptor deficiency and IFN- α stimulation. Knockdown of lncRNA-ACOD1 significantly reduces viral load of VSV in macrophages and VSV replication is remarkably reduced in the lncRNA-ACOD1-deficient mice, indicating that the lncRNA promotes virus replication (80). Microarray transcriptome analysis shows that the lncRNA-ACOD1 deficiency leads to changes in the expressions of many metabolism-related genes, indicating the potential role of the lncRNA in regulation of metabolism upon viral infection. RNA immunoprecipitation assay suggests that lncRNA-ACOD1 directly binds to the metabolic enzyme glutamic-oxaloacetic transaminase 2 (GOT2) near the substrate niche, enhancing its catalytic activity. It has been shown that lncRNA-ACOD1 overexpression promotes viral replication in control cells, while has no effect in GOT2-knockdown cells. Taken together, these results demonstrate that lncRNA-ACOD1 facilitates viral replication through promoting GOT2 activity.

VIRALLY ENCODED lncRNAs INHIBIT ANTIVIRAL RESPONSES

The existence of virus-encoded lncRNAs has been identified for years (81, 82). However, only recently, their roles in virus pathobiology and host responses have been explored. The viral lncRNAs are generally transcribed from RNA polymerase II or III, and some of lncRNAs can even be polyadenylated, similar to host mRNA (11, 83). Interestingly, some viral lncRNAs even need unique maturation steps using host cell transcription machineries.

A polyadenylated nuclear RNA (lncRNA PAN) expressed by Kaposi's sarcoma-associated herpesvirus is localized within the cell nucleus and accumulated largely during lytic infection. Several studies demonstrate that PAN represses host gene transcription through a variety of mechanisms. Interferon regulatory factor 4 (IRF4) is a transcription factor that can bind to and transactivate the IL-4 promoter along with PU.1 (84). However, the expression of PAN interferes with the transcription of IL-4 through preventing PU.1 binding to IL-4 promoter (85). In addition, the results also suggest that PAN decreases the expression of several immune regulators, including IL-18, RNase L, IFN-16, and IFN- γ . This mechanism is closely connected to the extensive binding capacity of PAN and host transcriptional proteins, such as histones H1 and H2A, and mitochondrial and cellular single-stranded binding proteins. Another similar study indicates that PAN suppresses the expression of host antiviral genes by activating the PRC2 (83). Besides broadly inhibiting actions of immunity-related genes, PAN also participates in regulating the virus life cycle. In this context, it has been shown that PAN is able to bind to ubiquitously transcribed tetratricopeptide repeat X chromosome (UTX) and jumonji domain containing 3 (JMJD3) to remove the H3K23me3 from the viral genome, resulting in the change of virus life cycle from latent to lytic infection (86, 87). In addition, PAN interacts with the latency-associated nuclear antigen protein (LANA) to

maintain latent infection. Collectively, the viral lncRNA PAN regulates both host and viral gene expression to inhibit antiviral responses and regulate virus life cycle.

Another lncRNA Beta2.7, transcribed from the human cytomegalovirus genome, exists at the early stages of viral infection (88, 89). Beta2.7 and GRIM19 (gene associated with retinoid/IFN-induced mortality-19) are combined together to form a subunit of mitochondrial complex I, which is key for stabilizing the mitochondrial membrane potential, leading to continued production of adenosine triphosphate, which is critical for the completion of the virus life cycle (90–92). Beta2.7 may also protect mitochondrial complex I against stress-induced apoptosis and prevent neuron death.

The 5'-3' exonuclease Xrn1 functions in mRNA decay as well as degradation of flavivirus genomic RNA (84, 93). Most of the RNAs, even the ones with strong secondary or tertiary structures, cannot resist Xrn1 degradation. Surprisingly, the subgenomic flavivirus RNAs (sfrRNAs), generated from viral genome, accumulate to a high level in cells and repress the activation of Xrn1 (94–96). A further study demonstrates that the lncRNA sfrRNAs are transcribed at the 3'-terminus of flavivirus genome. Based on the special stem-loop structure, the lncRNA sfrRNAs bind to the Xrn1 and inhibit its cascade function. Moreover, Xrn1 can also be used to form new 5'-terminus of transcripts to improve viral gene expression *via* the generation of the lncRNA sfrRNAs (95). The lncRNAs from hepaciviruses (e.g., HCV) and pestiviruses (e.g., bovine viral diarrhea virus) are shorter than those from arthropod-borne flaviviruses, which implies that they may play unique roles in the virus life cycle. The transcription and function of the lncRNA sfrRNAs indicate that flaviviruses repress host immune system with virus-encoded lncRNAs (Table 1).

CONCLUDING REMARKS AND PROSPECTS

Formerly, lncRNAs were considered as non-functional gene transcripts in cells and the studies on host–virus interactions were mainly focus on the genomic DNA and proteins of hosts or viruses. However, in the past few years, powerful evidence supports that some lncRNAs from hosts or viruses are actively involved in host–virus interactions. On one hand, host-encoded lncRNAs are supposed to exert antiviral functions *via* different immune response processes, including innate and adaptive immune responses and ISG expression through completely different mechanisms. On the other hand, viruses seem to hijack host lncRNAs or to exploit viral lncRNAs for inhibition of antiviral responses and virus persistence. Thus, besides DNA and proteins, lncRNAs are a new kind of actors in host immune defense and virus survival.

Here, we raise a question: how to identify functional lncRNAs? To obtain the potential lncRNAs, conventionally researchers analyze the transcriptome and screen the differential expression of mRNAs and lncRNAs induced by viral infections. However, a leading challenge is how to separate lncRNAs from mRNAs in large-scale transcriptome data, since hundreds or even thousands of differentially expressed lncRNAs will be obtained using

TABLE 1 | Characteristics of lncRNAs involved in host–virus interactions.

Functions of lncRNAs	Names	Mechanisms	Sources	References
Antiviral responses	NeST	NeST interacts with WDR5 to alter histone 3 methylation at the IFN- γ locus to induce the IFN- γ expression	Host	(72)
	NRAV	NRAV inhibits the initial transcription of IFITM3 and MxA by regulating the histone modifications of these ISG genes	Host	(62)
	lncRNA#32	lncRNA#32 significantly increases the expression of IRF7, CCL5, CXCL11, OASL, RSAD2, and IP-10 through its interaction with ATF2 and hnRNPU	Host	(68)
	BISPR	BISPR induces the transcription of BST2 gene in trans by counteracting the repressive action of PRC2	Host	(56)
	Cox2	Cox2, induced by TLR, can interact with hnRNP-A/B and hnRNP-A2/B1 to mediate the immune responses in both positive and negative regulatory signaling pathways	Host	(41)
	THRIL	THRIL binds to hnRNPL and TNF promoter/enhancer region to induce TNF- α expression and is downregulated by TNF activation through a negative feedback mechanism	Host	(38)
	NEAT1	NEAT1 activates the transcription of IL-8, RIG-I, and DDX60 through removal of the transcriptional inhibitory effects of SFPQ from promoter region by relocating SFPQ to paraspeckles	Host	(34, 44, 46)
	CMPK2	lncRNA CMPK2, as a negative regulatory factor in ISGs response, is involved in the regulation of ISGs transcription by forming RNA-protein complexes with chromatin remodeling or transcription factors	Host	(25)
Virus infections	NRON	The HIV-1 Nef and Vpu proteins reduce or increase the expression of NRON at different times postinfection to regulate the virus life cycle, resulting in persistent infection	Host	(77)
	NeST	Overexpression of NeST has been shown to increase the persistence of Theiler's virus and reduce the host resistance	Host	(75)
	lncRNA-ACOD1	lncRNA-ACOD1 is induced during viral infection and facilitates viral replication through promoting the catalytic activity of GOT2	Host	(80)
	PAN	PAN is a key regulator in controlling gene expression by multiple mechanisms. Many immunity-related genes, such as IL-4, IFN- γ , IL-18, and IFN- α , are regulated by lncRNA PAN. In addition, PAN participates in regulating the virus life cycle through removing the suppressive H3K23me3 from the viral genome and interacting with LANA to maintain latent infection	Virus	(83–87)
	Beta2.7	Beta2.7 and GRIM19 are combined together to form a subunit of mitochondrial complex I, leading to continued production of adenosine triphosphate	Virus	(92)
	sfRNAs	Based on the special secondary structure, sfRNAs bind to Xrn1 and inhibit its degradation of flaviviral genomic RNA	Virus	(94, 95)

lncRNAs, long non-coding RNAs; WDR5, WD repeat-containing protein 5; IFN- γ , interferon- γ ; NRAV, negative regulator of antiviral response; ISG, interferon-stimulated gene; IRF7, interferon regulatory factor 7; BISPR, BST2 IFN-stimulated positive regulator; hnRNPU, heterogeneous nuclear ribonucleoprotein U; ATF2, activating transcription factor 2; PRC2, polycomb repression complex 2; TLR, toll-like receptor; hnRNP, heterogeneous nuclear ribonucleoprotein; THRIL, TNF- α and hnRNPL-related immunoregulatory lncRNA; NEAT1, nuclear enriched abundant transcript 1; HIV, human immunodeficiency virus; GOT2, glutamic-oxaloacetic transaminase 2; sfRNAs, subgenomic flavivirus RNAs.

RNA-seq data, making it laborious to identify functional lncRNAs. Indeed, unlike mRNAs, the sequences of lncRNAs usually display poor evolutionary conservation among different species, thus it is difficult to use conventional bioinformatic tools to predict their functions. In addition, the sequences of lncRNAs are yet to be determined in most species. In spite of these limitations, many lncRNAs from viruses or hosts have been disclosed in recent years. We propose to establish bioinformatics pipelines to genetically annotate lncRNAs by incorporating our current understandings on the functions of lncRNAs in the future.

Since lncRNAs are associated with DNA, mRNA or proteins, it is worth thinking about the possible existence of potential links between lncRNAs and miRNAs. This speculation is supported by some studies that lncRNAs can act as efficient miRNA “sponges” to reduce miRNA levels or through binding to primary miRNAs to repress miRNA maturation (97, 98). However, the discovery about the functions of sncRNAs is scarcely reported in viral infections or host–virus interactions. Up to now, the interactions between miRNA and lncRNAs are a freshly new frontier research area.

Currently, relatively complete lncRNA databases have been established only for human and model animal species (mouse

and rat). However, based on the current findings, we believe that lncRNA databases for broader species will facilitate the study on natures and dynamics of lncRNAs-mediated antiviral responses and regulation of the virus life cycle.

In conclusion, growing evidence suggests that additional hosts- or viral-origin lncRNAs remain undiscovered, and systematic and novel probing approaches are required to characterize functional lncRNAs and identify clinically relevant lncRNAs with broader antiviral characteristics.

AUTHOR CONTRIBUTIONS

X-YM is the major contributor of the review. YL, MNA, YS, YG, HZ and MM participate in the modification of the article. H-JQ conceived and revised the paper.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (no. 31700139 and 31402194) and China Postdoctoral Science Foundation (no. 2016M591313). We appreciate Drs. Muhammad Abid and Teshale Teklue for editing the manuscript.

REFERENCES

- Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. Genome science: the transcriptional landscape of the mammalian genome. *Science* (2005) 309(5740):1559–63. doi:10.1126/science.1112014
- Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, et al. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* (2002) 420(6915):563–73. doi:10.1038/nature01266
- ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* (2012) 489(7414):57–74. doi:10.1038/nature11247
- Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* (2007) 447(7146):799–816. doi:10.1038/nature05874
- Cech TR, Steitz JA. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* (2014) 157(1):77–94. doi:10.1016/j.cell.2014.03.008
- Katsarou K, Rao ALN, Tsagris M, Kalantidis K. Infectious long non-coding RNAs. *Biochimie* (2015) 117:37–47. doi:10.1016/j.biochi.2015.05.005
- Prasanth KV, Spector DL. Eukaryotic regulatory RNAs: an answer to the ‘genome complexity’ conundrum. *Genes Dev* (2007) 21(1):11–42. doi:10.1101/gad.1484207
- Amaral PP, Dinger ME, Mercer TR, Mattick JS. The eukaryotic genome as an RNA machine. *Science* (2008) 319(5871):1787–9. doi:10.1126/science.1155472
- Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* (2011) 470(7333):284–8. doi:10.1038/nature09701
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* (2007) 316(5830):1484–8. doi:10.1126/science.1138341
- Fortes P, Morris KV. Long noncoding RNAs in viral infections. *Virus Res* (2016) 212:1–11. doi:10.1016/j.virusres.2015.10.002
- Heward JA, Lindsay MA. Long non-coding RNAs in the regulation of the immune response. *Trends Immunol* (2014) 35(9):408–19. doi:10.1016/j.it.2014.07.005
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* (2012) 22(9):1775–89. doi:10.1101/gr.132159.111
- Caley DP, Pink RC, Trujillo D, Carter DR. Long noncoding RNAs, chromatin, and development. *ScientificWorld Journal* (2010) 10:90–102. doi:10.1100/tsw.2010.7
- Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, et al. Landscape of transcription in human cells. *Nature* (2012) 489(7414):101–8. doi:10.1038/nature11233
- Ponjavic J, Ponting CP, Lunter G. Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* (2007) 17(5):556–65. doi:10.1101/gr.6036807
- Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell* (2009) 136(4):629–41. doi:10.1016/j.cell.2009.02.006
- Huarte M, Rinn JL. Large non-coding RNAs: missing links in cancer? *Hum Mol Genet* (2010) 19(R2):R152–61. doi:10.1093/hmg/ddq353
- Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* (2011) 43(6):904–14. doi:10.1016/j.molcel.2011.08.018
- Gupta RA, Shah N, Wang KC, Kim J, Horlings HM, Wong DJ, et al. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* (2010) 464(7291):1071–6. doi:10.1038/nature08975
- Yoon JH, Abdelmohsen K, Gorospe M. Posttranscriptional gene regulation by long noncoding RNA. *J Mol Biol* (2013) 425(19):3723–30. doi:10.1016/j.jmb.2012.11.024
- Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* (2006) 21(5):617–28. doi:10.1016/j.molcel.2006.01.028
- Carnero E, Barriocanal M, Segura V, Guruceaga E, Prior C, Borner K, et al. Type I interferon regulates the expression of long non-coding RNAs. *Front Immunol* (2014) 5:548. doi:10.3389/fimmu.2014.00548
- Josset L, Tchitchek N, Gralinski LE, Ferris MT, Einfeld AJ, Green RR, et al. Annotation of long non-coding RNAs expressed in collaborative cross founder mice in response to respiratory virus infection reveals a new class of interferon-stimulated transcripts. *RNA Biol* (2014) 11(7):875–90. doi:10.4161/rna.29442
- Kambara H, Niaz F, Kostadinova L, Moonka DK, Siegel CT, Post AB, et al. Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res* (2014) 42(16):10668–80. doi:10.1093/nar/gku713
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, et al. Divergent transcription from active promoters. *Science* (2008) 322(5909):1849–51. doi:10.1126/science.1162253
- Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta stone of a hidden RNA language? *Cell* (2011) 146(3):353–8. doi:10.1016/j.cell.2011.07.014
- Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* (2011) 147(2):344–57. doi:10.1016/j.cell.2011.09.029
- Mercer TR, Dinger M, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* (2009) 10(3):155–9. doi:10.1038/nrg2521
- Luo S, Lu JY, Liu L, Yin Y, Chen C, Han X, et al. Divergent lncRNAs regulate gene expression and lineage differentiation in pluripotent cells. *Cell Stem Cell* (2016) 18(5):637–52. doi:10.1016/j.stem.2016.01.024
- Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, et al. Antisense transcription in the mammalian transcriptome. *Science* (2005) 309(5740):1564–6. doi:10.1126/science.1112009
- Peng X, Gralinski L, Armour CD, Ferris MT, Thomas MJ, Proll S, et al. Unique signatures of long noncoding RNA expression in response to virus infection and altered innate immune signaling. *MBio* (2010) 1(5):e00206–10. doi:10.1128/mBio.00206-10
- Sonkoly E, Bata-Csorgo Z, Pivarcsi A, Polyanka H, Kenderessy-Szabo A, Molnar G, et al. Identification and characterization of a novel, psoriasis susceptibility-related noncoding RNA gene. *PRINS J Biol Chem* (2005) 280(25):24159–67. doi:10.1074/jbc.M501704200
- Zhang Q, Chen CY, Yedavalli VS, Jeang KT. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *MBio* (2013) 4(1):e00596–12. doi:10.1128/mBio.00596-12
- Ahanda ML, Ruby T, Wittzell H, Bed'Hom B, Chausse AM, Morin V, et al. Non-coding RNAs revealed during identification of genes involved in chicken immune responses. *Immunogenetics* (2009) 61(1):55–70. doi:10.1007/s00251-008-0337-8
- Razozyk BS, Obermayer B, O'May JB, Tarakhovskiy A. Viral infection identifies micropeptides differentially regulated in smORF-containing lncRNAs. *Genes (Basel)* (2017) 8(8):206–16. doi:10.3390/genes8080206
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* (2009) 458(7235):223–7. doi:10.1038/nature07672
- Li Z, Chao TC, Chang KY, Lin N, Patil VS, Shimizu C, et al. The long non-coding RNA THRIL regulates TNF-alpha expression through its interaction with hnRNPL. *Proc Natl Acad Sci U S A* (2014) 111(3):1002–7. doi:10.1073/pnas.1313768111
- Rapicavoli NA, Qu K, Zhang J, Mikhail M, Laberge RM, Chang HY. A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics. *Elife* (2013) 2:e00762. doi:10.7554/eLife.00762
- Wang P, Xue Y, Han Y, Lin L, Wu C, Xu S, et al. The STAT3-binding long noncoding RNA lnc-DC controls human dendritic cell differentiation. *Science* (2014) 344(6181):310–3. doi:10.1126/science.1251456
- Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, et al. A long noncoding RNA mediates both activation and repression of immune response genes. *Science* (2013) 341(6147):789–92. doi:10.1126/science.1240925
- Zhang J, Sun P, Gan L, Bai W, Wang Z, Li D. Genome-wide analysis of long noncoding RNA profiling in PRRSV-infected PAM cells by RNA sequencing. *Sci Rep* (2017) 7(1):4952–62. doi:10.1038/s41598-017-05279-z
- Garmire LX, Garmire DG, Huang W, Yao J, Glass CK, Subramaniam S. A global clustering algorithm to identify long intergenic non-coding RNA – with applications in mouse macrophages. *PLoS One* (2011) 6(9):e24051. doi:10.1371/journal.pone.0024051

44. Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* (2009) 33(6):717–26. doi:10.1016/j.molcel.2009.01.026
45. Lazar DC, Morris KV, Saayman SM. The emerging role of long non-coding RNAs in HIV infection. *Virus Res* (2016) 212:114–26. doi:10.1016/j.virusres.2015.07.023
46. Imamura K, Imamachi N, Akizuki G, Kumakura M, Kawaguchi A, Nagata K, et al. Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol Cell* (2014) 53(3):393–406. doi:10.1016/j.molcel.2014.01.009
47. Promadej-Lanier N, Hanson DL, Srinivasan P, Luo W, Adams DR, Guenther PC, et al. Resistance to simian HIV infection is associated with high plasma interleukin-8, RANTES and eotaxin in a macaque model of repeated virus challenges. *J Acquir Immune Defic Syndr* (2010) 53(5):574–81. doi:10.1097/QAI.0b013e3181d3521f
48. Ma H, Han P, Ye W, Chen H, Zheng X, Cheng L, et al. The long noncoding RNA NEAT1 exerts anti-hantaviral effects by acting as positive feedback for RIG-I signaling. *J Virol* (2017) 91(9):2250–2216. doi:10.1128/JVI.02250-16
49. Puthothu B, Bierbaum S, Kopp MV, Forster J, Heinze J, Weckmann M, et al. Association of TNF-alpha with severe respiratory syncytial virus infection and bronchial asthma. *Pediatr Allergy Immunol* (2009) 20(2):157–63. doi:10.1111/j.1399-3038.2008.00751.x
50. Torigo S, Ihara T, Kamiya H, IL-12, IFN-gamma, and TNF-alpha released from mononuclear cells inhibit the spread of varicella-zoster virus at an early stage of varicella. *Microbiol Immunol* (2000) 44(12):1027–31. doi:10.1111/j.1348-0421.2000.tb02599.x
51. Neuzil KM, Tang YW, Graham BS. Protective role of TNF-alpha in respiratory syncytial virus infection *in vitro* and *in vivo*. *Am J Med Sci* (1996) 311(5):201–4. doi:10.1016/S0002-9629(15)41695-7
52. Poynter SJ, DeWitte-Orr SJ. Fish interferon-stimulated genes: the antiviral effectors. *Dev Comp Immunol* (2016) 65:218–25. doi:10.1016/j.dci.2016.07.011
53. Schneider WM, Chevillotte MD, Rice CM. Interferon-stimulated genes: a complex web of host defenses. *Annu Rev Immunol* (2014) 32:513–45. doi:10.1146/annurev-immunol-032713-120231
54. Zhang YB, Gui JF. Molecular regulation of interferon antiviral response in fish. *Dev Comp Immunol* (2012) 38(2):193–202. doi:10.1016/j.dci.2012.06.003
55. Haller O, Kochs G. Interferon-induced mx proteins: dynamin-like GTPases with antiviral activity. *Traffic* (2002) 3(10):710–7. doi:10.1034/j.1600-0854.2002.31003.x
56. Kambara H, Gunawardane L, Zebrowski E, Kostadinova L, Jobava R, Krokowski D, et al. Regulation of interferon-stimulated gene BST2 by a lncRNA transcribed from a shared bidirectional promoter. *Front Immunol* (2014) 5:676. doi:10.3389/fimmu.2014.00676
57. Barriocanal M, Carnero E, Segura V, Fortes P. Long non-coding RNA BST2/BISPR is induced by IFN and regulates the expression of the antiviral factor tetherin. *Front Immunol* (2014) 5:655. doi:10.3389/fimmu.2014.00655
58. Dafa-Berger A, Kuzmina A, Fassler M, Yitzhak-Asraf H, Shemer-Avni Y, Taube R. Modulation of hepatitis C virus release by the interferon-induced protein BST-2/tetherin. *Virology* (2012) 428(2):98–111. doi:10.1016/j.virol.2012.03.011
59. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature* (2008) 451(7177):425–30. doi:10.1038/nature06553
60. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* (2012) 81:145–66. doi:10.1146/annurev-biochem-051410-092902
61. Rinn JL. lncRNAs: linking RNA to chromatin. *Cold Spring Harb Perspect Biol* (2014) 6(8):a018614. doi:10.1101/cshperspect.a018614
62. Ouyang J, Zhu X, Chen Y, Wei H, Chen Q, Chi X, et al. NRAV, a long noncoding RNA, modulates antiviral responses through suppression of interferon-stimulated gene transcription. *Cell Host Microbe* (2014) 16(5):616–26. doi:10.1016/j.chom.2014.10.001
63. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* (2012) 484(7395):519–23. doi:10.1038/nature10921
64. Fensterl V, Wetzel JL, Ramachandran S, Ogino T, Stohlman SA, Bergmann CC, et al. Interferon-induced Ifit2/ISG54 protects mice from lethal VSV neuro-pathogenesis. *PLoS Pathog* (2012) 8(5):e1002712. doi:10.1371/journal.ppat.1002712
65. Liu XY, Chen W, Wei B, Shan YF, Wang C. IFN-induced TPR protein IFIT3 potentiates antiviral signaling by bridging MAVS and TBK1. *J Immunol* (2011) 187(5):2559–68. doi:10.4049/jimmunol.1100963
66. Manz B, Dornfeld D, Gotz V, Zell R, Zimmermann P, Haller O, et al. Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. *PLoS Pathog* (2013) 9(3):e1003279. doi:10.1371/journal.ppat.1003279
67. Huang Z, Luo Q, Yang F, Qing C, Ye J, Deng Y, et al. Identification of differentially expressed long non-coding RNAs in polarized macrophages. *Sci Rep* (2016) 6:19705–14. doi:10.1038/srep19705
68. Nishitsuji H, Ujino S, Yoshio S, Sugiyama M, Mizokami M, Kanto T, et al. Long noncoding RNA#32 contributes to antiviral responses by controlling interferon-stimulated gene expression. *Proc Natl Acad Sci U S A* (2016) 113(37):10388–93. doi:10.1073/pnas.1525022113
69. Pang KC, Dinger ME, Mercer TR, Malquori L, Grimmond SM, Chen W, et al. Genome-wide identification of long noncoding RNAs in CD8+ T cells. *J Immunol* (2009) 182(12):7738–48. doi:10.4049/jimmunol.0900603
70. Hu G, Tang Q, Sharma S, Yu F, Escobar TM, Muljo SA, et al. Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation. *Nat Immunol* (2013) 14(11):1190–8. doi:10.1038/ni.2712
71. Mourtada-Maarabouni M, Hedge VL, Kirkham L, Farzaneh F, Williams GT. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J Cell Sci* (2008) 121(7):939–46. doi:10.1242/jcs.024646
72. Gomez JA, Wapinski OL, Yang YW, Bureau JF, Gopinath S, Monack DM, et al. The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* (2013) 152(4):743–54. doi:10.1016/j.cell.2013.01.015
73. Collier SP, Collins PL, Williams CL, Boothby MR, Aune TM. Cutting edge: influence of Tmevpg1, a long intergenic noncoding RNA, on the expression of ifng by Th1 cells. *J Immunol* (2012) 189(5):2084–8. doi:10.4049/jimmunol.1200774
74. Collier SP, Henderson MA, Tossberg JT, Aune TM. Regulation of the Th1 genomic locus from ifng through Tmevpg1 by T-bet. *J Immunol* (2014) 193(8):3959–65. doi:10.4049/jimmunol.1401099
75. Vigneau S, Rohrlach PS, Brahic M, Bureau JF. Tmevpg1, a candidate gene for the control of Theiler's virus persistence, could be implicated in the regulation of gamma interferon. *J Virol* (2003) 77(10):5632–8. doi:10.1128/JVI.77.10.5632-5638.2003
76. Willingham AT, Orth AP, Batalov S, Peters EC, Wen BG, Aza-Blanc P, et al. A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* (2005) 309(5740):1570–3. doi:10.1126/science.1115901
77. Sharma S, Findlay GM, Bandukwala HS, Oberdoerffer S, Baust B, Li Z, et al. Dephosphorylation of the nuclear factor of activated T cells (NFAT) transcription factor is regulated by an RNA-protein scaffold complex. *Proc Natl Acad Sci U S A* (2011) 108(28):11381–6. doi:10.1073/pnas.1019711108
78. Imam H, Bano AS, Patel P, Holla P, Jameel S. The lncRNA NRON modulates HIV-1 replication in a NFAT-dependent manner and is differentially regulated by early and late viral proteins. *Sci Rep* (2015) 5:8639. doi:10.1038/srep08639
79. Aune TM, Spurlock CF. Long non-coding RNAs in innate and adaptive immunity. *Virus Res* (2016) 212:146–60. doi:10.1016/j.virusres.2015.07.003
80. Wang P, Xu J, Wang Y, Cao X. An interferon-independent lncRNA promotes viral replication by modulating cellular metabolism. *Science* (2017). doi:10.1126/science.aao0409
81. Marx JL. “Viroids”: a new kind of pathogen? *Science* (1972) 178(4062):734. doi:10.1126/science.178.4062.734
82. Mathews MB. Genes for VA-RNA in adenovirus 2. *Cell* (1975) 6(2):223–9. doi:10.1016/0092-8674(75)90013-6
83. Rossetto CC, Tarrant-Elorza M, Verma S, Purushothaman P, Pari GS. Regulation of viral and cellular gene expression by Kaposi's sarcoma-associated herpesvirus polyadenylated nuclear RNA. *J Virol* (2013) 87(10):5540–53. doi:10.1128/JVI.03111-12
84. Ahly AN, Chang HC, Dent AL, Nutt SL, Kaplan MH, et al. IFN regulatory factor 4 regulates the expression of a subset of Th2 cytokines. *J Immunol* (2009) 183(3):1598–1596. doi:10.4049/jimmunol.0803302
85. Rossetto CC, Pari GS. Kaposi's sarcoma-associated herpesvirus noncoding polyadenylated nuclear RNA interacts with virus- and host cell-encoded proteins and suppresses expression of genes involved in immune modulation. *J Virol* (2011) 85(24):13290–7. doi:10.1128/JVI.05886-11

86. Rossetto CC, Pari GS. KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome. *PLoS Pathog* (2012) 8(5):e1002680. doi:10.1371/journal.ppat.1002680
87. Borah S, Darricarrere N, Darnell A, Myoung J, Steitz JA. A viral nuclear noncoding RNA binds re-localized poly(A) binding protein and is required for late KSHV gene expression. *PLoS Pathog* (2011) 7(10):e1002300. doi:10.1371/journal.ppat.1002300
88. Gatherer D, Seirafian S, Cunningham C, Holton M, Dargan DJ, Baluchova K, et al. High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A* (2011) 108(49):19755–60. doi:10.1073/pnas.1115861108
89. Poole E, Kuan WL, Barker R, Sinclair J. The human cytomegalovirus non-coding Beta2.7 RNA as a novel therapeutic for Parkinson's disease – translational research with no translation. *Virus Res* (2016) 212:64–9. doi:10.1016/j.virusres.2015.05.007
90. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, et al. Decoding human cytomegalovirus. *Science* (2012) 338(6110):1088–93. doi:10.1126/science.1227919
91. Bergamini G, Reschke M, Battista MC, Boccuni MC, Campanini F, Ripalti A, et al. The major open reading frame of the Beta2.7 transcript of human cytomegalovirus: *in vitro* expression of a protein posttranscriptionally regulated by the 5' region. *J Virol* (1998) 72(10):8425–9.
92. Reeves MB, Davies AA, McSharry BP, Wilkinson GW, Sinclair JH. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* (2007) 316(5829):1345–8. doi:10.1126/science.1142984
93. Lee JE, Lee JY, Tremblay J, Wilusz J, Tian B, Wilusz CJ. The PARN deadenylase targets a discrete set of mRNAs for decay and regulates cell motility in mouse myoblasts. *PLoS Genet* (2012) 8(8):e1002901. doi:10.1371/journal.pgen.1002901
94. Moon SL, Anderson JR, Kumagai Y, Wilusz CJ, Akira S, Khromykh AA, et al. A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability. *RNA* (2012) 18(11):2029–40. doi:10.1261/rna.034330.112
95. Moon SL, Blackinton JG, Anderson JR, Dozier MK, Dodd BJ, Keene JD, et al. XRN1 stalling in the 5' UTR of hepatitis C virus and bovine viral diarrhea virus is associated with dysregulated host mRNA stability. *PLoS Pathog* (2015) 11(3):e1004708. doi:10.1371/journal.ppat.1004708
96. Charley PA, Wilusz J. Standing your ground to exoribonucleases: function of flavivirus long non-coding RNAs. *Virus Res* (2016) 212:70–7. doi:10.1016/j.virusres.2015.09.009
97. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* (2013) 495(7441):333–8. doi:10.1038/nature11928
98. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA cross-talk and competition. *Nature* (2014) 505(7483):344–52. doi:10.1038/nature12986

Conflict of Interest Statement: 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.

Copyright © 2017 Meng, Luo, Anwar, Sun, Gao, Zhang, Munir and Qiu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Intra-Dependence of Viruses and the Holobiont

Juris A. Grasis^{1,2*}

¹ Department of Biology, San Diego State University, San Diego, CA, United States, ² School of Natural Sciences, University of California at Merced, Merced, CA, United States

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, United States

Reviewed by:

Jonathan L. Klassen,
University of Connecticut,
United States
Mercedes Berlanga,
University of Barcelona, Spain
Kevin R. Theis,
Wayne State University School of
Medicine, United States

*Correspondence:

Juris A. Grasis
jagrasis@ucmerced.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 29 April 2017

Accepted: 24 October 2017

Published: 09 November 2017

Citation:

Grasis JA (2017) The Intra-
Dependence of Viruses and the
Holobiont.
Front. Immunol. 8:1501.
doi: 10.3389/fimmu.2017.01501

Animals live in symbiosis with the microorganisms surrounding them. This symbiosis is necessary for animal health, as a symbiotic breakdown can lead to a disease state. The functional symbiosis between the host, and associated prokaryotes, eukaryotes, and viruses in the context of an environment is the holobiont. Deciphering these holobiont associations has proven to be both difficult and controversial. In particular, holobiont association with viruses has been of debate even though these interactions have been occurring since cellular life began. The controversy stems from the idea that all viruses are parasitic, yet their associations can also be beneficial. To determine viral involvement within the holobiont, it is necessary to identify and elucidate the function of viral populations in symbiosis with the host. Viral metagenome analyses identify the communities of eukaryotic and prokaryotic viruses that functionally associate within a holobiont. Similarly, analyses of the host in response to viral presence determine how these interactions are maintained. Combined analyses reveal how viruses interact within the holobiont and how viral symbiotic cooperation occurs. To understand how the holobiont serves as a functional unit, one must consider viruses as an integral part of disease, development, and evolution.

Keywords: holobiont, virome, symbiosis, viral metagenomics, host-microbe interactions, innate immunity, antiviral immunity, bacteriophage

INTRODUCTION

All animals interact with a consortium of microbes at all times and have done so since the dawn of animal life (1). Animal life has evolved from and in intimate association with microorganisms, while these same microorganisms have evolved in part to the resources provided by their animal surroundings. This symbiosis allows for a sharing of resources, including metabolic products and genes. These interactions have been of intense research and speculation; however, an important player in these symbiotic interactions is often overlooked, the effects of viruses. None of these interactions occur in the absence of viruses, so to inquire about symbioses requires discussion of viruses.

Viruses are seemingly universal in the biosphere (2). Their numbers are so staggering that when speaking of large numbers, one should use the term “viral” rather than “astronomical.” There are an estimated 10^{31} viruses on the planet, which may be an underestimation due to our inability to properly enumerate RNA viruses and viral elements that persist in cells and genomes (3). Further, viral genomes are worldwide reservoirs of genetic diversity (4). Considering viral abundances, diversity, and ubiquitous presence (5), understanding symbioses is lacking without taking into account the effects of viruses on host and associated microbe metabolism, and genetic flow between organisms.

Viruses infect all animals, from Poriferans to Cnidarians to Bilaterans to Chordates. There is ever-increasing evidence that viral infections have occurred during all of cellular life, as the presence of viral elements are often found in genomes throughout evolution (6). Host–viral infections or associations are not adequately quantified, but in most host-associated systems it seems that the number of viruses is equivalent to or slightly less than the number of bacteria associating with a eukaryotic host (2, 7). In most cases, the enumerable viral populations are the free DNA prokaryotic viruses, which are likely involved with the regulation of the host-associated bacteria. In host-associated systems, it seems that Lotka–Volterra “kill-the-winner” predator–prey dynamics of the prokaryotic virus and bacteria are atypical. Many prokaryotic viruses found in these systems display temperate lifestyles in which the virus becomes latent and integrates into a host chromosome or exists as an episomal element, as indicated by the large abundance of integrase genes in viral genomes (8, 9). Additionally, the presence of latent viruses may allow for bacterial dominance of a niche in the presence of related strains (10). Experimental evidence in non-host-associated systems supports this idea, as increasing concentrations of bacteria favor prokaryotic virus temperate lifestyles (11). While most viral research focuses on lytic/virulent infections, it is useful to explore both the temperate dynamics of prokaryotic viruses and latent eukaryotic viral infection, and their role in symbiosis.

The functional association between a host, prokaryotic, eukaryotic, and viral entities within a particular environment is the holobiont. This functional association helps to define the phenotypic unit. Casual associations may not define the phenotype, so functional associations (and the genes used) help define the phenotype. This functional symbiosis is involved in animal development (12), nervous system regulation (13, 14), immune system development and regulation (15, 16), and many other biological processes (17). When this functional association breaks down, a dysbiotic state occurs, leading to grave effects on animal health, ranging from coral bleaching (18), to stunted immune system development (19), to nervous and immunological disorders (20), to effects on human health (21). Further, the holobiont is not static; it is in a constant state of genetic flux. Viruses predominantly affect this genetic flow and the acquisition of evolutionary traits (22, 23). Therefore, understanding the holobiont requires investigation of the effects viruses have on gene flow occurring within it. This is evaluated through viral metagenomics (viromics), where culture-independent viral isolations from host systems are sequenced and the viral genomes are analyzed. Not only can host-associated viral populations be identified, but how these populations change under dysbiotic conditions (24, 25), the identification of new viruses (26, 27), and the effects these viruses have on cellular systems (28) can all be learned through viral metagenome (virome) analyses.

VIRAL SYMBIOSES AS PARASITISM

Viruses act as parasites; they infect and either replicate within the host cell or integrate within the host genome. Viruses propagate by one of two different lifestyles, either lytic/virulent or temperate/latent. The lytic/virulent lifestyle involves the infection,

replication, and lysis of the cell, leading to the death of the cell and release of viral progeny. The temperate/latent lifestyle involves the integration of the virus into the genome in a proviral form, which can be activated at a later time to become a lytic/virulent replicative virus. Either one of these scenarios affects the host; replication leads to cellular damage, while integration leads to genomic damage. The host defense against parasitism limits cellular or genomic damage (29). These viral parasitic lifestyles cause a molecular arms race, the virus seeking a new host to continue propagation, while the host immune system recognizes the virus to minimize damage (30).

There are many direct causes of pathogenesis by parasitic viruses, but there are many indirect causes as well. Proviral endogenous retroelements can have negative effects on the genome by inserting, deleting, or rearranging portions of the genome (31). The large number of freely associating viruses found interacting with host systems also presents a conundrum, that the presence of large amounts of viral material, be it nucleic acid or protein, makes it unlikely that they would not cause an immune response. Microbial-associated molecular patterns (MAMPs) on prokaryotic and eukaryotic viruses can cause immune system recognition that can lead to immune related pathogenesis. Further, lysis of cells, be it of a bacterial cell or of a eukaryotic cell, or apoptosis of a virally infected cell can cause activation of the immune system leading to pathogenesis (32). Cellular lysis is often considered in the aftermath of eukaryotic viral infection, but prokaryotic lysis of bacteria is commonly overlooked. Release of bacterial antigens, such as LPS, peptidoglycans, lipopeptides, lipoteichoic acid, flagellin, and bacterial DNA, can easily activate the immune system, and in extreme cases lead to sepsis (33). There are many direct and indirect causes of viral pathogenesis, but given the sheer numbers of viruses within a holobiont, and the limited pathogenesis that actually occurs, it seems more likely that viral pathogenesis is not as common as viral commensalism and mutualism.

VIRAL SYMBIOSES AS COMMENSALISM AND MUTUALISM

Most consider viruses to be parasites, where infection benefits the virus, but decreases the fitness of the host. Now consider other scenarios, such as commensals and mutualists. A virus can be commensal, the virus benefits while host fitness is unaffected. A virus can be mutualistic, in which both organisms benefit and fitness increases. Such viral associations may provide advantages that promote evolution and biodiversity (34, 35). Also consider that one virulent virus among a sea of non-virulent viruses does not equate to pathogenesis. Unless transmission and recovery rates are high, pathogenicity may be an evolutionarily poor strategy for viral survival. More likely, pathogenesis is the exception and not the rule, with more instances being discovered of viruses having cooperative roles with the host (34, 36).

There are many instances where an organism cannot exist without beneficial viruses. Polydnavirus integration into parasitoid wasp genomes counters the effects of the caterpillar host immune system where the wasp has laid its eggs (37). Without this polydnavirus presence, the caterpillar immune system would

eliminate the wasp eggs, but when the polydnavirus endogenous viral element becomes active upon egg deposition, the host immune response to the eggs is negated. Similarly, endogenous retrovirus syncytin expression in the placenta of mammals allows for the development of the placental syncytium (38). This syncytial fusion creates a barrier for the placenta, which in part keeps the fetus from being rejected by the mother's immune system. Viruses can also modulate the immune system and restore dysbiotic conditions. Kernbauer et al. have shown that an enteric murine norovirus can restore normal mucosal immunity and intestinal morphology in germ-free mice, essentially replacing the immune stimulatory effects of gut microbiota (39). Viruses can also protect against or impede further infection or pathogenesis, such as Hepatitis G virus slowing the progress of HIV infection (40), and latent herpesviruses protecting against bacterial infections (41). It is becoming evident that viruses have the potential to be something more than parasites in a holobiont, which revises conceptions of how viruses impact host interactions.

I AM ONE WITH THE VIRUSES, THE VIRUSES ARE WITH ME

Viruses can also integrate into cellular genomes and act as genetic elements associating with genomes. The amount of DNA of viral origin within the human genome is similar to that of human coding domains (42). One major discovery in viromics is the persistence of viral genetic elements, either latently integrated into host genomes or surviving as chromosomal episomes. Host-associated viral populations seem to be dominated by temperate prokaryotic viruses or latent eukaryotic viruses. This is attributed to a large abundance of integrase sequences in prokaryotic viromes (8) and a large abundance of transposase sequences in eukaryotic viromes (43).

Integrated viral DNA in the host genome are endogenous viral elements (EVEs), which have the potential to drive evolutionary processes, such as speciation, resulting in the emergence of new traits (44–46). In addition to these evolutionary transitions, EVE integration can affect gene expression through their long terminal repeats (LTRs). These LTRs are repetitive viral DNA sequences that flank integrated EVEs, serving as promoters to both viral and host genes. These LTRs can affect stem cells (47), development (48), and immunity (49, 50). There are many individual genes affected by EVEs, though their major impact on evolutionary traits may be on gene regulatory networks, or the cellular regulators that impact RNA and protein expression (51, 52). The effects of EVEs and transposable elements in all these biological processes are being recognized as vitally important (53).

Genomically integrated viral elements are reminders that viruses affect everything in biology, but what about free viruses that associate with hosts? Viromics allow researchers to analyze the viral populations and effects these viruses have on the holobiont. These studies have been conducted in many host systems, from the base of animal life in the Cnidarian phylum (54) to mammals (55). Often, the viruses found freely associating are prokaryotic viruses, which regulate the number and strains of bacteria in a holobiont (56). These viruses are

likely selected by the host to maintain bacterial populations (26). Further, viromics show the sphere of viral involvement in gene flow and gene shuffling in an ever-changing environment, often from within bacterial cells and sometimes from within eukaryotic cells.

THE ETERNAL STRUGGLE OF HOST–VIRAL INTERACTIONS

Many viruses can persist in host cells and influence the host without symptoms of disease. Chronic systemic viruses continuously stimulate the immune system (57), driving the emergence of many viral recognition systems over evolutionary time (58). These recognition systems give a host integrity to coexist with viruses while minimizing pathogenesis and protecting genomic information. Antisense RNA encoded by genomic transposable elements allows for specific regulation of viral amplification products (59). This evolved into use of antisense RNAs with Argonaute nucleases. Piwi-interacting RNAs utilize transposon derived small RNAs to defend against integration events by binding to complementary RNAs and cleaving the complex with a bound Argonaute nuclease. This system seems to be restricted to the germ-line and protects genomic integrity. Similarly, the RNAi system processes RNAs by binding to small RNA fragments and cleaving these complexes with an RNase III nuclease, Dicer (60). While controversial, it appears that chordates may not have retained RNAi antiviral function. However, there are many immune functions additionally used in both chordates and non-chordates to regulate viral presence (**Figure 1**). These systems rely on host pattern-recognition receptors (PRRs) evolved to recognize MAMPs. These PRRs include the Toll-like receptors (TLRs), retinoic acid-inducible gene I (RIG-I)-like receptors (RLRs), cGAS-STING pathway, NOD-like receptors (NLRs), C-type lectin receptors (CLRs), and absent-in-melanoma-like receptors (ALRs). TLRs recognize viruses endosomally once viral nucleic acids are released (61), cytoplasmic RLRs recognize viral genomic RNA or double-stranded RNA intermediates (62), cGAS-STING senses retroviral and double-stranded DNA (63), NLRs recognize viral DNA genomes (64), ALRs can also recognize viral genomic DNA (65), while CLRs recognize carbohydrates (66). In the biological arms race that caused the development of the adaptive immune system capable of tracking evolutionary changes in pathogens, antiviral cytokines such as interferons (IFNs) became prominent signals alerting the host of viral infection and inhibit viral propagation (67). With IFNs came recombination events to generate antibodies and major histocompatibility complexes in vertebrates to increase the recognition possibilities that came with increased pathogen complexity. Although viral recognition research is often focused on the adaptive immune system in mammals, the overwhelming majority of animals has multiple pathways to recognize, regulate, and maintain viral associations and may not necessarily use canonical adaptive systems to structure the holobiont. Continuing research will involve the 95% of Metazoans that do not possess such an adaptive immune system to recognize viruses, yet are able to adapt to ever-changing viral populations through mechanisms, such as trained innate immunity (68).

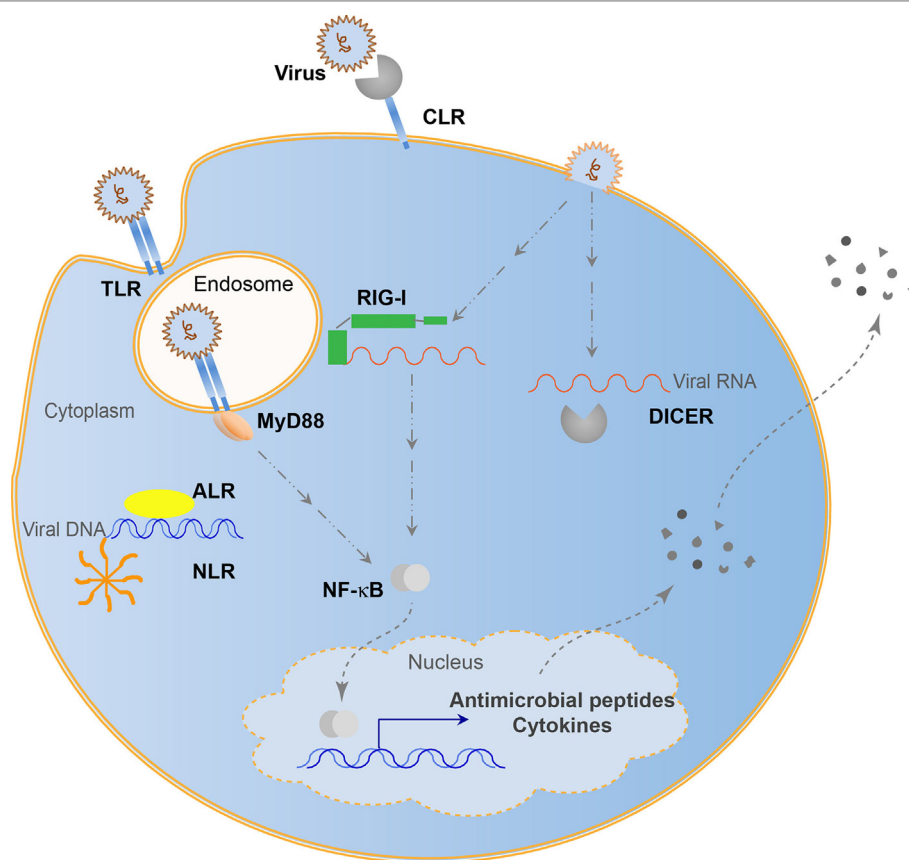


FIGURE 1 | Evolutionarily conserved antiviral innate immune systems. Toll-like receptors (TLRs) recognize endosomal viral nucleic acids, NOD-like receptors (NLRs) form an inflammasome and recognize viral DNA, absent-in-melanoma-like receptors (ALRs) recognize viral DNA, retinoic acid-inducible gene I (RIG-I) and RNAi (Dicer) pathways recognize viral RNA, while C-type lectins (CLRs) recognize viral carbohydrates. Some pathways can lead to the direct elimination of viral entities, while others lead to transcriptional activation resulting in cytokine and antimicrobial peptide secretion.

HAIL *Hydra*: THE IMPORTANCE OF A SIMPLE MODEL SYSTEM TO EVALUATE HOLOBIONT INTERACTIONS

Holobiont studies are complex. If one considers the sheer number of associated prokaryotes, eukaryotes, viruses, and all of their respective genomes, the number of potential interactions is overwhelming. Therefore, if one can use a model system with a limited number of microbial partners to deconstruct the holobiont and if this can be studied in an ancient animal phylum for conserved holobiont interactions, it could simplify these studies while retaining informative and predictive capabilities. The use of a basal metazoan allows research on mechanisms of holobiont assembly, holobiont effects on microbiota and host health, and metabolic interactions between the host and microbiota. This helps to elucidate symbiosis in healthy states and dysbiosis in disease states.

There are many useful systems that meet the above criteria to investigate the holobiont, including ascidians (69), anemones (70), and sponges (71). The basal model organism *Hydra* is another useful system. *Hydra* are freshwater Cnidarians practical for developmental, neural, aging, and stem cell studies

(72). Importantly, the findings made using *Hydra* translate well into host–microbe interaction studies due to its diploblastic morphology (73), conserved mucosal immunity (74), and limited number of microbial partners (75). Additionally, *Hydra* are clonal, have a well-annotated genome (76), can be made transgenic (77), germ-free (78), and due to its limited number of microbial interactions, *Hydra* can be used in symbiosis studies (79). *Hydra* display distinct microbial colonization patterns dependent on host factors (78), which are primarily driven by antimicrobial peptide selection at the epithelium (80). *Hydra* have many evolutionarily conserved receptor pathways to regulate microbial interactions, including a TLR pathway (81) and a large repertoire of NLRs (82). Further, *Hydra* utilize many uniquely identified classes of antimicrobial peptides to regulate its microbial interactions (81, 83, 84). Finally, 57% of the *Hydra* genome are transposable elements, one of the largest percentages found in an animal genome (76). These factors make *Hydra* a useful system to deconstruct and reconstruct an organismal holobiont (Figure 2).

Understanding the complete *Hydra*-associated virome has commenced. The *Hydra* DNA virome consists primarily of prokaryotic viruses in the Caudovirales order, the majority of

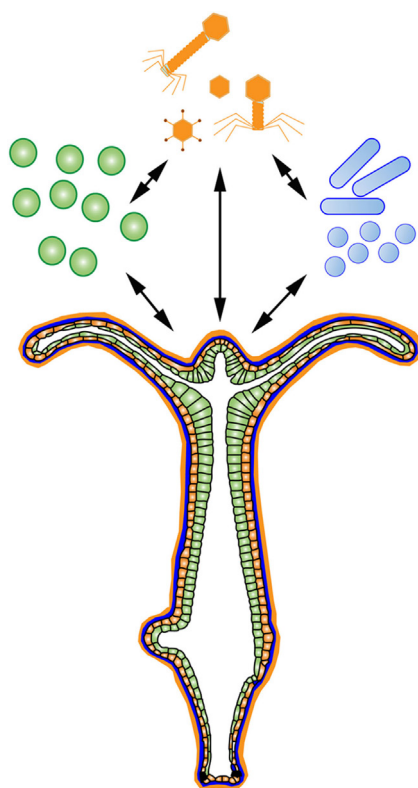


FIGURE 2 | The *Hydra* Holobiont. *Hydra* are an ideal system to deconstruct and reconstruct an organismal holobiont consisting of associated eukaryotes (green), prokaryotes (blue), and viruses (orange) at an exposed epithelium.

the eukaryotic viruses are of the Herpesviridae family, the diversity of the viruses increases upon environmental heat stress, and each species of *Hydra* associates with a specific community of viruses (25). Further, these *Hydra*-associated viruses affect *Hydra*-microbiome metabolism (25, 85). Studies on the RNA virome, germ-free eukaryotic virome, and prokaryotic virome of *Hydra*-associated bacteria are ongoing to create a comprehensive *Hydra* virome [J. Grasis, in preparation; (86)]. Combining the virome with *in vivo* viral infection transcriptomes and the ability to induce inflammatory conditions makes *Hydra* a useful system to structure viral–holobiont interactions related to animal health conditions. The *Hydra* model system may shed light on novel aspects of holobiont formation, maintenance, and dysbiosis, while integrating viral involvement within the holobiont.

VIRUSES BRING BALANCE TO THE HOLOBIONT

There has been much discussion about the holobiont recently, particularly as it relates to selective units of animal host and microbiome (87–93). Much of the focus has been placed on host–bacterial associations, but what of the viruses? They are intrinsically part of the genome and part of the holobiont, and

yet, extrinsically exist beyond the genome and the holobiont. This duality exists because both the host and the microbiome are under their own selective pressures, each are selecting for the environment that benefits them, establishing or propagating a phenotype, and allowing for co-existence to continue. It is neither eukaryo-centric, prokaryo-centric, nor viro-centric, each member has a role to play within the holobiont. Therefore, the holobiont is a coordination of integrated functions by all members to suit adaptation to an environment.

Viruses are genetic parasites constantly sampling their environments. Functional aspects of their genomes can be selected by their prokaryotic and eukaryotic hosts, and in this way, viruses are symbionts to these hosts. Viruses can also transfer DNA in the form of lateral gene transfer, which can be important for adaptations to new environments (94). For example, prophages can promote genetic transfer between prokaryotic viruses and eukaryotes. *Wolbachia* prophage WO in arthropods contains a eukaryotic association module, which among other genes, contains a spider toxin gene that can form pores in both prokaryotic and eukaryotic membranes to facilitate viral escape (95). There are many more instances of viral drivers of adaptation (96), which makes viral dynamics in the holobiont fluid. Free viruses can be acquired from the environment through horizontal transmission, while viral elements can be vertically transmitted through genomically integrated viral elements and episomes. Such horizontal and vertical transmissions allow for a fully functional range of symbioses, from obligate (both need each other to survive) to facultative (both benefit from the association, but it is not absolute).

Viruses are remarkable symbionts. Viral elements exist intra-genomically, intra-cellularly, extra-cellularly, and environmentally. They persist in all of these realms, and yet, are vital to the holobiont. As mentioned earlier, viromics teaches us that viruses are involved in gene flow and shuffling in a changing environment, and that the elements in the holobiont are in a constant ecological flux. In all cases, viruses provide balance to the holobiont, keeping the host and associating prokaryotes and eukaryotes functioning together as a unit.

AUTHOR CONTRIBUTIONS

JG wrote, did the artwork, and is responsible for the content of this manuscript.

ACKNOWLEDGMENTS

JG would like to thank Dr. Benjamin Knowles for critical reading of the manuscript and to the reviewers of the manuscript who provided fantastic feedback. Support for JG is through NIH F32AI098418. JG would also like to acknowledge the many related articles not cited here due to word restrictions, particularly in plant research areas, which have been vital in our burgeoning understanding of viral symbioses.

FUNDING

Support for JG is through NIH F32AI098418.

REFERENCES

- Moran NA. Symbiosis. *Curr Biol* (2006) 16:R866–71. doi:10.1016/j.cub.2006.09.019
- Cobián-Güemes AG, Youle M, Cantú VA, Felts B, Nulton J, Rohwer F. Viruses are winners in the game of life. *Annu Rev Virol* (2016) 3:197–214. doi:10.1146/annurev-virology-100114-054952
- Suttle CA. Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol* (2007) 5:801–12. doi:10.1038/nrmicro1750
- Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering earth's virome. *Nature* (2016) 536:425–30. doi:10.1038/nature19094
- Wasik BR, Turner PE. On the biological success of viruses. *Annu Rev Microbiol* (2013) 67:519–41. doi:10.1146/annurev-micro-090110-102833
- Katzourakis A. Paleovirology: inferring viral evolution from host genome sequence data. *Philos Trans R Soc Lond B Biol Sci* (2013) 368:20120493. doi:10.1098/rstb.2012.0493
- Kim M-S, Park E-J, Roh SW, Bae J-W. Diversity and abundance of single-stranded DNA viruses in human feces. *Appl Environ Microbiol* (2011) 77:8062–70. doi:10.1128/AEM.06331-11
- Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* (2010) 466:334–40. doi:10.1038/nature09199
- Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* (2011) 21:1616–25. doi:10.1101/gr.122705.111
- Duerkop BA, Clements CV, Rollins D, Rodrigues JL, Hooper LV. A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proc Natl Acad Sci U S A* (2012) 109:17621–6. doi:10.1073/pnas.1206136109
- Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, et al. Lytic to temperate switching of viral communities. *Nature* (2016) 531:466–70. doi:10.1038/nature17193
- Gilbert SF, Bosch TCG, Ledon-Rettig C. Eco-evo-devo: developmental symbiosis and developmental plasticity as evolutionary agents. *Nat Rev Genet* (2015) 16:611–22. doi:10.1038/nrg3982
- Cryan JF, Dinan TG. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat Rev Neurosci* (2012) 13:701–12. doi:10.1038/nrn3346
- Eisthen HL, Theis KR. Animal-microbe interaction and the evolution of nervous systems. *Philos Trans R Soc Lond B Biol Sci* (2016) 371:20150052. doi:10.1098/rstb.2015.0052
- Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol* (2016) 16:341–52. doi:10.1038/nri.2016.42
- Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. *Nature* (2016) 535:65–74. doi:10.1038/nature18847
- Dheilly NM. Holobiont-holobiont interactions: redefining host-parasite interactions. *PLoS Pathog* (2014) 10:e1004093. doi:10.1371/journal.ppat.1004093
- Egan S, Gardiner M. Microbial dysbiosis: rethinking disease in marine ecosystems. *Front Microbiol* (2016) 7:991. doi:10.3389/fmicb.2016.00991
- Lee YK, Mazmanian SK. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* (2010) 330:1768–73. doi:10.1126/science.1195568
- Fung TC, Olson CA, Hsiao EY. Interactions between the microbiota, immune and nervous systems in health and disease. *Nat Neurosci* (2017) 20:145–55. doi:10.1038/nn.4476
- Levy M, Kolodziejczyk AA, Thaiss CA, Elinav E. Dysbiosis and the immune system. *Nat Rev Immunol* (2017) 17:219–32. doi:10.1038/nri.2017.7
- Poole AM. Horizontal gene transfer and the earliest stages of the evolution of life. *Res Microbiol* (2009) 160:473–80. doi:10.1016/j.resmic.2009.07.009
- Boto L. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc Biol Sci* (2014) 281:20132450. doi:10.1098/rspb.2013.2450
- Vega Thurber RL, Barrott KL, Hall D, Liu H, Rodriguez-Mueller B, Desnues C, et al. Metagenomic analysis indicates that stressors induce production of herpes-like viruses in the coral *Porites compressa*. *Proc Natl Acad Sci U S A* (2008) 105:18413–8. doi:10.1073/pnas.0808985105
- Grasis JA, Lachnit T, Anton-Erxleben F, Lim YM, Schmieder R, Fraune S, et al. Species-specific viromes in the ancestral holobiont *Hydra*. *PLoS One* (2014) 9:e109952. doi:10.1371/journal.pone.0109952
- Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* (2012) 2:63–77. doi:10.1016/j.coviro.2011.12.004
- Dutilh BE, Cassman N, McNair K, Sanchez SE, Silva GG, Boling L, et al. A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nat Commun* (2014) 5:4498. doi:10.1038/ncomms5498
- Quistad SD, Lim YW, Silva GG, Nelson CE, Haas AF, Kelly LW, et al. Using viromes to predict novel immune proteins in non-model organisms. *Proc Biol Sci* (2016) 283:20161200. doi:10.1098/rspb.2016.1200
- Aiewsakun P, Katzourakis A. Endogenous viruses: connecting recent and ancient viral evolution. *Virology* (2015) 479–480:26–37. doi:10.1016/j.virol.2015.02.011
- 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
- Volkman HE, Stetson DB. The enemy within: endogenous retroelements and autoimmune disease. *Nat Immunol* (2014) 15:415–22. doi:10.1038/ni.2872
- Duerkop BA, Hooper LV. Resident viruses and their interactions with the immune system. *Nat Immunol* (2013) 14:654–9. doi:10.1038/ni.2614
- Wiersinga WJ, Leopold SJ, Cranendonk DR, van der Poll T. Host innate immune responses to sepsis. *Virulence* (2014) 5:36–44. doi:10.4161/viru.25436
- Roossinck MJ. Move over bacteria! Viruses make their mark as mutualistic microbial symbionts. *J Virol* (2015) 89:6532–5. doi:10.1128/JVI.02974-14
- Pradeu T. Mutualistic viruses and the heteronomy of life. *Stud Hist Philos Biol Biomed Sci* (2016) 59:80–8. doi:10.1016/j.shpsc.2016.02.007
- Roossinck MJ. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol* (2011) 9:99–108. doi:10.1038/nrmicro2491
- Herniou EA, Huguet E, Théze J, Bézier A, Periquet G, Drezén JM. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnviruses. *Philos Trans R Soc Lond B Biol Sci* (2013) 368:20130051. doi:10.1098/rstb.2013.0051
- Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, et al. Syncytin-A knockout mice demonstrate the critical role in placentalization of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc Natl Acad Sci U S A* (2009) 106:12127–32. doi:10.1073/pnas.0902925106
- Kernbauer E, Ding Y, Cadwell K. An enteric virus can replace the beneficial function of commensal bacteria. *Nature* (2014) 516:94–8. doi:10.1038/nature13960
- Tillmann HL, Heiken H, Knapik-Botor A, Heringlake S, Ockenga J, Wilber JC, et al. Infection with GB virus C and reduced mortality among HIV-infected patients. *N Engl J Med* (2001) 345:715–24. doi:10.1056/NEJMoa010398
- Barton ES, White DW, Cathelyn JS, Brett-McClellan KA, Engle M, Diamond MS, et al. Herpesvirus latency confers beneficial protection from bacterial infection. *Nature* (2007) 447:326–9. doi:10.1038/nature05762
- Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* (2012) 13:283–96. doi:10.1038/nrg3199
- Aziz RK, Breitbart M, Edwards RA. Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res* (2010) 38:4207–17. doi:10.1093/nar/gkq140
- Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet* (2008) 42:709–32. doi:10.1146/annurev.genet.42.110807.091501
- Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. *PLoS Genet* (2010) 6:e1001191. doi:10.1371/journal.pgen.1001191
- Huang CRL, Burns KH, Boeke JD. Active transposons in genomes. *Annu Rev Genet* (2012) 46:651–75. doi:10.1146/annurev-genet-110711-155616
- Schlesinger S, Goff SP. Retroviral transcriptional regulation and embryonic stem cells: war and peace. *Mol Cell Biol* (2015) 35:770–7. doi:10.1128/MCB.01293-14
- Emera D, Wagner GP. Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Brief Funct Genomics* (2012) 11:267–76. doi:10.1093/bfpg/els013
- Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* (2016) 351:1083–7. doi:10.1126/science.125497
- Kassiotis G, Stoye JP. Immune response to endogenous retroelements: taking the bad with the good. *Nat Rev Immunol* (2016) 16:207–19. doi:10.1038/nri.2016.27
- Shubin N, Tabin C, Carroll S. Deep homology and the origins of evolutionary novelty. *Nature* (2009) 457:818–23. doi:10.1038/nature0791

52. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* (2012) 46:21–42. doi:10.1146/annurev-genet-110711-155621
53. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* (2017) 18:71–86. doi:10.1038/nrg.2016.139
54. Vega Thurber R, Payet JP, Thurber AR, Correa AMS. Virus-host interactions and their roles in coral reef health and disease. *Nat Rev Microbiol* (2017) 15:205–16. doi:10.1038/nrmicro.2016.176
55. Virgin HW. The virome in mammalian physiology and disease. *Cell* (2014) 157:142–50. doi:10.1016/j.cell.2014.02.032
56. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasić L, Thingstad TF, Rohwer F, et al. Explaining microbial population genomics through phage predation. *Nat Rev Microbiol* (2009) 7:828–36. doi:10.1038/nrmicro2235
57. Virgin HW, Wherry EJ, Ahmed R. Redefining chronic viral infection. *Cell* (2009) 238:30–50. doi:10.1016/j.cell.2009.06.036
58. tenOever BR. The evolution of antiviral defense systems. *Cell Host Microbe* (2016) 19:142–9. doi:10.1016/j.chom.2016.01.006
59. Zhou R, Rana TM. RNA-based mechanisms regulating host-virus interactions. *Immunol Rev* (2013) 253:97–111. doi:10.1111/imr.12053
60. Ding SW. RNA-based antiviral immunity. *Nat Rev Immunol* (2010) 10:632–44. doi:10.1038/nri2824
61. Kawai T, Akira S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* (2011) 34:637–50. doi:10.1016/j.immuni.2011.05.006
62. Wu J, Chen ZJ. Innate immune sensing and signaling of cytosolic nucleic acids. *Annu Rev Immunol* (2014) 32:461–88. doi:10.1146/annurev-immunol-032713-120156
63. Margolis SR, Wilson SC, Vance RE. Evolutionary origins of cGAS-STING signaling. *Trends Immunol* (2017) 38:733–43. doi:10.1016/j.it.2017.03.004
64. Motta V, Soares F, Sun T, Philpott DJ. NOD-like receptors: versatile cytosolic sentinels. *Physiol Rev* (2015) 95:149–78. doi:10.1152/physrev.00009.2014
65. McFadden MJ, Gokhale NS, Horner SM. Protect this house: cytosolic sensing of viruses. *Curr Opin Virol* (2017) 22:36–43. doi:10.1016/j.coviro.2016.11.012
66. Geijtenbeek TBH, Gringhuis SI. Signalling through C-type lectin receptors: shaping immune responses. *Nat Rev Immunol* (2009) 9:465–79. doi:10.1038/nri2569
67. Fensterl V, Chattopadhyay S, Sen GC. No love lost between viruses and interferons. *Annu Rev Virol* (2015) 2:549–72. doi:10.1146/annurev-virology-100114-055249
68. Quintin J, Cheng SC, van der Meer JW, Netea MG. Innate immune memory: towards a better understanding of host defense mechanisms. *Curr Opin Immunol* (2014) 29:1–7. doi:10.1016/j.coi.2014.02.006
69. Dishaw LJ, Flores-Torres J, Lax S, Gemayel K, Leigh B, Melillo D, et al. The gut of geographically disparate *Ciona intestinalis* harbors a core microbiota. *PLoS One* (2014) 9:e93386. doi:10.1371/journal.pone.0093386
70. Mortzfeld BM, Urbanski S, Reitzel AM, Künzel S, Technau U, Fraune S. Response of bacterial colonization in *Nematostella vectensis* to development, environment and biogeography. *Environ Microbiol* (2015) 18:1764–81. doi:10.1111/1462-2920.12926
71. Thomas T, Moitinho-Silva L, Lurgi M, Björk JR, Easson C, Astudillo-García C, et al. Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* (2016) 7:11870. doi:10.1038/ncomms11870
72. Galliot B. *Hydra*, a fruitful model system for 270 years. *Int J Dev Biol* (2012) 56:411–23. doi:10.1387/ijdb.120094bg
73. Bosch TCG. Cnidarian-microbe interactions and the origin of innate immunity in metazoans. *Annu Rev Microbiol* (2013) 67:499–518. doi:10.1146/annurev-micro-092412-155626
74. Schröder K, Bosch TCG. The origin of mucosal immunity: lessons from the holobiont *Hydra*. *MBio* (2016) 7:e1184–1116. doi:10.1128/mBio.01184-16
75. Fraune S, Bosch TCG. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan *Hydra*. *Proc Natl Acad Sci U S A* (2007) 104:13146–51. doi:10.1073/pnas.0703375104
76. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, et al. The dynamic genome of *Hydra*. *Nature* (2010) 464:592–6. doi:10.1038/nature08830
77. Wittlieb J, Khalturin K, Lohmann JU, Anton-Erxleben F, Bosch TCG. Transgenic *Hydra* allow in vivo tracking of individual stem cells during morphogenesis. *Proc Natl Acad Sci U S A* (2006) 103:6208–11. doi:10.1073/pnas.0510163103
78. Franzenburg S, Fraune S, Künzel S, Baines JF, Domazet-Lozo T, Bosch TCG. MyD88-deficient *Hydra* reveal an ancient function of TLR signaling in sensing bacterial colonizers. *Proc Natl Acad Sci U S A* (2012) 109:19374–9. doi:10.1073/pnas.1213110109
79. Kovacevic G. Value of the *Hydra* model system for studying symbiosis. *Int J Dev Biol* (2012) 56:627–35. doi:10.1387/ijdb.123510gk
80. Franzenburg S, Walter J, Künzel S, Wang J, Baines JF, Bosch TC, et al. Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc Natl Acad Sci U S A* (2013) 110:E3730–8. doi:10.1073/pnas.1304960110
81. Bosch TCG, Augustin R, Anton-Erxleben F, Fraune S, Hemmrich G, Zill H, et al. Uncovering the evolutionary history of innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defence. *Dev Comp Immunol* (2009) 33:559–69. doi:10.1016/j.dci.2008.10.004
82. Lange C, Hemmrich G, Klostermeier UC, Lopez-Quintero JA, Miller DJ, Rahn T, et al. Defining the origins of the NOD-like receptor system at the base of animal evolution. *Mol Biol Evol* (2011) 28:1687–702. doi:10.1093/molbev/msq349
83. Augustin R, Anton-Erxleben F, Jungnickel S, Hemmrich G, Spudy B, Podschun R, et al. Activity of the novel peptide arminin against multidrug-resistant human pathogens shows the considerable potential of phylogenetically ancient organisms as drug sources. *Antimicrob Agents Chemother* (2009) 53:5245–50. doi:10.1128/AAC.00826-09
84. Fraune S, Augustin R, Anton-Erxleben F, Wittlieb J, Gelhaus C, Klimovich VB, et al. In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proc Natl Acad Sci U S A* (2010) 107:18067–72. doi:10.1073/pnas.1008573107
85. Bosch TCG, Grasis JA, Lahnit T. Microbial ecology in *Hydra*: why viruses matter. *J Microbiol* (2015) 53:193–200. doi:10.1007/s12275-015-4695-2
86. Deines P, Bosch TCG. Transitioning from microbiome composition to microbial community interactions: the potential of the metaorganism *Hydra* as an experimental model. *Front Microbiol* (2016) 7:1610. doi:10.3389/fmicb.2016.01610
87. Bordenstein SR, Theis KR. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. *PLoS Biol* (2015) 13:e1002226. doi:10.1371/journal.pbio.1002226
88. Douglas AE, Werren JH. Holes in the hologenome: why host-microbe symbioses are not holobionts. *MBio* (2016) 7:e2099. doi:10.1128/mBio.02099-15
89. Hester ER, Barott KL, Nulton J, Vermeij MJ, Rohwer FL. Stable and sporadic symbiotic communities of coral and algal holobionts. *ISME J* (2016) 10:1157–69. doi:10.1038/ismej.2015.190
90. Moran NA, Sloan DB. The hologenome concept: helpful or hollow? *PLoS Biol* (2015) 13:e1002311. doi:10.1371/journal.pbio.1002311
91. Rosenberg E, Zilber-Rosenberg I. Microbes drive evolution of animals and plants: the hologenome concept. *MBio* (2016) 7:e01395. doi:10.1128/mBio.01395-15
92. Shropshire JD, Bordenstein SR. Speciation by symbiosis: the microbiome and behavior. *MBio* (2016) 7:e01785. doi:10.1128/mBio.01785-15
93. Theis KR, Dheilly NM, Klassen JL, Brucker RM, Baines JF, Bosch TCG, et al. Getting the hologenome concept right: an eco-evolutionary framework for hosts and their microbiome. *mSystems* (2016) 1:e28–16. doi:10.1128/mSystems.00028-16
94. Gilbert C, Cordaux R. Viruses as vectors of horizontal transfer of genetic material in eukaryotes. *Curr Opin Virol* (2017) 25:16–22. doi:10.1016/j.coviro.2017.06.005
95. Bordenstein SR, Bordenstein SR. Eukaryotic association module in phage WO genomes from *Wolbachia*. *Nat Commun* (2016) 7:13155. doi:10.1038/ncomms13155
96. Koonin EV. Viruses and mobile elements as drivers of evolutionary transitions. *Philos Trans R Soc Lond B Biol Sci* (2016) 371:20150442. doi:10.1098/rstb.2015.0442

Conflict of Interest Statement: The 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.

Copyright © 2017 Grasis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



An Organismal Model for Gene Regulatory Networks in the Gut-Associated Immune Response

Katherine M. Buckley^{1*} and Jonathan P. Rast^{2,3,4*}

¹ Department of Biological Sciences, The George Washington University, Washington, DC, United States, ² Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, United States, ³ Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada, ⁴ Department of Immunology, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, United States

Reviewed by:

Lisa Rizzetto,
Fondazione Edmund Mach, Italy
Jeffrey A. Yoder,
North Carolina State
University, United States

*Correspondence:

Katherine M. Buckley
kshank@gwu.edu;
Jonathan P. Rast
jprast@emory.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 06 August 2017

Accepted: 27 September 2017

Published: 23 October 2017

Citation:

Buckley KM and Rast JP (2017)
An Organismal Model for
Gene Regulatory Networks
in the Gut-Associated
Immune Response.
Front. Immunol. 8:1297.
doi: 10.3389/fimmu.2017.01297

The gut epithelium is an ancient site of complex communication between the animal immune system and the microbial world. While elements of self-non-self receptors and effector mechanisms differ greatly among animal phyla, some aspects of recognition, regulation, and response are broadly conserved. A gene regulatory network (GRN) approach provides a means to investigate the nature of this conservation and divergence even as more peripheral functional details remain incompletely understood. The sea urchin embryo is an unparalleled experimental model for detangling the GRNs that govern embryonic development. By applying this theoretical framework to the free swimming, feeding larval stage of the purple sea urchin, it is possible to delineate the conserved regulatory circuitry that regulates the gut-associated immune response. This model provides a morphologically simple system in which to efficiently unravel regulatory connections that are phylogenetically relevant to immunity in vertebrates. Here, we review the organism-wide cellular and transcriptional immune response of the sea urchin larva. A large set of transcription factors and signal systems, including epithelial expression of interleukin 17 (IL17), are important mediators in the activation of the early gut-associated response. Many of these have homologs that are active in vertebrate immunity, while others are ancient in animals but absent in vertebrates or specific to echinoderms. This larval model provides a means to experimentally characterize immune function encoded in the sea urchin genome and the regulatory interconnections that control immune response and resolution across the tissues of the organism.

Keywords: inflammation, pigment cells, interleukin 17, gut immunology, phagocytosis, echinodermata, larva, sea urchins

The enormous progress made in the recent years in the field of pathology will surely also fertilize the field of pure zoology and at the same time the evolutionary standpoint of the latter field can provide solutions to medical problems in a comparative pathologic way. [Elya Metchnikoff (1)]

CONSERVATION AND INNOVATION IN ANIMAL IMMUNITY

Immune systems mediate complex interactions between animal hosts and a community of microbes that includes both pathogenic and beneficial strains (2). These ongoing processes occur in cells and tissues that are located across the animal and must be regulated at an organism-wide scale. In this context, immune response can be described as a distributed network of interconnecting

regulatory circuits that are coordinated to protect the host and stabilize interactions with microbiota. Given its central role in animal life, this integrated circuitry is, at some levels, subject to deep evolutionary conservation (3, 4). Consequently, causal connections gathered from experiments in morphologically simple invertebrate models have direct implications for understanding immunity in more complex vertebrates.

Most bilaterians harbor specialized immune cells that exhibit morphological or behavioral similarities (5, 6). One well-known example is the phylogenetically widespread phagocytic cells, which were first recognized and described in several invertebrates by Metchnikoff (7, 8). Dedicated phagocytes often exhibit similar motility and surveillance-like behaviors in different phyla. Many animal lineages also have granular cells that participate in immune sensing and control (5). Through intricately coordinated signaling mechanisms, these cell types cooperate to initiate and resolve immune response. In addition, immune cells express many rapidly evolving proteins such as non-self recognition receptors (9) and secreted effector molecules (10). Nonetheless, the characteristics of terminally differentiated immune cells (morphology and behavior) cannot be used to reliably infer evolutionary relationships among cell lineages. Instead, homology likely lies beyond cell lineages when comparing immunity in different phyla (i.e., the relevant unit of homology that is useful for understanding immune system evolution is likely to more often lie at the level of the regulatory subcircuitry within cells). Evolutionary pressure on immune systems manifests differently among gene types (11) but, in general, immune receptors and effectors tend to evolve quickly, and their relationships among phyla can be difficult to interpret. The regulatory circuitry that controls cell development and function can provide insight into this problem by defining the nature of homology in these systems across phyla.

ECHINODERM LARVAE: A NOT SO NOVEL MODEL SYSTEM IN IMMUNOLOGY

Echinoderms, together with the hemichordates, form a sister group to Chordata at the base of the deuterostomes (12). This evolutionary distance [echinoderms and chordates diverged ~530 million years ago (13)] provides the opportunity to investigate varying scales of immune system evolution, including (1) common mechanisms that regulate immunity throughout the deuterostomes, (2) ancestral strategies present in invertebrate deuterostomes or throughout Bilateria but specifically lost in vertebrates, and (3) evolutionary innovations that are specific to echinoderms. Examples of all three are evident in the sea urchin larval immune system.

Most sea urchins have biphasic life histories that include relatively long-lived, morphologically simple, planktonic larval stages. This form of development is ancestral to echinoderms (14). In the purple sea urchin (*Strongylocentrotus purpuratus*), a single female produces millions of eggs that, once fertilized, synchronously develop over 5 days into a free swimming, pluteus larva that feed for about 2 months before metamorphosis into a benthic juvenile form [reviewed in Ref. (15)]. Larvae have

a tripartite gut composed of an epithelial monolayer (16) and a cellular immune system of 80–150 mesenchymal cells that populate the blastocoel or are apposed to the ectodermal epithelia (17, 18). From an experimental standpoint, echinoderm larvae offer several advantages: transparency that enables organism-wide, *in vivo* imaging at single-cell resolution, and efficient transgenic strategies to precisely perturb protein function (19, 20). These characteristics can be exploited to investigate open questions in immunology.

A WEALTH OF ECHINODERM GENOMIC RESOURCES IS AVAILABLE

Experimental studies in echinoderms are supported by an extensive collection of genomic resources [www.echinobase.org (21)]. The purple sea urchin was the subject of the first assembled genome from an outbred, motile marine invertebrate and the largest invertebrate genome (814 Mb) sequenced at the time (22). Analysis of the *S. purpuratus* genome sequence identified many features previously believed to be vertebrate specific that were instead deuterostome or bilaterian innovations. One of the most striking findings was the expansive repertoire of genes encoding proteins with roles in immune recognition and defense (22–24).

Specifically, *S. purpuratus* has orthologs of most major transcription factor subfamilies important in vertebrate immunity (23). These include factors that regulate gene expression in the course of immune response (e.g., NF- κ B and IRF), as well as regulators of vertebrate hematopoiesis (25–27). Many homologs of vertebrate cytokines are absent, which is not surprising given the rapid evolution of these factors and their receptors even among vertebrates (28). However, the genome sequence contains homologs of tumor necrosis factor α , macrophage inhibitory factor and interleukin 17 (IL17), as well as IL1 receptors (23). This shared regulatory heritage between echinoderms and vertebrates enables experimental investigations into transcriptional control of immune cell development (25, 26) and immune response (17, 29) that can provide meaningful insight to vertebrate biology.

In contrast to this conservation, the *S. purpuratus* genome sequence contains surprisingly large families of genes that encode pattern recognition receptors. The repertoires of toll-like receptors (TLRs), NOD-like receptors, and proteins containing multiple scavenger receptor cysteine rich domains are significantly (~10-fold) larger than their counterparts in the well-characterized vertebrates and insects (23, 24, 30–32). The sea urchin TLRs form 10 subfamilies based on phylogenetic analysis (33). Genes within these subfamilies are differentially expressed in larval and adult tissues and are most highly expressed in the coelomocytes and gut tissue, which are both sites of dynamic immune activity. Residues predicted to be in close spatial proximity are subject to strong positive selection. The expression patterns, rapid evolution, and lack of expression during early development strongly suggest an immune role for the sea urchin TLRs (33). These and other immune innovations within the echinoderm lineage [e.g., the transformer (Trf, 185/333)]

proteins; reviewed in Ref. (34)] highlight the diversification of proteins that potentially interact directly with pathogens, as has been observed in other systems (11) and provide a rich platform to study the integration of these quickly evolving proteins with more conserved elements of regulatory circuitry.

The Sea Urchin Genome Project has also assembled genome sequences from two additional sea urchins, a sea star, sea cucumber, and brittle star (www.echinobase.org). Four high-quality and three less complete genome assemblies, as well as high coverage, unassembled whole genome sequencing reads are available from other echinoderm species (35–38). In total, the NCBI Short Read Archive hosts 206 projects in 75 echinoderm species that cover all five classes as of this writing. Collectively, these data provide deep coverage and broad evolutionary perspective for investigations of echinoderm immunity.

SEVERAL CELL TYPES MEDIATE THE LARVAL IMMUNE RESPONSE

To understand how this genomic complexity is deployed *in vivo*, immune response has been investigated in sea urchin adults [reviewed in Ref. (39)], as well as the embryonic and larval stages (17). Early life stages offer significant experimental advantages for characterizing the gene regulatory networks (GRNs) that control immune cell development and immunity. The many experimental strategies designed to investigate developmental GRNs in sea urchin embryos (40) can be applied to investigations targeting immunity in the larva.

The larval immune response is mediated by a collection of phagocytic and granular immune cells [~100 total cells at 10 days post-fertilization (dpf) (17)]. These cells are initially specified in the early blastula-stage embryo from a ring of non-skeletal mesodermal (NSM) cells that differentiate into pigment cells, a heterogeneous suite of blastocoelar cells, and several other cell types including pharyngeal muscle and celomic pouches [Figures 1A,B; Ref. (25, 41)]. Presumptive pigment cells activate the transcription factor *glial cells missing* (*gcm*) and a battery of differentiation genes early in development that remain upregulated in the aboral NSM ring by late blastula (42). These cells migrate into the blastocoel relatively early in gastrulation and migrate to the aboral ectoderm and larval arms in an Ephrin/Eph receptor-mediated system (43). Differentiated granular pigment cells produce the antimicrobial naphthoquinone echinochrome A (44), which can react to form peroxide in the presence of high calcium concentrations (45). Pigment cells are motile and exhibit a surveillance-like migratory behavior even in immunoquiescent conditions. However, in response to immune challenge (e.g., disturbance of gut bacteria or intracelomic bacterial injection), a subset of pigment cells increase motility enter the blastocoel and interact with other immune cells at sites of wounding or infection. These cells are morphologically and transcriptionally similar to the red spherule cells, which mediate wound healing and immune response in adults (25, 39).

During mid-blastula stage, a set of oral NSM cells are marked by expression of *gata1/2/3* and *scl* (25), transcription factors that are homologs of important vertebrate hematopoietic mediators.

These cells undergo epithelial–mesenchymal transition later in gastrulation (about 10–15 h after the pigment cells) and enter the blastocoelar cavity where they differentiate into several cell types with immune activities. These include phagocytic cell types (a subset of *filopodial cells* and rarer, motile *ovoid cells* that appear upon acute immune challenge), highly motile *amoeboid cells* travel rapidly throughout the blastocoel, interacting with other immune cells and epithelia, and *globular cells*, a set of motile vesicular cell that are marked by expression of perforin/MPEG-like genes (17, 25). The phagocytic filopodial cells express the sea urchin-specific *Trf* genes in response to bacterial challenge, which parallels similar responses in adult phagocytic coelomocytes (34). Together, this assemblage of immune cell types dynamically interacts in the course of larval immune response.

Blastocoelar injection of labeled bacteria, fluorescent beads, or Zymosan (particles derived from yeast cell walls) into sea urchin larvae elicits immune cell migrations and phagocytosis (17, 23, 46, 47). In purple sea urchin larvae, the response varies according to the particle: *E. coli* K12 elicits a weak response whereas *Vibrio* species and Zymosan elicit much stronger responses (17). Injected *Vibrio diazotrophicus* cells agglutinate within minutes and are quickly engulfed by filopodial cells. Pigment cells and sometimes amoeboid and globular cells migrate and accumulate in regions of high bacterial concentration but are not phagocytic. Injection of Zymosan particles and *Vibrio* spp. cells sometimes elicit large, highly phagocytic cells (ovoid cells) that may derive from the syncytial filopodial cell network. The larval response to bacteria and other foreign particles involves layers of coordinated response among phagocytic and non-phagocytic immune cells and humoral factors.

THE PURPLE SEA URCHIN LARVA AS A MODEL FOR GUT-ASSOCIATED IMMUNE RESPONSE

Four to five days after fertilization, the mouth opens and larvae begin to feed on algae and other planktonic organisms. Before this, the gut lumen is exposed to microbes through the open blastopore. Following the onset of feeding, however, the gut maintains significant contact with the microbial world. Immune cell activity at the gut epithelium and the complexity of immune gene expression in the epithelial cells highlight the importance of the gut in larval immunity. When larvae are cultured in freshly collected sea water (allowing them to feed on complex, natural food sources), pigment cells are commonly observed near the gut epithelium (rather than the ectoderm), indicating that the baseline state in wild populations is more immune activated than in quiescent laboratory animals.

An acute infection is induced by exposing larvae to high concentrations of the marine bacterium *V. diazotrophicus* [Figures 1C,D; Ref. (17)]. Within 6 h, the gut epithelium thickens and a subset of pigment cells, mainly those in the ectoderm nearest the midgut, migrate between the ectoderm and gut, making repeated filopodial contact with the midgut and hindgut epithelium. Amoeboid cells also increase contact

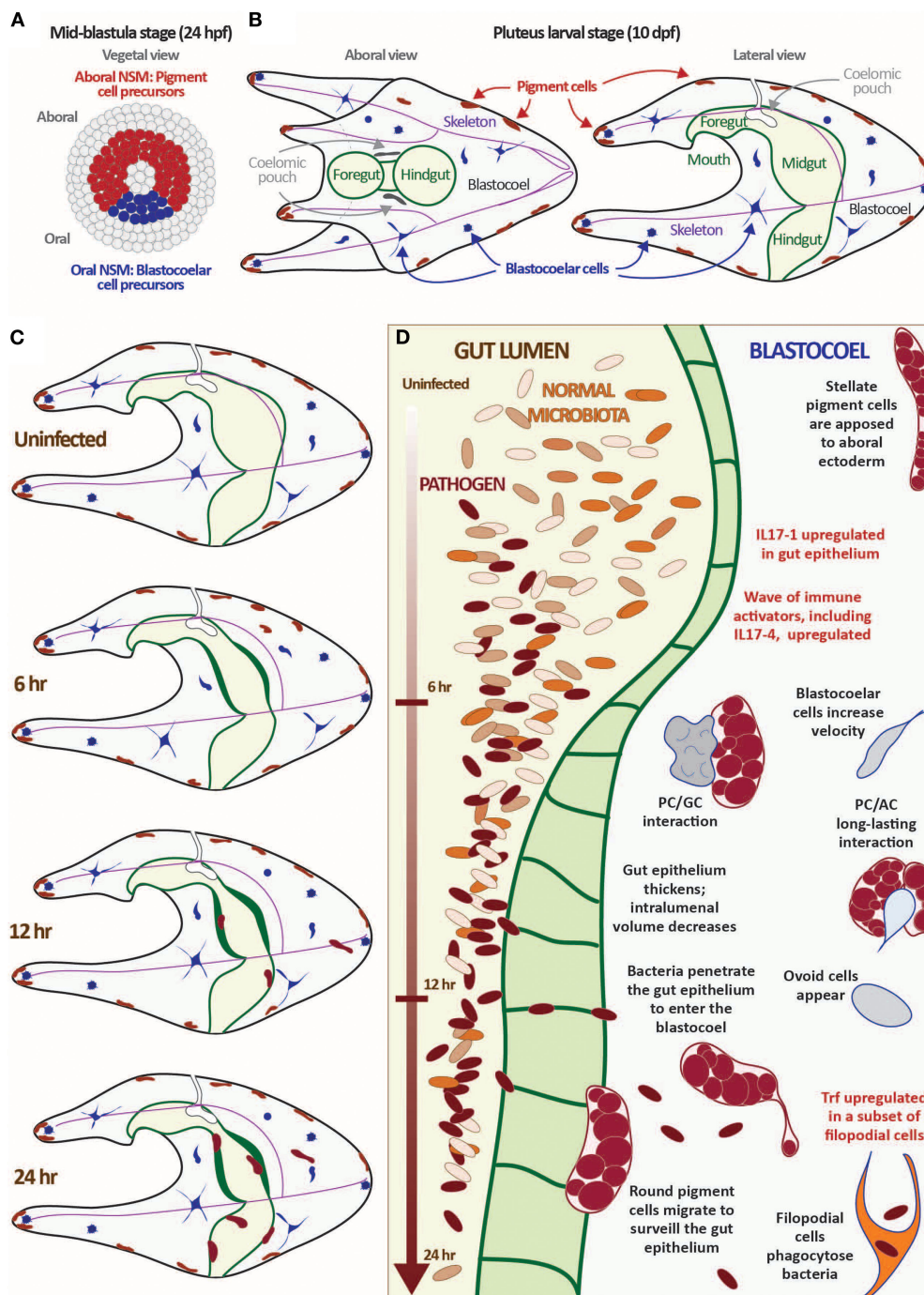


FIGURE 1 | Exposure to the marine bacterium *Vibrio diazotrophicus* induces an acute gut-associated inflammatory response in sea urchin larvae. (A,B) Sea urchin larvae exhibit a cellular immune response mediated by several mesodermally derived cell types. The mesenchyme blastula-stage embryo is shown from the vegetal view (A). In *Strongylocentrotus purpuratus*, embryos reach this stage about 24 hpf. The ring of non-skeletal mesoderm (NSM) cells is indicated by either red (aboral NSM) or blue (oral NSM). All other cell lineages are shown in gray. Aboral NSM cells differentiate into larval pigment cells; the oral NSM derivatives become the heterogeneous blastocoelar cells. Aboral and lateral views of the pluteus larvae are shown (B). Morphological features are indicated (pigment cells, blastocoelar cells, coelomic pouches, skeleton, and gut). The images shown in panels (A,B) are not to scale. (C,D) Larvae mount a cellular and transcriptional immune response to exposure to *V. diazotrophicus* in the sea water. In the first 24 h of exposure to *V. diazotrophicus*, the midgut epithelium thickens, reducing the volume of the gut lumen. Pigment cells change shape from a stellate to round morphology and migrate from the ectoderm to the gut. Cell motility increases, and complex cell:cell interactions occur. Bacteria begin to penetrate the gut epithelium and enter the epithelium, where they are phagocytosed by a subset of filopodial cells. One of the first transcriptional events is the acute upregulation of the *IL17-1* genes in the gut epithelium. This is followed by activation of a second wave of immune gene upregulation, including the *IL17-4* subtype. Immune effector genes, such as *Trf*, are activated in a subset of filopodial cells later in the response. Data are described in detail in Ref. (17, 29).

with the gut epithelium and make dynamic contacts with pigment cells that can last for hours. While it is unclear what is communicated during this process, it highlights the complex, cell-type interactions involved in immune response in this morphologically simple organism. After about 20 h, bacteria appear within the blastocoel of most larvae where they are quickly phagocytosed by *Trf*-expressing filopodial cells. This response requires live bacteria and is reversed by removing bacteria from the seawater. Because the response is relatively synchronous, tens of thousands of larvae can be analyzed in parallel to assess global transcription changes even of rare transcripts.

IL17 CYTOKINES MEDIATE THE LARVAL GUT-ASSOCIATED INFLAMMATORY RESPONSE

Within 2 h of exposure to *V. diazotrophicus*, changes in gene expression are evident in peripheral pigment cells near the ectoderm (17). However, at this point in infection, bacteria are restricted to the gut lumen and are not observed in the blastocoel until much later (12–24 h post-exposure). This suggests the possibility that gut epithelial cells communicate the perturbed state in the gut lumen to the wider organism. To identify an early immune signal, an RNA-Seq assay was used to quantify system-wide transcript levels in larvae over a time-course of exposure to *V. diazotrophicus*. From these data, a small family of genes orthologous to vertebrate IL17 cytokines emerged as the most highly upregulated transcripts across the entire genome (29). The mammalian IL17 signaling molecules [IL17A–F (48)] are expressed in Th17 cells, and other lymphocytes, myeloid cell types, and barrier tissues (49–51), including gut epithelial cells (52–54). IL17 expression in epithelia, particularly IL17C in the gut, maintains barrier integrity and regulates microbiota composition (52–56).

The *S. purpuratus* genome contains 30 genes predicted to encode functional IL17 factors (and five pseudogenes) (23, 29). Ten subtypes (IL17-1–10) are differentially expressed in the sea urchin immune response. These genes are transcriptionally inactive in immunoquiescent animals and are absent from non-challenged *S. purpuratus* transcriptome data. In the larval response to *V. diazotrophicus*, genes within two subtypes are rapidly upregulated. The *IL17-1* genes (11 nearly identical genes) are activated within 2 h of exposure and then rapidly attenuated by 8–12 h. The single *IL17-4* gene is activated with a moderate delay relative to *IL17-1* and coincides with the upregulation of a battery of other immune genes. Both IL17 subtypes are expressed exclusively in the mid- and hindgut epithelium (29). Although some cells express only one subtype at any one time as assessed by *in situ* hybridization, these IL17 subfamilies are often co-expressed (Figure 2). The successive expression of these IL17 subtypes in the gut epithelium suggests the possibility of a feedback mechanism to regulate the response.

Genes within a third subfamily, *IL17-9*, are upregulated in adult sea urchin coelomocytes. Transcript quantification in coelomocytes collected from adult sea urchins challenged with either live *V. diazotrophicus* or sham injection controls indicates

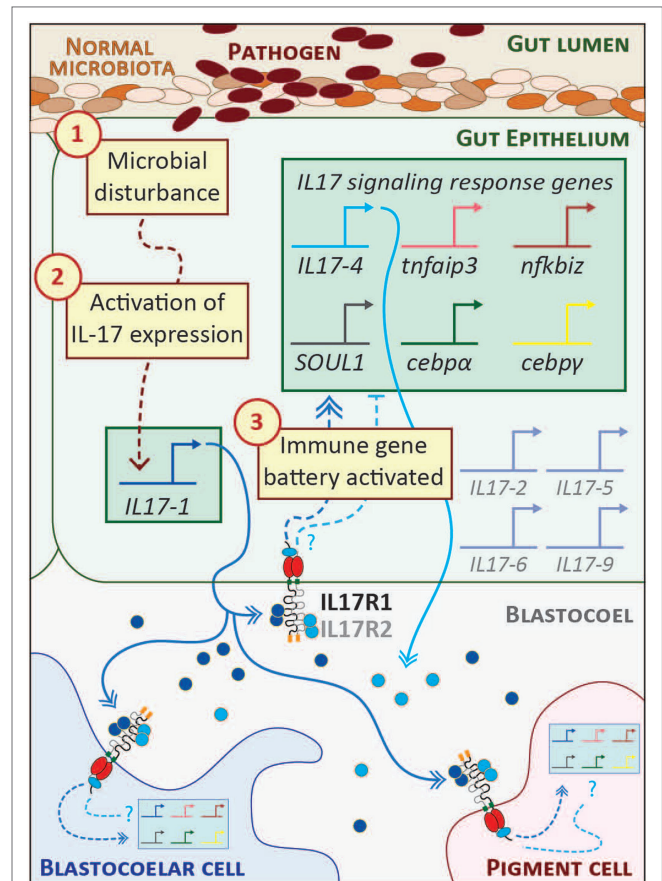


FIGURE 2 | Interleukin 17 (IL17) signaling mediates the larval immune response. A hypothetical scheme of the signaling molecules and transcriptional events that occur during the initial phase of the larval gut-associated immune response is shown. The community of normal microbiota is shown within the gut lumen in shades of brown. The introduction of pathogenic bacteria (indicated in dark red) to the gut is sensed by receptors the gut epithelial cells as a microbial disturbance [indicated by step (1)]. A signaling cascade is initiated that results in the transcriptional upregulation of the *IL17-1* genes [step (2)]. This is evident within 2 h of seawater exposure to *Vibrio diazotrophicus*. IL17-1 protein (dark blue) is secreted, where it can interact with widely expressed IL17 receptors and affect gene expression in cells distributed across the organism. IL17-R1 and -R2 are shown here as heterodimers, although they may also homodimerize. Upon activation, these receptors initiate intracellular signaling pathways that result in the upregulation of an IL17-dependent gene battery [step (3); shown in the green box]. These genes were identified using *in vivo* perturbation of IL17-R1 signaling (29). Notably, this includes the *IL17-4* gene, which is always activated subsequent to *IL17-1*. This linkage may point to regulatory feedback between the two subtypes and, given the rapid attenuation of *IL17-1* transcripts, the IL17-4 protein (light blue) may serve as an inhibitory mechanism. Given the broad expression patterns of the IL17 receptors, it is likely that immune cells (blastocoelar cells are shown in blue; pigment cells, pink) contain cell-type specific regulatory circuitry that controls immune gene expression in response to IL17 signaling. Spliced messages from the other IL17 subtypes (gray) can be recovered from larvae, although the levels are very low. These may be activated under different immune challenge conditions.

that challenged animals rapidly activated the *IL17-9* transcripts (peak expression 4–6 h post-infection). By contrast, the *IL17-9* genes were upregulated more slowly in sham-injected controls

(12–24 h post-infection), which is consistent with a more attenuated expression of the *Trf* genes.

Vertebrate IL17 receptors are characterized by an intracellular signaling domain known as a SEF/interleukin-1 receptor (SEFIR) domain (57). Five widely expressed IL17 receptors (IL17RA–E) (58) dimerize to mediate signaling in mammals (58–60). The *S. purpuratus* genome contains two genes that encode SEFIR domains; domain architecture and phylogenetic analysis indicate that both are IL17 receptors (IL17-R1 and IL17-R2) (23, 29). Consistent with observations in vertebrates, the sea urchin IL17 receptors are expressed at low levels; whole mount *in situ* hybridization suggests a broad expression pattern with some enrichment in the gut.

The functional consequences of IL17 signaling were investigated within the context of the larval inflammatory response using morpholino antisense oligonucleotides to perturb IL17-R1 signaling (29). These reagents were microinjected into fertilized eggs, which were grown to larval stage and exposed to *V. diazotrophicus*. Candidate genes for expression analysis were chosen based on their expression patterns (a sharp upregulation just following *IL17-1* activation) or known transcriptional links in other systems. Larvae subjected to IL17-R1 perturbation exhibit decreased expression of immune genes in response to immune challenge relative to controls. In the absence of IL17-R1 signaling, immune-challenged larvae expressed reduced levels of *IL17-4*, which may point to regulatory or feedback interactions between the two IL17 subtypes. In addition, reduced expression was also evident for two IL17 target genes in vertebrates: *tumor necrosis factor α induced protein 3* (*tnfaip3*; also known as A20), which encodes a ubiquitin-editing enzyme that inhibits NF- κ B activation (61), and *NF- κ B inhibitor ζ* (*nfkbi2*), an IL17 target gene in vertebrates that also regulates NF- κ B activity (62). Two IL17 associated transcription factors, *cebpa* and *cebpb*, are exhibit reduced activation in the larval inflammatory response in the presence of perturbed of IL17R signaling (29). Finally, IL17 signaling regulates the expression of a gene known as *soul1* (29). This transcript encodes a protein that contains a heme-binding SOUL domain (PF04832). The functions of these evolutionarily widespread domains are not well understood in

mammals (63). However, limiting iron availability is a known mechanism to suppress pathogen growth (64). The association between IL17 and SOUL1 may therefore represent an ancient regulatory connection yet to be identified in vertebrates. Together, these results indicate that highly regulated *IL17* expression in the sea urchin gut epithelium and signaling through IL17-R1 form a central axis of larval gut-associated immunity.

CONCLUSION AND PERSPECTIVES

The opening words in this review from Metchnikoff are now over 130 years old. Although Metchnikoff focused on cellular functions and we have long since moved to proteins and the genes encoding them, their relevance now holds renewed meaning. As we focus on genomes and the networks of regulatory interactions programmed therein, simple animal models offer novel strategies to investigate open problems in biology. These immune GRNs have evolved over hundreds of millions of years. Their highly complex and distributed nature requires that they be studied within intact organisms. The phylogenetic positions and experimental characteristics of well-chosen invertebrate models can be tailored to address specific questions. Here, we present the view that understanding these GRNs can shed light on how immune systems evolved on broad phylogenetic scales, a subject that remains poorly understood. The sea urchin larva is a morphologically simple model to experimentally characterize the system-wide GRNs that regulate immune cell development and immune response and is in an appropriate phylogenetic position to inform our understanding of vertebrate biology.

AUTHOR CONTRIBUTIONS

KB and JR wrote and edited the manuscript.

FUNDING

This work was funded by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2017-06247) to JR.

REFERENCES

1. Metchnikoff E. Untersuchungen über die Intracelluläre verdauung bei wirbellosen Tieren. *Arbeiten aus dem Zoologischen Institut der Universität Wien und der Zoologischen Station in Triest*. (1884) Vol 5 p. 141–168.
2. McFall-Ngai MJ, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* (2013) 110:3229–36. doi:10.1073/pnas.1218525110
3. Aderem A, Ulevitch RJ. Toll-like receptors in the induction of the innate immune response. *Nature* (2000) 406:782–7. doi:10.1016/j.imlet.2014.10.010
4. Lemaitre B, Nicolas E, Michaut L, Reichhart J, Hoffmann JA. The dorsoventral regulatory gene cassette *spätzle/toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* (1996) 86:973–83. doi:10.1016/S0092-8674(00)80172-5
5. Neely HR, Barreda DR, Flajnik MF. Evolution of myeloid cells. *Microbiol Spectr* (2016) 4. doi:10.1128/microbiolspec.MCHD-0007-2015
6. Gordon S. Phagocytosis: an immunobiologic process. *Immunity* (2016) 44:463–75. doi:10.1016/j.immuni.2016.02.026
7. Tauber A. Metchnikoff and the phagocytosis theory. *Nat Rev Mol Cell Biol* (2003) 4:897–901. doi:10.1038/nrm1244
8. Metchnikoff E. In: Starling FA, Starling Trans EHMD editors. *Lectures on the Comparative Pathology of Inflammation: Delivered at the Pasteur Institute in 1891*. Kegan Paul, Trench, Trubner & Co. Ltd (1893).
9. Ozinsky A, Underhill DM, Fontenot JD, Hajjar AM, Smith KD, Wilson CB, et al. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between toll-like receptors. *Proc Natl Acad Sci U S A* (2000) 97:13766–71. doi:10.1073/pnas.250476497
10. Hancock RE, Scott MG. The role of antimicrobial peptides in animal defenses. *Proc Natl Acad Sci U S A* (2000) 97:8856–61. doi:10.1073/pnas.97.16.8856
11. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Clark AG, Hultmark D. Dynamic evolution of the innate immune system in *Drosophila*. *Nat Genet* (2007) 39:1461–8. doi:10.1038/ng.2007.60
12. Dunn CW, Hejnos A, Matus DQ, Pang K, Browne WE, Smith SA, et al. Edgecombe GD, et al. Broad phylogenomic sampling improves

- resolution of the animal tree of life. *Nature* (2008) 452:745–9. doi:10.1038/nature06614
13. Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeck MA. Estimating metazoan divergence times with a molecular clock. *Proc Natl Acad Sci U S A* (2004) 101:6536–41. doi:10.1073/pnas.0401670101
 14. Raff RA. Origins of the other metazoan body plans: the evolution of larval forms. *Philos Trans R Soc Lond B Biol Sci* (2008) 363(1496):1473–9. doi:10.1098/rstb.2007.2237
 15. McClay DR. Evolutionary crossroads in developmental biology: sea urchins. *Development* (2011) 138:2639–48. doi:10.1242/dev.048967
 16. Burke RD. Structure of the digestive tract of the pluteus larva of *Dendraster excentricus* (Echinodermata: Echinoida). *Zoomorphology* (1981) 98:209–25. doi:10.1007/BF00312050
 17. 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* (2016) 94:861–74. doi:10.1038/icb.2016.51
 18. Smith MM, Cruz Smith L, Cameron RA, Urry LA. The larval stages of the sea urchin, *Strongylocentrotus purpuratus*. *J Morphol* (2008) 269:713–33. doi:10.1002/jmor.10618
 19. Materna SC. Using morpholinos to probe gene networks in sea urchin. *Methods in Molecular Biology*. Morpholino Oligomers Clifton, NJ. p. 87–104.
 20. Che-Yi Lin Y-HS. Genome editing in sea urchin embryos by using a CRISPR/Cas9 system. *Dev Biol* (2015) 409:1–9. doi:10.1017/CBO9781107415324.004
 21. Cameron RA, Samanta M, Yuan A, He D, Davidson EH. SpBase: the sea urchin genome database and web site. *Nucleic Acids Res* (2009) 37:D750–4. doi:10.1093/nar/gkn887
 22. Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, et al. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* (2006) 314:941–52. doi:10.1126/science.1133609
 23. 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
 24. Rast JP, Smith LC, Loza-Coll M, Hibino T, Litman GW. Genomic insights into the immune system of the sea urchin. *Science* (2006) 314:952–6. doi:10.1126/science.1134301
 25. Solek CM, Oliveri P, Loza-Coll M, Schrankel CS, Ho ECH, Wang G, et al. An ancient role for Gata-1/2/3 and Scl transcription factor homologs in the development of immunocytes. *Dev Biol* (2013) 382:280–92. doi:10.1016/j.ydbio.2013.06.019
 26. Schrankel CS, Solek CM, Buckley KM, Anderson MK, Rast JP. A conserved alternative form of the purple sea urchin HEB/E2-2/E2A transcription factor mediates a switch in E-protein regulatory state in differentiating immune cells. *Dev Biol* (2016) 416:149–61. doi:10.1016/j.ydbio.2016.05.034
 27. Rizzo F, Fernandez-Serra M, Squarzonni P, Archimandritis A, Arnone MI. Identification and developmental expression of the ets gene family in the sea urchin (*Strongylocentrotus purpuratus*). *Dev Biol* (2006) 300:35–48. doi:10.1016/j.ydbio.2006.08.012
 28. 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
 29. 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. doi:10.7554/eLife.23481
 30. Messier-Solek C, Buckley KM, Rast JP. Highly diversified innate receptor systems and new forms of animal immunity. *Semin Immunol* (2010) 22:39–47. doi:10.1016/j.smim.2009.11.007
 31. 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
 32. Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proc Natl Acad Sci U S A* (2000) 97:13156–61. doi:10.1073/pnas.230096397
 33. 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
 34. Smith LC, Lun CM. The *SpTransformer* gene family (formerly *Sp185/333*) in the purple sea urchin and the functional diversity of the anti-pathogen rSpTransformer-E1 protein. *Front Immunol* (2017) 8:725. doi:10.3389/fimmu.2017.00725
 35. Long KA, Nossa CW, Sewell MA, Putnam NH, Ryan JF. Low coverage sequencing of three echinoderm genomes: the brittle star *Ophionereis fasciata*, the sea star *Patiriella regularis*, and the sea cucumber *Australostichopus mollis*. *Gigascience* (2016) 5:20. doi:10.1186/s13742-016-0125-6
 36. Hall MR, Kocot KM, Baughman KW, Fernandez-Valverde SL, Gauthier MEA, Hatleberg WL, et al. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* (2017) 544:231–4. doi:10.1038/nature22033
 37. Sergiev PV, Artemov AA, Prokhortchouk EB, Dontsova OA, Berezkin GV. Genomes of *Strongylocentrotus franciscanus* and *Lytechinus variegatus*: are there any genomic explanations for the two order of magnitude difference in the lifespan of sea urchins? *Aging* (2016) 8:260–71. doi:10.18632/aging.100889
 38. Kober KM, Pogson GH. Genome-wide signals of positive selection in strongylocentrotid sea urchins. *BMC Genomics* (2017) 18:555. doi:10.1186/s12864-017-3944-7
 39. Smith LC, Ghosh J, Buckley KM, Clow LA, Dheilly NM, Haug T, et al. Echinoderm immunity. *Adv Exp Med Biol* (2010):260–301.
 40. Oliveri P, Tu Q, Davidson EH. Global regulatory logic for specification of an embryonic cell lineage. *Proc Natl Acad Sci U S A* (2008) 105:5955–62. doi:10.1073/pnas.0711220105
 41. Materna SC, Ransick A, Li E, Davidson EH. Diversification of oral and aboral mesodermal regulatory states in pregastrular sea urchin embryos. *Dev Biol* (2013) 375:92–104. doi:10.1016/j.ydbio.2012.11.033
 42. Ransick A, Davidson EH. Cis-regulatory logic driving glial cells missing: self-sustaining circuitry in later embryogenesis. *Dev Biol* (2012) 364:259–67. doi:10.1016/j.ydbio.2012.02.003
 43. Krupke OA, Burke RD. Eph-Ephrin signaling and focal adhesion kinase regulate actomyosin-dependent apical constriction of ciliary band cells. *Development* (2014) 141:1075–84. doi:10.1242/dev.100123
 44. Service M, Wardlaw AC. Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L.). *Comp Biochem Physiol B Comp Biochem* (1984) 79:161–5. doi:10.1016/0305-0491(84)90008-7
 45. Perry G, Epel D. Ca²⁺-stimulated production of H₂O₂ from naphthoquinone oxidation in *Arbacia* eggs. *Exp Cell Res* (1981) 134:65–72. doi:10.1016/0014-4827(81)90463-8
 46. Silva JR. The onset of phagocytosis and identity in the embryo of *Lytechinus variegatus*. *Dev Comp Immunol* (2000) 24:733–9. doi:10.1016/S0145-305X(00)00028-8
 47. Furukawa R, Takahashi Y, Nakajima Y, Dan-Sohkawa M, Kaneko H. Defense system by mesenchyme cells in bipinnaria larvae of the starfish, *Asterina pectinifera*. *Dev Comp Immunol* (2009) 33:205–15. doi:10.1016/j.dci.2008.08.011
 48. Pappu R, Ramirez-Carrozzi V, Ota N, Ouyang W, Hu Y. The IL-17 family cytokines in immunity and disease. *J Clin Immunol* (2010) 30:185–95. doi:10.1007/s10875-010-9369-6
 49. 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
 50. Marks BR, Craft J. Barrier immunity and IL-17. *Semin Immunol* (2009) 21:164–71. doi:10.1016/j.smim.2009.03.001
 51. Takatori H, Kanno Y, Watford WT, Tato CM, Weiss G, Ivanov II, et al. Lymphoid tissue inducer-like cells are an innate source of IL-17 and IL-22. *J Exp Med* (2009) 206:35–41. doi:10.1084/jem.20072713
 52. Song X, Zhu S, Shi P, Liu Y, Shi Y, Levin SD, et al. IL-17RE is the functional receptor for IL-17C and mediates mucosal immunity to infection with intestinal pathogens. *Nat Immunol* (2011) 12:1151–8. doi:10.1038/ni.2155
 53. Ramirez-Carrozzi V, Sambandam A, Luis E, Lin Z, Jeet S, Lesch J, et al. IL-17C regulates the innate immune function of epithelial cells in an autocrine manner. *Nat Immunol* (2011) 12:1159–66. doi:10.1038/ni.2156
 54. Reynolds JM, Lee Y-H, Shi Y, Wang X, Angkasekwinai P, Nallaparaju KC, et al. Interleukin-17B antagonizes interleukin-25-mediated mucosal inflammation. *Immunity* (2015) 42:692–703. doi:10.1016/j.immuni.2015.03.008

55. Reynolds JM, Martinez GJ, Nallaparaju KC, Chang SH, Wang Y-H, Dong C. Regulation of intestinal inflammation and barrier function by IL-17C. *J Immunol* (2012) 189:4226–30. doi:10.4049/jimmunol.1103014
56. Pappu R, Rutz S, Ouyang W. Regulation of epithelial immunity by IL-17 family cytokines. *Trends Immunol* (2012) 33:1–7. doi:10.1016/j.it.2012.02.008
57. Novatchkova M, Leibbrandt A, Werzowa J, Neubuser A, Eisenhaber F. The STIR-domain superfamily in signal transduction, development and immunity. *Trends Biochem Sci* (2003) 28:226–9. doi:10.1016/S0968-0004(03)00067-7
58. Gaffen SL. Structure and signalling in the IL-17 receptor family. *Nat Rev Immunol* (2009) 9:556–67. doi:10.1038/nri2586
59. Chang SH, Park H, Dong C. Act1 adaptor protein is an immediate and essential signaling component of interleukin-17 receptor. *J Biol Chem* (2006) 281:35603–7. doi:10.1074/jbc.C600256200
60. Wu L, Zepp JA, Qian W, Martin BN, Ouyang W, Yin W, et al. A novel IL-25 signaling pathway through STAT5. *J Immunol* (2015) 194:4528–34. doi:10.4049/jimmunol.1402760
61. Garg AV, Ahmed M, Vallejo AN, Ma A, Gaffen SL. The deubiquitinase A20 mediates feedback inhibition of interleukin-17 receptor signaling. *Sci Signal* (2013) 6:ra44. doi:10.1126/scisignal.2003699
62. Shen F, Ruddy MJ, Plamondon P, Gaffen SL. Cytokines link osteoblasts and inflammation: microarray analysis of interleukin-17- and TNF-induced genes in bone cells. *J Leukoc Biol* (2004) 77:388–99. doi:10.1189/jlb.0904490
63. Fortunato AE, Sordino P, Andreakis N. Evolution of the SOUL heme-binding protein superfamily across eukarya. *J Mol Evol* (2016) 82:279–90. doi:10.1007/s00239-016-9745-9
64. Sakamoto K, Kim Y-G, Hara H, Kamada N, Caballero-Flores G, Tolosano E, et al. IL-22 controls iron-dependent nutritional immunity against systemic bacterial infections. *Sci Immunol* (2017) 2:eaai8371. doi:10.1126/sciimmunol.aai8371

Conflict of Interest Statement: 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.

Copyright © 2017 Buckley and Rast. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Specific Pathogen Recognition by Multiple Innate Immune Sensors in an Invertebrate

Guillaume Tetreau^{*†}, Silvain Pinaud[‡], Anaïs Portet[‡], Richard Galinier, Benjamin Gourbal and David Duval^{*}

University of Perpignan, IHPE UMR 5244, CNRS, IFREMER, University of Montpellier, Perpignan, France

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Katherine Buckley,
George Washington University,
United States

Robert Braidwood Sim,
University of Leicester,
United Kingdom

*Correspondence:

Guillaume Tetreau,
guillaume.tetreau@gmail.com;
David Duval
david.duval@univ-perp.fr

[†]Present address:

Guillaume Tetreau,
University of Grenoble Alpes, CNRS,
CEA, IBS, Grenoble, France

[‡]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 07 July 2017

Accepted: 20 September 2017

Published: 05 October 2017

Citation:

Tetreau G, Pinaud S, Portet A,
Galinier R, Gourbal B and Duval D
(2017) Specific Pathogen Recognition
by Multiple Innate Immune Sensors in
an Invertebrate.
Front. Immunol. 8:1249.
doi: 10.3389/fimmu.2017.01249

Detection of pathogens by all living organisms is the primary step needed to implement a coherent and efficient immune response. This implies a mediation by different soluble and/or membrane-anchored proteins related to innate immune receptors called PRRs (pattern-recognition receptors) to trigger immune signaling pathways. In most invertebrates, their roles have been inferred by analogy to those already characterized in vertebrate homologs. Despite the induction of their gene expression upon challenge and the presence of structural domains associated with the detection of pathogen-associated molecular patterns in their sequence, their exact role in the induction of immune response and their binding capacity still remain to be demonstrated. To this purpose, we developed a fast interactome approach, usable on any host-pathogen couple, to identify soluble proteins capable of directly or indirectly detecting the presence of pathogens. To investigate the molecular basis of immune recognition specificity, different pathogens (Gram-positive bacterium, *Micrococcus luteus*; Gram-negative, *Escherichia coli*; yeast, *Saccharomyces cerevisiae*; and metazoan parasites, *Echinostoma caproni* or *Schistosoma mansoni*) were exposed to hemocyte-free hemolymph from the gastropod *Biomphalaria glabrata*. Twenty-three different proteins bound to pathogens were identified and grouped into three different categories based on their primary function. Each pathogen was recognized by a specific but overlapping set of circulating proteins in mollusk's hemolymph. While known PRRs such as C-type lectins were identified, other proteins not known to be primarily involved in pathogen recognition were found, including actin, tubulin, collagen, and hemoglobin. Confocal microscopy and specific fluorescent labeling revealed that extracellular actin present in snail hemolymph was able to bind to yeasts and induce their clotting, a preliminary step for their elimination by the snail immune system. Aerolysin-like proteins (named biomphalysins) were the only ones involved in the recognition of all the five pathogens tested, suggesting a sentinel role of these horizontally acquired toxins. These findings highlight the diversity and complexity of a highly specific innate immune sensing system. It paves the way for the use of such approach on a wide range of host-pathogen systems to provide new insights into the specificity and diversity of immune recognition by innate immune systems.

Keywords: invertebrate innate immunity, interactome, pathogen sensing, *Biomphalaria glabrata*, pattern-recognition receptor, proteomic profiling, immune specificity, hemocyte-free hemolymph

INTRODUCTION

The innate immune system allows the host to sense pathogens and mount an appropriate anti-pathogenic defense. Confronted with a large variety of pathogens, ranging from viruses to multicellular parasites, the animals' immune systems did not converge to a unique system with shared features but they emerged independently to provide an optimal protection of the host from infection (1). However, they all tend toward the genesis of a restricted repertoire of pathogen recognition molecules, named pattern-recognition receptors (PRRs), allowing to identify a determined diversity of pathogens (2). In vertebrates, pathogens recognition ability can be complemented by somatic recombination and hypermutation of a large repertoire of genes encoding immune receptors that lead to the production of soluble or membrane-bound antibodies (3, 4). Twelve years ago, Hargreaves and Medzhitov described the innate immune system in vertebrates as a complex of several recognition molecules capable of triggering one or more pathways to eliminate a given pathogen (1). Concepts highlighting the cooperation and complementation between the different recognition molecules leading to the activation of immune responses have since been supported by functional studies in vertebrates and in some model species (5, 6).

In invertebrates, and despite the lack of a vertebrate-like adaptive immunity, an increasing number of studies reported different repertoires of surprisingly highly diversified immune receptors within the innate immune system. This molecular diversity appears to be an essential basis for developing a fine and specific immune response against a large range of pathogens (7). The diversified arthropods' Down syndrome cell adhesion molecule (Dscam) generated by different splicing events, the somatic hypermutated snail fibrinogen-related proteins (FREPs), the C-type lectins, or the sea urchin 185/333 proteins whose diversity is generated by RNA editing and post-translational modifications are the most well-known diversified immune molecules (8–10). However, they are not the only critical factors involved in pathogen recognition since their knock-out by RNA interference did not result in a complete lack of protection (11, 12).

Many additional actors have been characterized with the increasing use of high-throughput sequencing. Their annotation as “immune-like receptors” was based on the induction of their gene expression following infectious challenges and/or on the presence in their gene sequence of homologous domains already characterized in known immune receptors. Indeed, most immunological processes in invertebrates are extrapolated based on protein sequence homology with other model species (13–15). Moreover, many transcriptomic experiments performed in invertebrates following challenges with different pathogens resulted in a list of differentially expressed immune genes, supposedly involved in pathogen recognition, for which the interaction with pathogens and the potential roles in immune recognition have never been validated (16–18). As a consequence, many molecular functions still remain to be clarified, particularly their real contribution in the effective host immune response and the nature of the pathogen and/or molecular target with whom they interact.

To solve these questions, we investigated the immune sensing ability for a wide range of pathogens, from bacteria to trematodes, by the schistosomiasis vector snail, *Biomphalaria glabrata*. The objective of this study was to identify which molecules from the snail host interacted with pathogen's surface determinants and their potential role in the specificity of the innate immune system. In this study, we report the repertoire of sensors from innate immunity constituted of previously characterized immune recognition factors (IRF) and of proteins involved in non-canonical immune pathways. These diverse and complementary molecules display a sentinel role by their constitutive expression in naïve animals. This circulating activity brings clues about the specificity and the mechanisms of pathogen detection in the host plasma. These results provide insights into the evolutionary selection of such factors and their role in specificity of invertebrate innate immunity that ultimately trigger an appropriate immune response, from inflammation to targeted clearance mechanisms.

MATERIALS AND METHODS

Snail Rearing

An albino strain of the freshwater snail *B. glabrata* originated from Recife, Brazil (BgBRE2) was used as the invertebrate host (19). The snail strain was maintained in rearing chambers at 26°C, 12/12 h light/dark period. The laboratory and experimenters possessed an official certificate from the French Ministry of National Education, Research, and Technology, CNRS and DRAAF Languedoc Roussillon for experiments on animals, animal housing, and animal breeding (# A66040; decree # 87–848, October 19, 1987; and authorization # 007083).

Hemolymph Extraction and Interaction with Pathogens

The interactome procedure used in this study consists in comparing the proteomic profile of the pathogen alone with the proteomic profile of the pathogen that was in contact with the cell-free hemolymph from the snail (**Figure 1**). This allows identifying the native proteins from the hemolymph that interact with outer proteins from the entire living pathogen. Hemolymph was collected from the head-foot region of twenty 9- to 10-mm snails (**Figure 1**, 1) as previously described (20). 5 and 2 mL of hemolymph from a pool of snails were used for each replicate for interactome with bacteria and yeast and with metazoan parasites, respectively. Hemolymph was centrifuged at $2,000 \times g$ for 10 min and the supernatant, constituting the cell-free hemolymph, was recovered for further interaction (**Figure 1**, 2). All plasma preparations were used immediately after their collection.

Integrity of the cells was verified by confocal microscopy to ensure that the procedure was not damaging the hemocytes, which could bias downstream analyses. Three conditions were tested: 1. freshly collected hemocytes were centrifuged at $2,000 \times g$ for 10 min and used as a control for intact cells; 2. hemocytes vortexed and centrifuged at $2,000 \times g$ for 10 min corresponded to the hemolymph preparation procedure of the interactome; 3. hemolymph sonicated (70% for 5 s) and then centrifuged ($2,000 \times g$ for 10 min) was the control of disrupted cells. Hemolymphatic cells

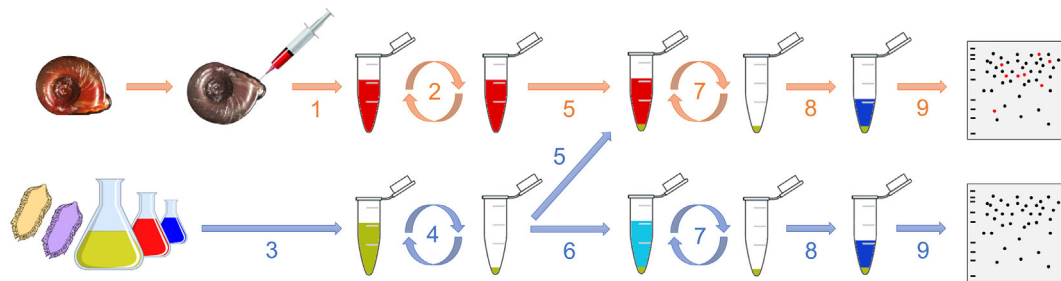


FIGURE 1 | Graphical representation of the interactome procedure. The hemolymph is collected (1) and then centrifuged (2). Meanwhile, the pathogen is also collected (3) and centrifuged (4). The cell-free hemolymph is put in contact with the pellet of pathogen (5; “pathogens + hemolymph”). A control is also performed consisting in adding a buffer that mimics the internal snail osmolarity to the pellet of pathogen (6; “pathogens only”). After 20 min, the suspension is centrifuged (7), the pellet is washed and proteins are extracted (8) for their proteomic profiling by 2D-PAGE (9).

were deposited on microscope slides to check their integrity and adhesion to surface. Cells were labeled with DAPI, which labels the DNA, and phalloidin, which labels the actin, by incubation for 20 and 2 min at 26°C in dark, respectively. Preparation was observed under a Zeiss LSM 700 microscope with two lasers at wavelengths of 405 and 488 nm for detection of DAPI and phalloidin labeling, respectively.

Five pathogens from three different kingdoms were used: the Gram-positive bacteria *Micrococcus luteus*, the Gram-negative bacteria *Escherichia coli*, the yeast *Saccharomyces cerevisiae*, and the two parasitic trematodes *Echinostoma caproni* and *Schistosoma mansoni*. *S. mansoni* and *E. caproni* have been maintained in the laboratory on *B. glabrata* BgBRE2 snails as previously described (12, 21).

The bacteria were plated and isolated on LB-agar Petri dishes. For each bacterium, one colony was introduced into a LB liquid medium and cultured overnight. Then, 150 µL of culture media, which contained approximately 35 million of bacteria, was sampled (Figure 1, 3) and centrifuged at $5,000 \times g$ for 10 min (Figure 1, 4). This quantity of bacterial cells was based on studies previously published (22, 23) and it was shown to be above the detection threshold of the 2D-SDS-PAGE approach by preliminary tests (data not shown), which ensured a proper analysis of the interactome profiles. The supernatant was discarded and the pellet was washed twice with 1 mL of Chernin's balanced salt solution (CBSS); NaCl, 48 mM; KCl, 2 mM; Na_2HPO_4 , 0.5 mM; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1.8 mM; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 3.6 mM; NaHCO_3 , 0.6 mM; pH 7.4. This buffer was chosen to mimic the internal snail osmolarity (24). The pellet was then resuspended in 1 mL of cell-free hemolymph and incubated on a rotating agitator for 20 min at 26°C (snail rearing chamber temperature) (Figure 1, 5). As a control, the bacterial pellet was incubated with 1 mL of filtered-CBSS in the same conditions (Figure 1, 6). After the incubation, the suspension was centrifuged at $5,000 \times g$ for 10 min and the pellet was washed twice with 1 mL of CBSS (Figure 1, 7). Three biological replicates of each condition (“pathogen alone” and “pathogen + hemolymph”) were performed.

The yeast culture was performed on a unique colony in Sabouraud liquid medium (dextrose, 20 g L⁻¹; pancreatic digest of casein, 5 g L⁻¹; peptic digest of animal tissue, 5 g L⁻¹, pH 5.6) at 26°C for 4 days. One hundred microliters of culture media, which

contained approximately 30 million yeast cells, was collected as described above for bacteria.

Schistosoma mansoni eggs were recovered as previously described (12), then exposed to water and light for 2 h to let miracidia hatch. *E. caproni* adults were recovered on the digestive tracts of mice, cultured *in vitro* in RPMI solution supplemented with penicillin and streptomycin (SP4458, Sigma) at 37°C for 2 days. Eggs were recovered, washed, and stored in water in the dark at 26°C with air injector. Twenty days later, eggs were put in fresh water and exposed to light for 2 h for miracidia hatching. One thousand five hundred miracidia from *S. mansoni* and *E. caproni* were individually counted by using a glass pipette and processed as described for bacteria until protein extraction.

Protein Extraction and 2D-SDS-PAGE Profiling

Proteins were extracted by resuspending the pellet of CBSS-washed pathogens in 70 µL of denaturing UTTC buffer (urea, 7 M; thiourea, 2 M; Tris, 30 mM; CHAPS, 4%; pH 8.5) (Figure 1, 8). After 2 h incubation at room temperature on a rocking agitator, the sample was centrifuged at $10,000 \times g$ for 5 min and the supernatant was transferred to a low protein binding tube for its analysis by 2D-electrophoresis (Figure 1, 9).

Then, 280 µL of rehydration buffer (urea, 7 M; thiourea, 2 M; CHAPS, 4%; DTT, 65 mM) containing 0.2% of Bio-Lyte 3/10 ampholyte (Bio-Rad) was added. The sample was then loaded on a tray channel for 5 h of passive rehydration followed by 14 h of active rehydration (50 V) of a 17 cm ReadyStrip IPG strip with a non-linear 3–10 pH gradient (Bio-Rad). Focusing was performed using the following program: 50 V for 1 h, 250 V for 1 h, 8,000 V for 1 h, and a final step at 8,000 V for a total of 90,000 V h with a slow ramping voltage (quadratically increasing voltage) at each step. Rehydration and focusing were both performed on a Protean IEF Cell system (Bio-Rad). Focused proteins were reduced by incubating the strip twice with equilibration buffer (Tris, 1.5 M; urea, 6 M; SDS, 2%; glycerol, 30%; bromophenol blue; pH 8.8) containing DTT (130 mM) at 55°C and they were alkylated by an incubation with equilibration buffer containing iodoacetamide (135 mM) on a rocking agitator (400 rpm) at room temperature protected from light.

Proteins were separated in function of their molecular weight on a 12%/0.32% acrylamide/piperazine diacrylamide gel run at 25 mA/gel for 30 min followed by 75 mA/gel for 8 h using a Protean II XL system (Bio-Rad). Protein standards were loaded with Whatman paper impregnated with 3 μ L of Unstained Precision Plus Protein Standards (Bio-Rad) on the left part of the gels. Gels were stained following a regular silver staining procedure: sensitizing using sodium acetate (68 g L⁻¹) and sodium thiosulfate (2 g L⁻¹), marking with 2.5 g L⁻¹ of silver nitrate, and then developing with sodium carbonate (25 g L⁻¹) in a 7.5% formaldehyde solution. Staining was stopped by replacing the developing solution by a solution of glycine (5 g L⁻¹) in 0.1% acetic acid. Gels were scanned using a ChemiDoc MP Imaging System (Bio-Rad) associated with Image Lab software version 4.0.1 (Bio-Rad). The qualitative comparative analysis of digitized proteome maps was conducted using the image analysis software PDQuest 7.4.0 (Bio-Rad). Only spots present in all the three replicates of “pathogens + hemolymph” samples and absent from all the profiles of pathogens alone were selected and picked in a mass spectrometry (MS)-compatible silver stained gel for further identification.

Spot Picking and Trypsin Digestion

Spots were excised from the gels using a Onetouch Plus Spot Picker Disposable (Harvard Apparatus), equipped with specific 1.5-mm methanol-washed tips. The gel plug containing the spot was disposed into a methanol-washed low protein binding tube and stored at -80°C until further processing. Gel plug was first destained by incubating it in 150 μ L of a solution of potassium ferricyanide (15 mM) and sodium thiosulfate (50 mM) at room temperature for 10 min on a rocking agitator (500 rpm). The destaining solution was discarded and this step was repeated once. Then, the plug was washed twice by adding 150 μ L of ammonium bicarbonate (25 mM) and it was incubated at room temperature for 30 min on a rocking agitator (500 rpm). Finally, 150 μ L of a solution of ammonium bicarbonate (12.5 mM) and acetonitrile (50%) was added to the spot. After incubation at room temperature for 10 min on a rocking agitator (500 rpm), the solution was discarded and the gel plug lyophilized for 30 min. The plug was rehydrated with 50 μ L of sequencing grade modified trypsin (Promega) and incubated on ice for 30 min. The excess of trypsin was discarded and 50 μ L of ammonium bicarbonate (25 mM) was added. Digestion was performed overnight at 30°C. The 50 μ L of solution were put in a new methanol-washed low-protein binding tube and the peptides were extracted from the plug by washing it three times with 100 μ L of a solution of formic acid (1%) and acetonitrile (50%) and by incubating 15 min at room temperature on a rocking agitator (500 rpm). The solution was collected at each washing step and mixed together in the same tube (final volume: 350 μ L). The solution was flash-frozen in liquid nitrogen, lyophilized for 3 h and stored at -80°C until further processing.

MS/MS Identification

Peptides were resuspended in 10 μ L of 3% (v/v) acetonitrile and 0.1% (v/v) formic acid, and then analyzed with a nano-LC1200 system coupled to a Q-TOF 6550 mass spectrometer equipped with a nanospray source and an HPLC-chip cube interface

(Agilent Technologies). A 34-min linear gradient (3–75% acetonitrile in 0.1% formic acid), at a flow rate of 350 nL min⁻¹, was used to separate peptides on a polaris-HR-Chip C18 column (150 mm long \times 75 μ m inner diameter). Full autoMS1 scans from 290 to 1700 *m/z* and autoMS2 from 59 to 1700 *m/z* were recorded. In every cycle, a maximum of five precursors sort by charge state (2+ preferred and single-charged ions excluded) were isolated and fragmented in the collision cell that was automatically adjusted depending on the *m/z*. Active exclusion of these precursors was enabled after 1 spectrum within 0.2 min, and the absolute threshold for precursor selection was set to 1,000 counts (relative threshold 0.001%). For protein identification, peak lists were extracted (merge MS_n scans with the same precursor at \pm 30 s retention time window and \pm 50 ppm mass tolerance) and compared with specific databases by using the PEAKS studio 7.5 proteomics workbench (Bioinformatics Solutions Inc., build 20150615). The searches were performed with the following specific parameters: enzyme specificity, trypsin; three missed cleavages permitted; fixed modification, carbamidomethylation (C); variable modifications, oxidation (M), pyro-glu from E and Q; monoisotopic; mass tolerance for precursor ions, 20 ppm; mass tolerance for fragment ions, 50 ppm; MS scan mode, quadrupole; and MS/MS scan mode, time of flight. For each interactome experiment, each spot identification was performed against the *B. glabrata* translated transcriptome (12, 25) and against the corresponding pathogen proteome. Only significant hits with a false discovery rate (FDR \leq 1) for peptide and protein cutoff (-logP \geq 20 and number of unique peptides \geq 2) were considered. For ensuring a proper identification of the proteins found by the interactome approach, a BLAST search against NCBI nr database was performed and the conserved domains of the sequence were retrieved using the NCBI CD-search available at <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> (26). For each protein, pI and molecular mass were also calculated with the ExPASy Compute pI/Mw tool (available at http://web.expasy.org/compute_pi) to compare with their location on the gel and provide an additional confirmation of their proper identification.

Validation of Actin As an Extracellular Immune Factor

Integrity of the cells was verified by confocal microscopy prior to actin localization in the plasma to ensure that the preparative procedure was not damaging the hemocytes, which could bias downstream analyses. The same three samples of hemolymph used for cell integrity (centrifuged hemolymph, vortexed and centrifuged hemolymph, and sonicated and centrifuged hemolymph) were used. 40 μ L of hemolymph from each sample were extracted in Laemmli buffer (Bio-Rad) containing β -mercaptoethanol and denaturated at 99°C for 5 min. Proteins were separated in a 12% acrylamide gel using the Mini-Protean Tetra Cell machinery (Bio-Rad) powered by PowerPac HC (Bio-Rad) at 110 V for 80 min. Proteins were then transferred onto a 0.2 μ m PVDF membrane using Trans-Blot Turbo Transfer Pack for 3 min at 25 V and 2.5 A (Bio-Rad). After saturation during 1 h at 37°C in TBSTM [1 \times TBS (500 mM Tris-HCl, 1.5 M NaCl, pH 7.5), 0.05% Tween20, 5% non-fat milk], the membrane was incubated for 90 min at

RT in TBSTM containing a mouse actin monoclonal antibody (mAbGEa, ThermoFisher) at a 1:1,000 dilution. The membrane was washed three times with TBST (TBSTM without milk), and further incubated for 70 min at RT with manufactured horseradish peroxidase-conjugated goat anti-mouse IgG antibody (Agrisera) at a 1:4,000 dilution. The membrane was washed three times with TBST. Actin presence was revealed by incubating the membrane in an enhanced chemiluminescent reagent (Super Signal West Pico Chemiluminescent Substrate, ThermoScientist) for 5 min at RT. The membrane was scanned using a ChemiDoc MP Imaging System (Bio-Rad) associated with Image Lab software version 4.0.1 (Bio-Rad).

Yeast Clotting by Incubation with Cell-Free Hemolymph

Yeast cells were cultured in Sabouraud medium as described above. They were washed twice with CBSS. Yeasts were then resuspended either in CBSS or in cell-free hemolymph for 20 min or 3 h. Preparations were deposited on microscope slides for plating and were then labeled with DAPI and phalloidin as described above. They were observed using a Zeiss LSM 700 microscope.

RESULTS AND DISCUSSION

An Original and Simple Method

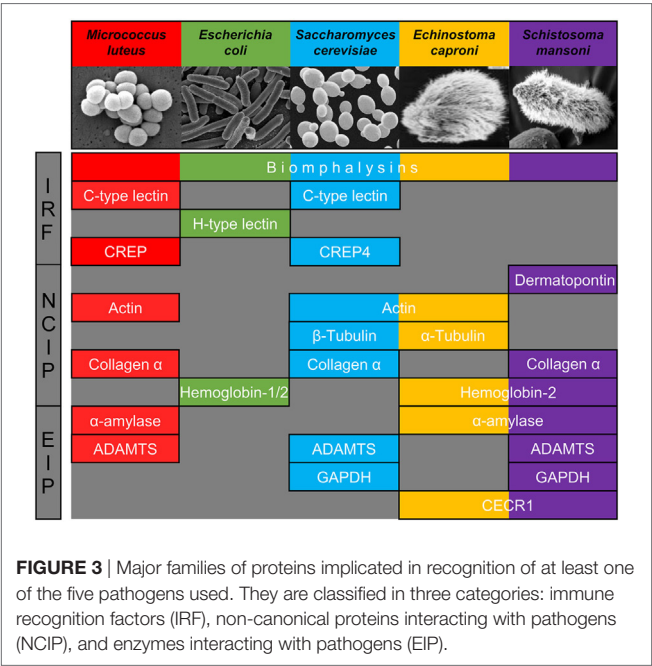
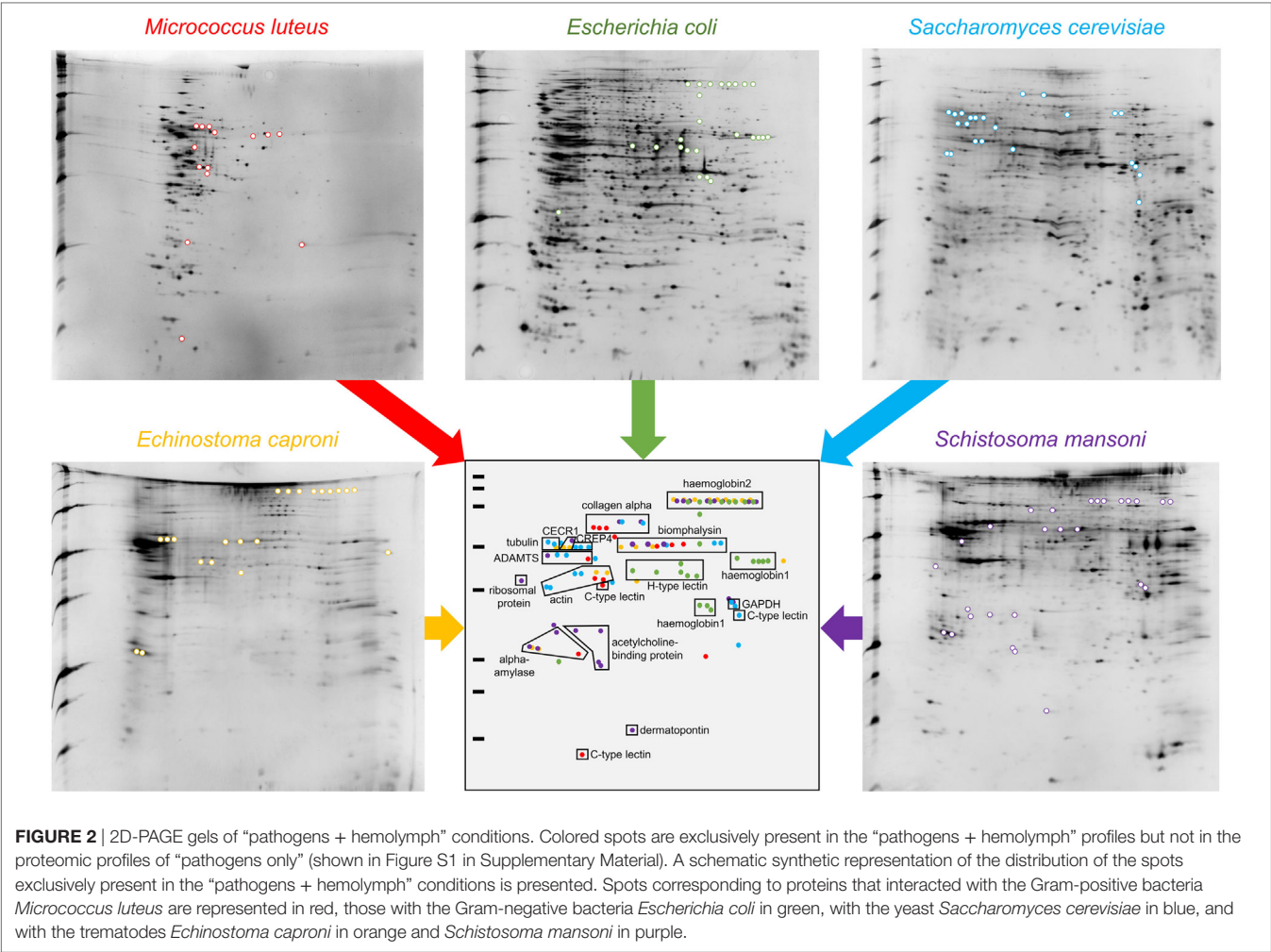
Generally, the identification of host molecules that can bind or recognize a set of pathogen determinants is performed by global pull-down assays. Such global interactome approach consists in the incubation of native or denatured protein extracts from both the host and the pathogen. The resulting interacting protein complexes are then separated through differential centrifugation steps, revealed by SDS-PAGE and identified by MS (27–29). Although powerful, this strategy suffers from several flaws, mainly associated with the extraction procedure itself which might (i) affect the nature of protein interactions by changing their conformation and (ii) promote forced interaction between proteins that would not encounter each other *in vivo*. Therefore, a part of the interactions observed can be essentially artificial and experimentally biased. To bypass these problems, we propose a new and simple interactome procedure in a cell-free hemolymph context that tends to mimic biological interactions between pathogens and soluble host proteins (Figure 1). Indeed, entire living pathogens were exposed to circulating humoral factors already present in cell-free hemolymph freshly extracted from naïve snails and they were incubated at 26°C, which corresponds to the environmental and internal temperature of this ectothermic organism. Therefore, only proteins present at the surface of the pathogen are recognized in a biologically realistic context. Moreover, the short time (20 min) chosen allows focusing exclusively on the very first step of innate immune response and avoiding the pathogen to respond to the attack from the immune factors, which could affect pathogens' proteomic profiles and bias the analysis. As a control, only spots that were present in the three “pathogen + hemolymph” replicates and absent in the three “pathogen only” replicates were considered for the analysis of each pathogen studied.

Each MS/MS profile was confronted to both the databases of the host and of the pathogen. This ensured that the approach reliably enabled the identification of host's interacting molecules while limiting the risk of false positives. No significant matches were observed against any of the pathogen databases, which confirms that all spots exclusively identified in the analysis of “pathogen + hemolymph” samples and not in the “pathogens only” gels were proteins from the snail's plasma. The benefit of this approach relies on its universality: it can be used with most host and parasite systems and gives rise to reliable qualitative differences within just few hours, which represents a great step forward for studies focusing on model and non-model systems.

Identification of a Large Variety of Interacting Proteins

This approach allowed the identification of a total of 109 spots exclusively identified in “pathogen + hemolymph” samples for the five pathogens tested (Figure 2; Figure S1 in Supplementary Material). These spots provided a significant match to 34 unique accession numbers, referring to 23 different proteins (Table S1 in Supplementary Material). Each pathogen was recognized by a specific, but overlapping, set of circulating proteins in mollusk's hemolymph (Figure 3). Specific recognition proteins to a given pathogen must be expected since each class of pathogen express at their surface specific and different structural motif also called pathogen-associated molecular patterns (PAMPs). The best known PAMPs are lipopolysaccharide from Gram-negative bacteria, lipoteichoic acid or peptidoglycan from Gram-positive bacteria, mannan-derived molecules or glycan from fungi, and fucosylated or glycoprotein receptors from *Schistosoma* sp. (30). Surprisingly, we identified numerous proteins not known to be involved in pathogen recognition and/or killing (extracellular matrix proteins, protease, and carbohydrase enzyme). Considering that some of these proteins are generally considered as intracellular molecules, a possible explanation for their presence could be that host's cells were damaged during the hemolymph collection (although non-invasive) and/or during the centrifugation step. A dual-staining with DAPI and phalloidin of hemocytes revealed no difference between fresh hemocytes and vortexed ones that were intact, as compared to sonicated hemocytes that were totally disrupted (Figure 4). This confirmed that the procedure of preparation of cell-free hemolymph did not damage the cells and that all interacting proteins from the snails were naturally present in the extracellular compartment of the hemolymph.

We, thus, propose to classify the snail interacting proteins identified into three different categories based on their nomenclature and known primary function: (i) molecules previously described as primary pathogen recognition molecules able to trigger an immunological response, with potential additional lytic activity [immune recognition factors (IRF)], (ii) proteins whose primary role is not pathogen sensing but are involved in other physiological functions [non-canonical proteins interacting with pathogens (NCIP)], and (iii) enzymes implicated in the metabolism of a wide range of molecules enzymes interacting with pathogens (EIP).



Pathogen Sensing by Soluble Immune Receptors and Atypical Toxins (IRF)

Among the IRE, two different families of proteins are identified: lectins and biomphalysin (Figure 2). Lectins represent a large family with a wide variety of evolutionarily conserved structures and some of them have been described as involved in immune recognition (7, 31). Among them, calcium-dependent (C-type) lectins were considered the most promising pattern-recognition proteins involved in the specific recognition of pathogens in the invertebrate immune system. This specificity is due to their high level of polymorphism and/or diversification to face up pathogens’ antigenic diversity (31). In addition to their role as soluble receptors, they can also limit the spreading of the pathogen in the host’s tissues and participate to its elimination (32, 33). Two different C-type lectins were interacting with the bacterium *M. luteus* and the yeast *S. cerevisiae* but not with the three other pathogens (Figure 2; Table S1 in Supplementary Material). Another C-type lectin-related protein (CREP4), recently characterized in *B. glabrata* from transcriptomic data (25), was apparently able to bind to *S. cerevisiae*. By contrast, the recognition of the bacterium *E. coli* involved a totally different category of lectin, the hyal-adherins (H-type), which are also

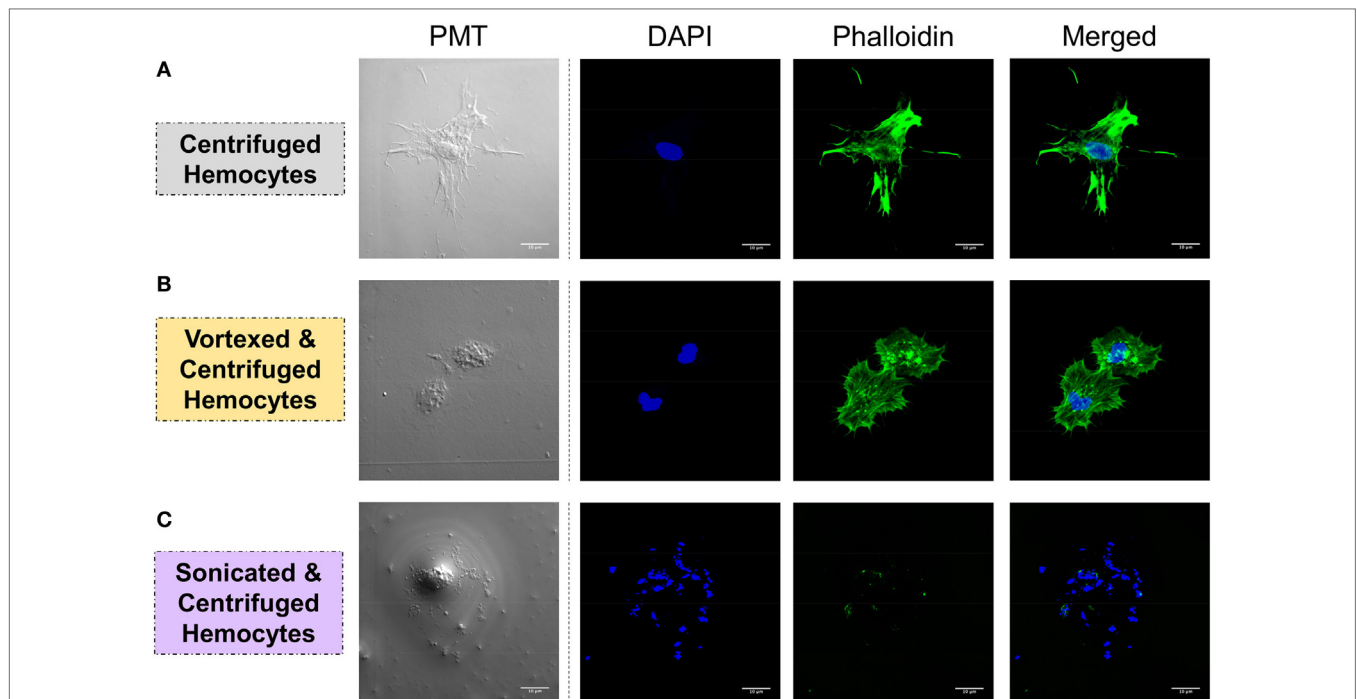


FIGURE 4 | Hemocyte integrity was tested by analyzing the spreading capacity and by observing the nuclear/cytoplasmic ratio. Hemolymph was carefully collected and was either **(A)** slowly centrifuged, **(B)** vortexed and centrifuged, or **(C)** sonicated and centrifuged. Hemocytes were stained with DAPI, which colors nucleicacids contained in the nucleus in blue, and with phalloidin, which colors f-actin in green. White bar = 10 μ m.

carbohydrate-binding proteins but data are missing concerning their role in pathogen recognition. Among the lectins, FREPs are proteins containing immunoglobulin-like domains whose role in the interaction between snails and metazoan parasites has been suggested (34, 35). Surprisingly, FREPs were not identified in the interaction with both metazoan parasites in our study while they were evidenced in previous transcriptomic and proteomic studies (27, 29). Such discrepancy with previous results likely comes from the different developmental stage of the parasites used in the different studies, i.e., miracidia herein and sporocysts in other studies. Several proteomic and glycomic studies showed that the glycan elements harbored by *Schistosoma*, to which FREPs bind, differ from one developmental stage to another (36, 37). This would suggest a subtle ability for the snail immune machinery to distinguish various intramolluscan developmental stages of the parasite (miracidium to primary and secondary sporocysts or even cercariae) and FREPs might not be involved in the recognition of all stages. Moreover, FREPs were previously identified by interactome experiments after 2.5 h of contact between protein extracts from sporocyst and snail cell-free hemolymph (27) while our procedure includes a 20-min contact of outer pathogen membrane proteins with circulating snail hemolymph proteins. Of note, it has been observed that some FREPs can form multimers and that they can interact with other proteins such as thioester-containing proteins (TEPs), which could both modulate their recognition ability (27, 34, 38). It is, therefore, possible that these processes are mandatory for the recognition by FREPs of the pathogens used in this study. A longer exposure time between pathogens with proper membrane-bound glycan

antigens and the cell-free hemolymph would then be required for the complexes to form and for their detection by our interactome approach.

The second class of IRF identified is the biomphalysin toxin, which is an aerolysin-like protein that has been acquired by a putative horizontal gene transfer from a bacterium (39) (Figure 3). This protein is constituted of two domains: one large domain that shares structure similarities with β -pore-forming toxins whose role is to perforate cell membranes by forming transmembrane pores and a small domain potentially involved in pathogens' carbohydrate motifs recognition (39). Biomphalysin is a dual protein: it has recently been shown to directly bind to *S. mansoni* sporocysts and to have a lytic activity enhanced by snail plasmatic factors (39). Herein, we demonstrate for the first time that this anti-schistosome toxin is also able to interact with other pathogens and suggest a role in bacterial clearance. One (*E. coli*) and three (*M. luteus*, *S. cerevisiae*, *E. caproni*, and *S. mansoni*) spots were identified as biomphalysins in 2D gels (Figure 2; Table S1 in Supplementary Material). Even if they were all of the same size (65–70 kDa), the expected size of biomphalysin (39), they exhibited a large range of isoelectric points, from slightly acid/neutral for *E. caproni* and *S. mansoni* to basic for *E. coli* and *S. cerevisiae* (Figure 2). Altogether, this suggests that different protein isoforms of biomphalysins must be involved in the recognition/clearance of the same pathogen but also of different pathogens. Interestingly, different biomphalysin genes were predicted in the recently sequenced genome of *B. glabrata* (BioProject: PRJNA290623 on NCBI database) (40), which suggests that they might be different genes rather than different isoforms (39). This

biomphalysin family could be a major player of the specificity of the *Biomphalaria* innate immune response together with lectins.

Biomphalysins were the only proteins that interacted with all pathogens. There is a growing number of evidence that aerolysin-like proteins have been horizontally transferred within many different invertebrate phyla acquiring in the same time potentially new and varied functions but details of their involvement in the invertebrate immunity remain largely unknown (41). The interactome approach developed herein suggests that biomphalysins might be a key component of the pathogen sensing system, and potentially of its specificity. Indeed, heterogeneous assembly from these different monomeric isoforms to the heptameric biomphalysin pore complex may generate a high degree of pathogen-binding specificity. In *Anopheles gambiae*, two C-type lectins, CTL4 and CTLMA2, form a disulfide-linked heterodimer to specifically kill *E. coli* (42). The ability to form heterodimers could greatly expand the repertoire of recognition molecules (43, 44). Further experiments are now required to understand how biomphalysin gene expression is regulated in response to exposure with different pathogens and how the different proteins are recruited to respond to a specific pathogen encounter.

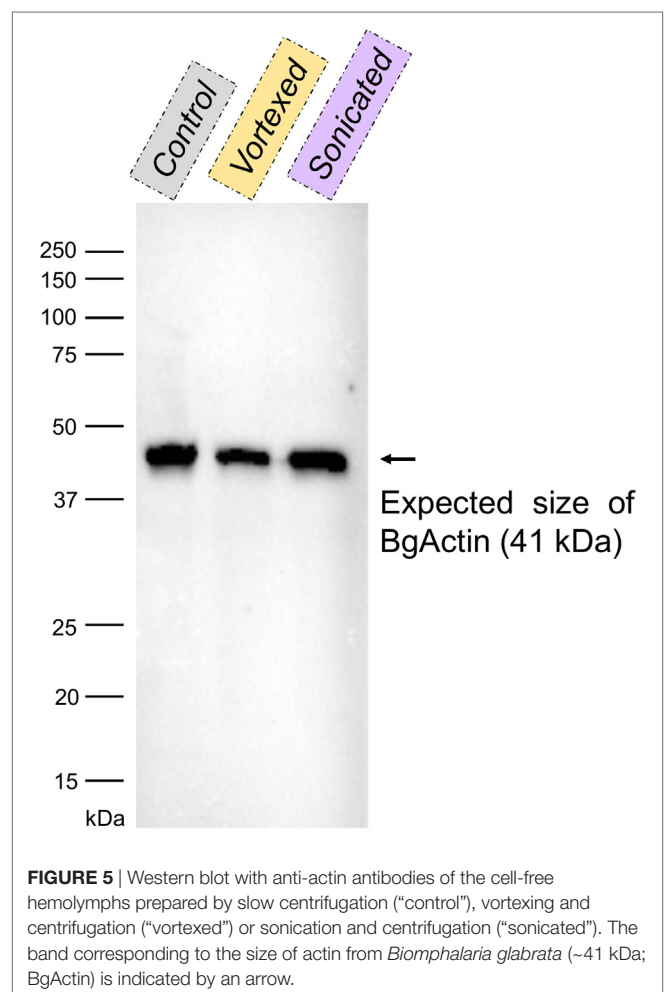
Pathogen Sensing by Major Extracellular Matrix Components (NCIP)

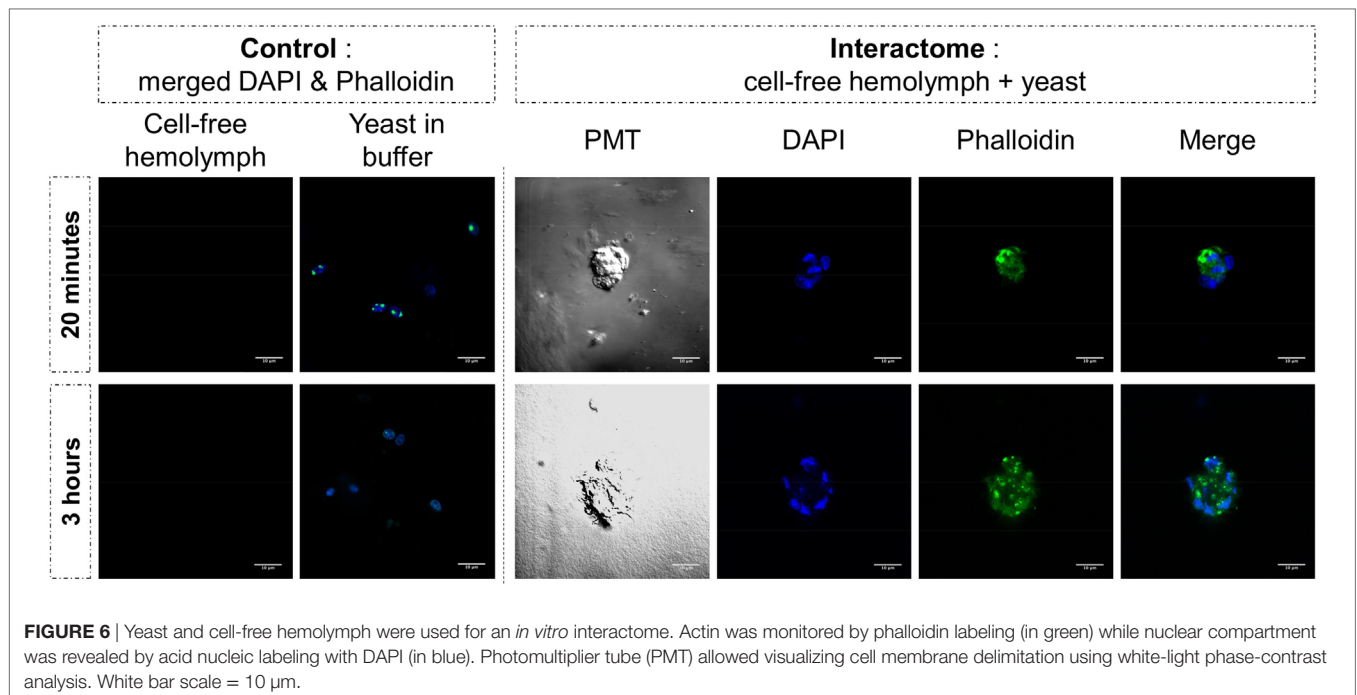
The category of NCIP includes proteins whose primary function is not immunity, such as cell-matrix junction proteins (dermatopontin, collagen) and cytoskeleton extracellular matrix proteins (actin, tubulin). Concerning the dermatopontin, its gene expression can be increased after immune challenge with *E. caproni* (21, 45) and *S. mansoni* (45) but not with *E. coli*, *B. cereus*, and *S. cerevisiae* (46). While its role was unknown at this time, our results suggest that it might be involved in a hemolymph coagulation-like system to prevent parasite establishment through the tissue of the host (Figure 3).

The same type of molecular process is expected for other extracellular proteins such as actin. Western blot analyses of cell-free hemolymph using anti-actin antibodies revealed its presence in the extracellular compartment of the hemolymph (Figure 5). Considering that the procedure of hemolymph collection and preparation did not damage the cells (Figure 4), this actin must be considered as a real extracellular actin (ECA) present in snail hemolymph. Interestingly, the amount of ECA present in hemolymph was similar between the three conditions tested in western blot, which suggests that ECA is an important component of hemolymph released by a process still unknown in mollusk. In insects, some isoforms are secreted from cells through an exosome-independent pathway (47) while monocyte cells can release some extracellular vesicles (ectosome) containing β -actin and actinin in vertebrates (48). Observation of yeasts by confocal microscopy shows that in CBSS buffer, some actin is located inside the yeast, revealed as small precisely localized green dots (Figure 6). In the presence of cell-free hemolymph, these intra-yeast dots of actin are still visible but there is a large amount of ECA surrounding the yeast cells, which appears as early as 20 min and seems even more intense after 3 h of incubation (Figure 6). Considering that yeasts were still intact after 20 min of contact with cell-free hemolymph, this actin surrounding the yeasts is likely the ECA from snail that is able to bind and participate

to yeast clotting prior to its elimination. The triggering of the destruction of yeast cells by these immune complexes is indicated by their nuclear destructure visible at 3 h (Figure 6). This finding is consistent with recent studies that demonstrated an active role of actin in extracellular trap for pathogens clotting, facilitating their elimination by phagocytosis in the mosquito *A. gambiae* for example (47). Until now, these soluble molecules were considered as damage-associated molecular patterns (DAMPs) potentially involved in the “danger theory” where self-constituents could trigger an immune response (49). Based on our results and particularly on the short time of our interaction that prevents the pathogen from circumventing host immune factors, these molecules must be considered as soluble immune sensing factors rather than just DAMPs.

The case of hemoglobin is particularly interesting. Two different classes of hemoglobin were identified against *E. coli* (hemoglobin-1 and -2) while only hemoglobin-2 was interacting with *E. caproni* and *S. mansoni* (Figure 2). Many different isoforms were identified (same size, different isoelectric points) but they were at a much lower size (55–60 and 100–120 kDa for hemoglobin-1 and -2, respectively) than the predicted full-size hemoglobin protein predicted from *B. glabrata* genome (514 and 582 kDa, respectively) (Figure 2). Such peptides with enhanced or alternative





functionality that can be liberated from larger proteins are named cryptides. Those derived from hemoglobin have already been associated with immune modulation, hematopoiesis, signal transduction, and microbicidal activities in metazoans (50). Although identified as differentially expressed upon *S. mansoni* exposure in *B. glabrata* (45), these highly abundant proteins were excluded from previous interactome approaches by ultracentrifugation of plasma as they were thought to interfere with pathogen recognition and not be directly implicated in it (27). Also, the role of this major protein in hemolymph has been largely neglected as its function was expected to be mostly pleiotropic. Hemoglobin and/or hemoglobin cryptides could directly interfere with the pathogen and limit its growth, as it has been shown for the “classical swine fever virus” (51), and/or they could reinforce the interaction between pathogen and extracellular matrix proteins, as it has been shown between human fibronectin and the pathogenic yeast *Candida albicans* (52). The binding of hemoglobin to the major virulence factor of *Salmonella typhi* has also been shown to promote the production of proinflammatory cytokines from monocytes (53).

Host Plasmatic Enzymes Involved in Pathogen Surface Binding (EIP)

Many different EIPs were identified in this interactome approach (Figure 3). α -amylases have already been identified after co-immunoprecipitation of *B. glabrata* plasmatic proteins with *S. mansoni* protein extracts but they were considered as mucus contamination at this time (27). Present data challenge this contamination hypothesis since α -amylase was only detected after interaction of hemolymph with *M. luteus*. α -amylases would, thus, be critical for the host’s specific response to certain pathogens. For the other EIPs, reports on the involvement of

ADAMTS, GAPDH, and CECR1 in invertebrate immunity are scarce. However, GAPDH has been demonstrated to modulate immune responses against bacteria in plants (54) and metalloproteases have been characterized as key actors of many diverse immune and inflammatory processes in vertebrates (55). Results obtained in this study demonstrate that their binding to the pathogen surface can no longer be considered as artifactual. Further experiments are now required to understand if EIPs can bind directly to surface pathogens’ factors or if their involvement is related to their enzyme activities to mediate the maturation of immune complexes after association with other IRFs and/or NCIPs.

Experimental Support to Theoretical Concepts Opens New Perspectives for Studying Pathogen Sensing by Invertebrates

Although extensively investigated and well documented in vertebrates, the factors involved in invertebrate immune recognition rather constitute a black-box in which many different proteins with a wide range of functions, often referred to as PRRs, can be found (56, 57). Some responses have arisen from model species essentially from insects such as *Drosophila* for which the Gram-negative bacteria-specific Imd pathway and the fungi and Gram-positive bacteria-specific Toll pathway have been first identified (8). However, data remain scarce in non-model species mostly due to the absence of reliable knock-out technology, which may fail in demonstrating the full richness and the role of invertebrate pathogen recognition molecules (7, 58).

In this study, we developed a simple interactome approach to identify soluble plasmatic molecules that bind directly or

indirectly to pathogen surfaces and to gain access rapidly to the biological functions of the candidate proteins. Here, we focused on the sentinel role of molecules that interacted with pathogens since they were constitutively present in hemolymph of uninfected (naïve) snails. Indeed, most of the studies are based on the differential analysis (i.e., uninfected vs infected, or infected by different pathogens) of the host immune response (efficient or not) leading to a list of genes whose immunological function is rarely demonstrated. Moreover, if functional invalidation (gene knock-out, siRNA-mediated gene silencing, and mutants) already demonstrated the requirement of such molecules during the immune response, the first step of pathogens binding is still rarely studied (11, 12, 33).

Each pathogen was recognized by a specific, although partially overlapping set of interacting proteins from the mollusk (Figure 3; Table S1 in Supplementary Material). Most of pathogens' perception involved at least three different families of proteins from two of the three protein categories described (Figure 3). Such contrasting sets of binding proteins, in terms of diversity and quantity, suggest that specificity of immune detection quickly occurs at a fine scale. The recognition of the same pathogen by several different sensors with a high degree of specificity suggests that these molecules are part of different host defense pathways that can interact with each other (1). Such interactions can take three different forms: by cooperation, leading to the more efficient engagement of the same effector mechanism, by complementation, allowing to trigger different complementing effector mechanisms or by compensation, where one pathway compensates the deficiency of another one (59). The real involvement of these proteins in pathogen recognition, as expected in parasite antigen/host receptor interaction, is still not demonstrated and will require specific investigation of downstream process for each candidate identified. Thus, these pathways might contribute to assess the danger for which they have been exposed and leading *in fine* to discriminate symbiotic organisms from pathogens (60). Simultaneous activation of distinct recognition pathways would enable a concerted and appropriate response to tolerate or eliminate such or such intruder. Another aspect of the molecular interaction not yet described and evaluated in invertebrates is the temporal dynamic of pathogen perception by soluble immune factors. Can this recognition be immediate and frozen once and for all, or require gradual biochemical and structural maturation to recruit other more specialized immune factors? The dynamic of structuration of soluble immune complexes by analyzing interactome at different times must be explored to answer this question. In this study, we show that different biochemical interactions between the external surface of pathogens and host molecules occur within just 20 min of interaction. This supports the idea of a first wave of pathogen detection that we called "sensing," a prerequisite for the subsequent activation of immune system. This sensing step appears additive but also epistatic by the number of various biological functions involved and suggests a cooperative crosstalk for a specific immune response (1). The relative function of the IRE, NCIP, and EIP, whether they are implicated in pathogen recognition, immune complex maturation, and/or triggering of immune response, will require further investigation. The method developed herein allowed reaching the early step of pathogen sensing, validating the binding ability of

several IRE, and opening opportunity in model systems to deeper study their activity in the immune response pathways.

In summary, the present data constituted an important step toward a better understanding of the pathogen sensing and immune specificity in invertebrates. It clearly demonstrates that innate immune response in invertebrate is not supported by a unique class of immune factors but rather by a panel of molecules involved in diverse biological functions and able to bind specifically to a range of distinct pathogens. Notably, it involves some dual immune proteins able to play a role in both pathogen binding and clearance. This work does not intend to provide an extensive description of all sensing molecules but it definitely opens the way to a better integrative biological overview of molecules necessary to initiate an orchestrated immune response against pathogens in both model and non-model organisms.

AUTHOR CONTRIBUTIONS

GT, SP, and DD designed the research; AP, RG, and BG substantially participated in conception and improvement of research; SP, AP, and DD performed interaction experiments; GT performed the 2D-SDS-PAGE experiments and qualitative analysis; SP and AP performed the Western blots, fluorescent labeling, and microscope observations; all authors contributed to the analysis and interpretation of the results; GT, SP, and DD led the manuscript writing; all authors participated to manuscript writing, editing, and critical reviewing; and they all approved the final draft.

ACKNOWLEDGMENTS

The authors want to thank Philippe Chan, Marie-Laure Walet-Balieu and David Vaudry from PISSARO Proteomic Platform for 2D spots protein identification and Nathalie Arancibia, Cécile Saint-Beat and Anne Rognon for the animal breeding facilities. The authors would like to thank the members of EcoEvI's groups of IHPE laboratory for helpful discussions.

FUNDING

This work was funded by ANR JCJC INVIMORY (number ANR-13-JSV7-0009) from the French National Research Agency (ANR).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://www.frontiersin.org/article/10.3389/fimmu.2017.01249/full#supplementary-material>.

FIGURE S1 | 2D-PAGE gels of "pathogens + hemolymph" and "pathogens only" for each of the five pathogens used. Arrows are indicating spots exclusively present in the "pathogens + hemolymph" profiles but not in the proteomic profiles of "pathogens only," which represents proteins from *Biomphalaria glabrata* hemolymph that participated in the recognition of pathogen's proteins.

TABLE S1 | Protein identification of the 109 spots revealed only in "pathogens + hemolymph" gels as compared to "pathogens only" gels. For each spot, the $-10\log P$ values of proteins and peptides are indicated, together with the top BLAST hit in NCBI nr database, the conserved domains of the sequence retrieved (performed with NCBI CD-search available at <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) and the pI and molecular mass (calculated with the ExPASy "Compute pI/Mw tool" available at http://web.expasy.org/compute_pi).

REFERENCES

- Hargreaves DC, Medzhitov R. Innate sensors of microbial infection. *J Clin Immunol* (2005) 25:503–10. doi:10.1007/s10875-005-8065-4
- Janeway CA Jr, Medzhitov R. Innate immune recognition. *Annu Rev Immunol* (2002) 20:197–216. doi:10.1146/annurev.immunol.20.083001.084359
- Matthews AJ, Zheng S, Dimenna LJ, Chaudhuri J. Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Adv Immunol* (2014) 122:1–57. doi:10.1016/B978-0-12-800267-4.00001-8
- Methot SP, Di Noia JM. Molecular mechanisms of somatic hypermutation and class switch recombination. *Adv Immunol* (2017) 133:37–87. doi:10.1016/bs.ai.2016.11.002
- Delaney JR, Stöven S, Uvell H, Anderson KV, Engström Y, Mlodzik M. Cooperative control of *Drosophila* immune responses by the JNK and NF- κ B signaling pathways. *EMBO J* (2006) 25:3068–77. doi:10.1038/sj.emboj.7601182
- Slack E, Hapfelmeier S, Stecher B, Velykoredko Y, Stoel M, Lawson MA, et al. Innate and adaptive immunity cooperate flexibly to maintain host-microbiota mutualism. *Science* (2009) 325:617–20. doi:10.1126/science.1172747
- Du Pasquier L. Germline and somatic diversification of immune recognition elements in Metazoa. *Immunol Lett* (2006) 104:2–17. doi:10.1016/j.imlet.2005.11.022
- Lemaitre B, Hoffmann J. The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* (2007) 25:697–743. doi:10.1146/annurev.immunol.25.022106.141615
- 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
- Armitage SA, Peuss R, Kurtz J. Dscam and pancrustacean immune memory – a review of the evidence. *Dev Comp Immunol* (2015) 48:315–23. doi:10.1016/j.dci.2014.03.004
- Hanington PC, Forsys MA, Dragoo JW, Zhang SM, Adema CM, Loker ES. Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. *Proc Natl Acad Sci U S A* (2010) 107:21087–92. doi:10.1073/pnas.1011242107
- Pinaud S, Portela J, Duval D, Nowacki FC, Olive MA, Allienne JF, et al. A shift from cellular to humoral responses contributes to innate immune memory in the vector snail *Biomphalaria glabrata*. *PLoS Pathog* (2016) 12:e1005361. doi:10.1371/journal.ppat.1005361
- Cannarsa E, Zampicini G, Friard O, Santovito A, Cervella P. *De novo* transcriptome sequencing of a non-model polychaete species. *Mar Genomics* (2016) 29:31–4. doi:10.1016/j.margen.2016.06.001
- Clark KE, Greenwood SJ. Next-generation sequencing and the crustacean immune system: the need for alternatives in immune gene annotation. *Integr Comp Biol* (2016) 56:1113–30. doi:10.1093/icb/icw023
- Schultz JH, Adema CM. Comparative immunogenomics of molluscs. *Dev Comp Immunol* (2017) 75:3–15. doi:10.1016/j.dci.2017.03.013
- Coustau C, Gourbal B, Duval D, Yoshino TP, Adema CM, Mitta G. Advances in gastropod immunity from the study of the interaction between the snail *Biomphalaria glabrata* and its parasites: a review of research progress over the last decade. *Fish Shellfish Immunol* (2015) 46:5–16. doi:10.1016/j.fsi.2015.01.036
- Hussain A, Rizwan-Ul-Haq M, Al-Ayed H, Aljabr AM. Susceptibility and immune defence mechanisms of *Rhynchophorus ferrugineus* (Olivier) (Coleoptera: Curculionidae) against entomopathogenic fungal infections. *Int J Mol Sci* (2016) 17(9):1518. doi:10.3390/ijms17091518
- Zhu J, Fu Q, Ao Q, Tan Y, Jiang H, et al. Transcriptomic profiling analysis of tilapia (*Oreochromis niloticus*) following *Streptococcus agalactiae* challenge. *Fish Shellfish Immunol* (2017) 62:202–12. doi:10.1016/j.fsi.2017.01.023
- Theron A, Rognon A, Gourbal B, Mitta G. Multi-parasite host susceptibility and multi-host parasite infectivity: a new approach of the *Biomphalaria glabrata*/*Schistosoma mansoni* compatibility polymorphism. *Infect Genet Evol* (2014) 26:80–8. doi:10.1016/j.meegid.2014.04.025
- Sminia T, Barendsen L. A comparative morphological and enzyme histochemical study on blood cells of the fresh water snails *Lymnaea stagnalis*, *Biomphalaria glabrata*, and *Bulinus truncatus*. *J Morphol* (1980) 165:31–9. doi:10.1002/jmor.1051650104
- Bouchut A, Roger E, Coustau C, Gourbal B, Mitta G. Compatibility in the *Biomphalaria glabrata*/*Echinostoma caproni* model: potential involvement of adhesion genes. *Int J Parasitol* (2006) 36:175–84. doi:10.1016/j.ijpara.2005.09.009
- Hanelt B, Lun CM, Adema CM. Comparative ORESTES-sampling of transcriptomes of immune-challenged *Biomphalaria glabrata* snails. *J Invertebr Pathol* (2008) 99:192–203. doi:10.1016/j.jip.2008.06.002
- Adema CM, Hanington PC, Lun CM, Rosenberg GH, Aragon AD, Stout BA, et al. Differential transcriptomic responses of *Biomphalaria glabrata* (Gastropoda, Mollusca) to bacteria and metazoan parasites, *Schistosoma mansoni* and *Echinostoma paraense* (Digenea, Platyhelminthes). *Mol Immunol* (2010) 47:849–60. doi:10.1016/j.molimm.2009.10.019
- Eveland LK, Haseeb MA. Laboratory rearing of *Biomphalaria glabrata* snails and maintenance of larval schistosomes *in vivo* and *in vitro*. In: Toledo R, Fried B, editors. *Biomphalaria Snails and Larval Trematodes*. New York: Springer (2011). p. 33–55.
- Dheilly NM, Duval D, Mouahid G, Emans R, Allienne JF, Galinier R, et al. A family of variable immunoglobulin and lectin domain containing molecules in the snail *Biomphalaria glabrata*. *Dev Comp Immunol* (2015) 48:234–43. doi:10.1016/j.dci.2014.10.009
- Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Deweese-Scott C, et al. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res* (2011) 39:D225–9. doi:10.1093/nar/gkq1189
- Mone Y, Gourbal B, Duval D, Du Pasquier L, Kieffer-Jaquinod S, Mitta G. A large repertoire of parasite epitopes matched by a large repertoire of host immune receptors in an invertebrate host/parasite model. *PLoS Negl Trop Dis* (2010) 4:e813. doi:10.1371/journal.pntd.0000813
- Liu X, She XT, Zhu QF, Li H, Peng XX. Heterogeneous interactome between *Litopenaeus vannamei* plasma proteins and *Vibrio parahaemolyticus* outer membrane proteins. *Fish Shellfish Immunol* (2013) 34:192–8. doi:10.1016/j.fsi.2012.10.023
- Wu XJ, Dinguirard N, Sabat G, Lui HD, Gonzalez L, Gehring M, et al. Proteomic analysis of *Biomphalaria glabrata* plasma proteins with binding affinity to those expressed by early developing larval *Schistosoma mansoni*. *PLoS Pathog* (2017) 13:e1006081. doi:10.1371/journal.ppat.1006081
- Smit CH, Van Diepen A, Nguyen DL, Wührer M, Hoffman KF, Deelder AM, et al. Glycomic analysis of life stages of the human parasite *Schistosoma mansoni* reveals developmental expression profiles of functional and antigenic glycan motifs. *Mol Cell Proteomics* (2015) 14:1750–69. doi:10.1074/mcp.M115.048280
- Pees B, Yang W, Zarate-Potes A, Schulenburg H, Dierking K. High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates. *J Innate Immun* (2016) 8:129–42. doi:10.1159/000441475
- Yoshida S, Shimada T, Kondoh D, Kozuma Y, Ghosh AK, Jacobs-Lorena M, et al. Hemolytic C-type lectin CEL-III from sea cucumber expressed in transgenic mosquitoes impairs malaria parasite development. *PLoS Pathog* (2007) 3:e192. doi:10.1371/journal.ppat.0030192
- Wongpanya R, Sengprasert P, Amparyup P, Tassanakajon A. A novel C-type lectin in the black tiger shrimp *Penaeus monodon* functions as a pattern recognition receptor by binding and causing bacterial agglutination. *Fish Shellfish Immunol* (2017) 60:103–13. doi:10.1016/j.fsi.2016.11.042
- Gordy MA, Pila EA, Hanington PC. The role of fibrinogen-related proteins in the gastropod immune response. *Fish Shellfish Immunol* (2015) 46:39–49. doi:10.1016/j.fsi.2015.03.005
- Portet A, Pinaud S, Tetreau G, Galinier R, Cosseau C, Duval D, et al. Integrated multi-omic analyses in *Biomphalaria-Schistosoma* dialogue reveal the immunobiological significance of FREP-SmPoMuc interaction. *Dev Comp Immunol* (2017) 75:16–27. doi:10.1016/j.dci.2017.02.025
- Hokke CH, Fitzpatrick JM, Hoffmann KF. Integrating transcriptome, proteome and glycome analyses of *Schistosoma* biology. *Trends Parasitol* (2007) 23:165–74. doi:10.1016/j.pt.2007.02.007
- Peterson NA, Hokke CH, Deelder AM, Yoshino TP. Glycotope analysis in miracidia and primary sporocysts of *Schistosoma mansoni*: differential expression during the miracidium-to-sporocyst transformation. *Int J Parasitol* (2009) 39:1331–44. doi:10.1016/j.ijpara.2009.06.002
- Adema CM. Fibrinogen-related proteins (FREPs) in mollusks. *Results Probl Cell Differ* (2015) 57:111–29. doi:10.1007/978-3-319-20819-0_5

39. Galinier R, Portela J, Mone Y, Allienne JF, Henri H, Delbecq S, et al. Biomphalysin, a new beta pore-forming toxin involved in *Biomphalaria glabrata* immune defense against *Schistosoma mansoni*. *PLoS Pathog* (2013) 9:e1003216. doi:10.1371/journal.ppat.1003216
40. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, et al. Whole genome analysis of a schistosomiasis-transmitting freshwater snail. *Nat Commun* (2017) 8:15451. doi:10.1038/ncomms15451
41. Moran Y, Fredman D, Szczesny P, Grynberg M, Technau U. Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. *Mol Biol Evol* (2012) 29:2223–30. doi:10.1093/molbev/mss089
42. Schnitger AK, Yassine H, Kafatos FC, Osta MA. Two C-type lectins cooperate to defend *Anopheles gambiae* against Gram-negative bacteria. *J Biol Chem* (2009) 284:17616–24. doi:10.1074/jbc.M808298200
43. Oliveira-Nascimento L, Massari P, Wetzler LM. The role of TLR2 in infection and immunity. *Front Immunol* (2012) 3:79. doi:10.3389/fimmu.2012.00079
44. 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
45. Mitta G, Galinier R, Tisseyre P, Allienne JF, Girerd-Chambaz Y, Guillou F, et al. Gene discovery and expression analysis of immune-relevant genes from *Biomphalaria glabrata* hemocytes. *Dev Comp Immunol* (2005) 29:393–407. doi:10.1016/j.dci.2004.10.002
46. Deleury E, Dubreuil G, Elangovan N, Wajnberg E, Reichhart JM, Gourbal B, et al. Specific versus non-specific immune responses in an invertebrate species evidenced by a comparative de novo sequencing study. *PLoS One* (2012) 7:e32512. doi:10.1371/journal.pone.0032512
47. Sandiford SL, Dong Y, Pike A, Blumberg BJ, Bahia AC, Dimopoulos G. Cytoplasmic actin is an extracellular insect immune factor which is secreted upon immune challenge and mediates phagocytosis and direct killing of bacteria, and is a *Plasmodium* antagonist. *PLoS Pathog* (2015) 11:e1004631. doi:10.1371/journal.ppat.1004631
48. Bernimoulin M, Waters EK, Foy M, Steele BM, Sullivan M, Falet H, et al. Differential stimulation of monocytic cells results in distinct populations of microparticles. *J Thromb Haemost* (2009) 7:1019–28. doi:10.1111/j.1538-7836.2009.03434.x
49. Pradeu T, Cooper EL. The danger theory: 20 years later. *Front Immunol* (2012) 3:287. doi:10.3389/fimmu.2012.00287
50. Coates CJ, Decker H. Immunological properties of oxygen-transport proteins: hemoglobin, hemocyanin and hemerythrin. *Cell Mol Life Sci* (2017) 74:293–317. doi:10.1007/s00018-016-2326-7
51. Li D, Dong H, Li S, Munir M, Chen J, Luo Y, et al. Hemoglobin subunit beta interacts with the capsid protein and antagonizes the growth of classical swine fever virus. *J Virol* (2013) 87:5707–17. doi:10.1128/JVI.03130-12
52. Yan S, Rodrigues RG, Cahn-Hidalgo D, Walsh TJ, Roberts DD. Hemoglobin induces binding of several extracellular matrix proteins to *Candida albicans*. Identification of a common receptor for fibronectin, fibrinogen, and laminin. *J Biol Chem* (1998) 273:5638–44. doi:10.1074/jbc.273.10.5638
53. Garg R, Qadri A. Hemoglobin transforms anti-inflammatory *Salmonella typhi* virulence polysaccharide into a TLR-2 agonist. *J Immunol* (2010) 184:5980–7. doi:10.4049/jimmunol.0903512
54. Henry E, Fung N, Liu J, Drakakaki G, Coaker G. Beyond glycolysis: GAPDHs are multi-functional enzymes involved in regulation of ROS, autophagy, and plant immune responses. *PLoS Genet* (2015) 11:e1005199. doi:10.1371/journal.pgen.1005199
55. Apte SS, Parks WC. Metalloproteinases: a parade of functions in matrix biology and an outlook for the future. *Matrix Biol* (2015) 44–46:1–6. doi:10.1016/j.matbio.2015.04.005
56. Yuan S, Tao X, Huang S, Chen S, Xu A. Comparative immune systems in animals. *Annu Rev Anim Biosci* (2014) 2:235–58. doi:10.1146/annurev-animal-031412-103634
57. Brubaker SW, Bonham KS, Zanoni I, Kagan JC. Innate immune pattern recognition: a cell biological perspective. *Annu Rev Immunol* (2015) 33:257–90. doi:10.1146/annurev-immunol-032414-112240
58. Armitage SAO, Kurtz J, Brites D, Dong Y, Pasquier L, Wang HC. Dscam1 in pancrustacean immunity: current status and a look to the future. *Front Immunol* (2017) 8:662. doi:10.3389/fimmu.2017.00662
59. Nish S, Medzhitov R. Host defense pathways: role of redundancy and compensation in infectious disease phenotypes. *Immunity* (2011) 34:629–36. doi:10.1016/j.immuni.2011.05.009
60. Pradeu T, Jaeger S, Vivier E. The speed of change: towards a discontinuity theory of immunity? *Nat Rev Immunol* (2013) 13:764–9. doi:10.1038/nri3521

Conflict of Interest Statement: 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 funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Copyright © 2017 Tetreau, Pinaud, Portet, Galinier, Gourbal and Duval. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Value of a Comparative Approach to Understand the Complex Interplay between Microbiota and Host Immunity

Norma M. Morella* and Britt Koskella

Department of Integrative Biology, University of California, Berkeley, Berkeley, CA, United States

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, United States

Reviewed by:

Spencer V. Nyholm,
University of Connecticut,
United States
Silke Paust,
Baylor College of Medicine,
United States

*Correspondence:

Norma M. Morella
morella@berkeley.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 01 May 2017

Accepted: 24 August 2017

Published: 14 September 2017

Citation:

Morella NM and Koskella B (2017)
The Value of a Comparative
Approach to Understand the
Complex Interplay between
Microbiota and Host Immunity.
Front. Immunol. 8:1114.
doi: 10.3389/fimmu.2017.01114

The eukaryote immune system evolved and continues to evolve within a microbial world, and as such is critically shaped by—and in some cases even reliant upon—the presence of host-associated microbial species. There are clear examples of adaptations that allow the host to simultaneously tolerate and/or promote growth of symbiotic microbiota while protecting itself against pathogens, but the relationship between immunity and the microbiome reaches far beyond simple recognition and includes complex cross talk between host and microbe as well as direct microbiome-mediated protection against pathogens. Here, we present a broad but brief overview of how the microbiome is controlled by and interacts with diverse immune systems, with the goal of identifying questions that can be better addressed by taking a comparative approach across plants and animals and different types of immunity. As two key examples of such an approach, we focus on data examining the importance of early exposure on microbiome tolerance and immune system development and function, and the importance of transmission among hosts in shaping the potential coevolution between, and long-term stability of, host–microbiome associations. Then, by comparing existing evidence across short-lived plants, mouse model systems and humans, and insects, we highlight areas of microbiome research that are strong in some systems and absent in others with the hope of guiding future research that will allow for broad-scale comparisons moving forward. We argue that such an approach will not only help with identification of generalities in host–microbiome–immune interactions but also improve our understanding of the role of the microbiome in host health.

Keywords: timing of exposure, microbiome, defensive symbiont, microbiome transmission, microbiome variation

INTRODUCTION

Across kingdoms of life and branches of immunity, there are conserved characteristics in how hosts interact with their microbiome. Plants, mammals, and invertebrates are all able to differentiate between self and non-self, where they tolerate, and in some cases promote, associations with commensal or beneficial microbes while retaining the ability to sense and attack microbial pathogens. In many cases, beneficial microbes can even be considered an extension of the immune system through either competitive exclusion of pathogens or direct inhibition of their growth. Furthermore, non-pathogenic microbiota can both interact with and influence the adaptive and innate immune systems. Across these diverse host systems, the evidence for an interaction between

the microbiome and immunity is strong and unsurprising given that eukaryotic evolution has occurred entirely within a microbial world. The topic of immunity is highly complex and may seem inaccessible to those outside the discipline. However, from the perspective of evolutionary ecology, there is much that can be learned about host–microbe adaptation and coevolution through exploring topics in immunity. Therefore, our goal in this perspective piece is to broadly examine the key characteristics of known interplay between host immune systems and symbiotic bacteria across well-studied systems (the more detailed aspects of which, including microbiome variability among individuals, stability over time, mode of transmission, and evidence for host–microbiota co-speciation, we summarize in **Table 1**). We focus on the bacterial component of the microbiome but recognize the importance of fungal members and viruses, especially bacteriophages, given their known impact on the microbiome [e.g., Ref. (1, 2)] and possible role in host immunity [e.g., Ref. (3)].

The microbiome field is expanding rapidly, and doing so across systems, such as plants, mouse models, humans, and insects. We suggest that taking a broad comparative approach across the diverse mechanisms of immunity and host systems could offer unique insight to how host defenses are shaped by and shape the microbiome. Such an approach can, for example, help identify areas in which research is strong for certain systems but lacking in others. Here, we emphasize areas lacking in plant host systems, but which would likely elucidate important aspects of plant health and resilience against pathogens. Filling in such gaps across systems would allow for more powerful comparative studies and may inform predictions about host–microbe adaptations in light of larger issues such as antibiotic overuse and the spread of agricultural pathogens in a changing climate.

OVERVIEW OF HOST IMMUNE SYSTEMS

To begin, we offer a brief description of immunity in mammals, plants, and insects focusing primarily on the aspects of these systems that directly relate to known interactions with the microbiome [thorough and more discipline-specific descriptions of these immune systems exist elsewhere (103–106)]. The adaptive immune system is thought to have arisen in jawed fish ≈ 500 million years ago (107), whereas the innate immune system likely dates back to early eukaryotic cells themselves (105, 108). As microbial communities greatly predate the existence of multicellular eukaryotes, both branches of the immune system, therefore, evolved in the presence of microbes, and it follows that tolerance for commensal or mutualistic microbiota (those associated with hosts, but which do not cause disease) must have been a key factor in shaping the evolution of immunity. Innate immunity, found across all kingdoms of life, is largely non-specific and responds broadly to “non-self” cells. Its hallmarks include protective physical barriers and general pattern recognition receptors that sense non-self signals known as microbe-associated or pathogen-associated molecular patterns (MAMPs/PAMPs) and elicit generalized host responses (such as phagocytic ingestion of invading cells in animals or a hypersensitive response in plants).

Adaptive immunity is unique to vertebrates and responds to specific pathogens through detection of antigens *via* somatically generated receptors and specialized white blood cells (B and T cells). Cellular recognition of a specific pathogen leads to clonal expansion of the lymphocyte, resulting in daughter cells that produce the same antigen-specific antibodies. Memory cells are also produced, resulting in specific and long-lasting immunological memory. Other versions of adaptive immunity may exist (discussed below), but broadly speaking, adaptive immune responses are highly specific to particular pathogens or antigens, and the immune response changes over the course of a host’s lifetime.

In many cases in vertebrates, innate immunity is the first line of defense that elicits an adaptive immune response (103), and the two systems work cooperatively to combat infection. In comparison, plants rely on an innate immunity consisting of two primary responses to microbes (106). The first branch of the immune system recognizes MAMPs/PAMPs, such as flagellin and lipopolysaccharides (LPS), through the use of transmembrane pattern recognition receptors and results in pattern-triggered immunity. However, many plant pathogens have evolved to overcome these defenses through the use of effectors. Plants with resistance genes for specific pathogens can detect the effectors through NB-LRR proteins, which represent the second response to microbes: effector-triggered immunity. In addition, plants have physical barriers to infection such as cell wall defenses (109) and can also secrete antimicrobial peptides to ward off infection (110). Insect immunology shares characteristics with both plants and mammals; responses to microbial pathogens are highly diverse among host species, but most are considered innate. Immune responses include production of antimicrobial peptides, pattern recognition receptors, and responding to pathogens *via* circulating phagocytic cells. Evidence accrued over the last few decades also shows responses reminiscent of adaptive-type immunity, such as immunological memory *via* virus-derived complementary DNAs that generate systemic immunity (111) and highly specific immune priming both within and across generations (112), but the extent of such adaptive-type immunity and similarity to vertebrate defenses remains an open question in the field (113, 114). Taking into account the type of host immunity is essential when making hypotheses about adaptation and coevolution between host and microbiota. For example, in contrast to adaptive immunity, the innate immune response is a general resistance that can only respond to selection across host generations and not within, an important distinction when considering how plants might adapt in response to microbiota as compared to vertebrates.

As is becoming increasingly evident, the immune system influences both the composition and abundance of non-pathogenic microbiota in addition to its well-studied role in preventing pathogen establishment. In mammals, this is best studied in the gut microbiome, where differentiating between these diverse symbionts and colonizing pathogens is clearly a complex problem. The human immune system maintains a homeostatic relationship with commensal microbiota through mechanisms that include stratification and compartmentalization of the intestine, production of a mucous layer and antimicrobial proteins, and

TABLE 1 | Characteristics of host/microbiota association.

Host	Site/organ	Core taxa	Max O.T.U. range ^a	Variability between individuals		
				Early development	Adult	
Humans	Skin	4 phyla	14–182	May depend on delivery method in first weeks	Yes: more similarity between sites on one body than between bodies; variation between bodies depends on skin site	
	Gut	3 phyla	237–395	Lower alpha diversity than adults; higher variability between individuals	Yes: tend to fall into three enterotypes	
	Oral	6 phyla	600–19,000	Lower alpha diversity than adults; may depend on delivery method in first weeks	Low at genus level; higher at species level; varying results across studies	
Honey bee (<i>Apis mellifera</i>)	Gut	3 phyla; 6–10 species	68–99	Bacteria limited or absent in larvae/newly emerged bees	No: core species across geography, “tasks,” diets, and time	
Termite	Gut	11 phyla	357–5413	Lower alpha diversity than adults, limited larval studies	Dominated by some phyla, but very diet dependent	
Aphid	Gut	1 primary species	3–67	?	Primary symbionts: low; secondary: varies with geography or host plant	
<i>Drosophila melanogaster</i>	Gut	2–4 genera	21–122	High	Primarily same genera but varies based on diet and wild/lab strain	
Tsetse fly (<i>Glossina</i> spp.)	Various (Gut, bacteriome, milk gland organ)	3 phyla; 3 primary species	25 (one study)	Some primary symbionts: low; secondary: ?	Some primary symbionts: low; secondary: variable, but limited studies	
<i>Arabidopsis thaliana</i>	Rhizosphere	3–7 phyla	778–1,262	?	Soil and/or genotype driven	
<i>Arabidopsis thaliana</i>	Rhizosphere	4 phyla	840–5,057	?	Abundance variable; may converge over time	
Maize	Phyllosphere	4 phyla	396–61,067	Abundance/diversity variable between genotypes; high consistency with synthetic community	Genotype, soil type, geography driven	
Host	Site/organ	Heritability of microbiota	Evidence for role of host genetics shaping microbiota	Transmission	Temporal stability	Co-speciation
Humans	Skin	Yes	Yes	Maternal, contact, environmental	Skin site dependent; stability shown up to 2 years	Yes
	Gut	Yes	Yes	Maternal, environmental	May stabilize after adolescence; diet has an impact	?
	Oral	Yes	?	Maternal, environmental	May stabilize after adolescence	?
Honey bee	Gut	Yes	?	Social hive interactions within 3 days	Change in abundance	Limited evidence
Termite	Gut	Yes	?	Early social exchange/exposure <i>via</i> proctodeal trophallaxis	Stable; diet has an impact	Yes
Aphid	Gut	Primary: yes	Primary: yes; secondary: ?	Primary: ovarian transmission; secondary: vertical or horizontal	Primary: low; secondary: may vary over time	Primary: yes
<i>Drosophila melanogaster</i>	Gut	Yes	Yes	Larval ingestion of bacteria-coated egg shells	Composition and density change with gut development and age	Maybe with endosymbionts
Tsetse Fly (<i>Glossina</i> spp.)	Various (Gut, bacteriome, milk gland organ)	Primary: yes	Limited	Primary: maternal milk, germline; secondary: ?	?	Primary: yes
<i>Arabidopsis thaliana</i>	Rhizosphere	?	Yes	Horizontal: soil	Changes with developmental stage of plant; may converge after senescence	?
<i>Arabidopsis thaliana</i>	Rhizosphere	?	Yes	Horizontal: air, soil	Communities may converge over time	?
Maize	Phyllosphere	Yes	Yes	Vertical: seed; horizontal: soil	Known successional dynamics	?

(Continued)

TABLE 1 | Continued

Host	Site/organ	Reference
Humans	Skin	(4–10)
	Gut	(10–19)
	Oral	(5, 20–29)
Honey Bee (<i>Apis mellifera</i>)	Gut	(30, 31–36)
Termite	Gut	(37–42)
Aphid	Gut	(43–58)
<i>Drosophila melanogaster</i>	Gut	(59–72)
Tsetse Fly (<i>Glossina</i> spp.)	Various (Gut, bacteriome, milk gland organ)	(73–83)
<i>Arabidopsis thaliana</i>	Rhizosphere	(84–89)
<i>Arabidopsis thaliana</i>	Rhizosphere	(90–94)
Maize	Phyllosphere	(34, 95–102)

"Described in literature as 97–99% OTU cutoff or "phylogroup."
? Unknown or unproven.
"Primary" is not used to explicitly indicate "obligate" in this table.

limiting epithelial exposure and immune response (115), and through antibody targeting, which can limit bacterial spread and virulence, among other mechanisms (116). Interactions between the immune system and microbiota in the gut is a heavily studied field (115, 117–121), but we are still learning the ways in which aberrations in cross talk can cause or contribute to conditions, such as inflammatory bowel disease, obesity, and even certain types of cancer (122–126).

In insects, immune system responses also contribute to homeostasis with endosymbionts, reviewed in Ref. (127, 128), and restriction of other commensal bacteria to specific host compartments, as in the gut symbionts of termites (129), bees (32, 33), drosophila (130), and aphids (43), may also help maintain invertebrate symbiotic communities. The plant immune system is also critical in shaping the non-pathogenic microbiome [recently reviewed by Zipfel and Oldroyd (131)]. Two studies in *Arabidopsis thaliana* demonstrate that disrupting components of the plant immune system, such as the signaling molecules: salicylic acid (SA) and jasmonic acid (JA), influences microbial community composition: the first shows evidence for altered root microbiome communities in plant hosts lacking genes controlling production of SA compared to control plants (132) and the second shows altered microbial communities in plants with mutations in genes controlling ethylene response (another signaling molecule) and cuticle formation (90). Recent work in wheat also demonstrates a role for JA in shaping composition of the microbiome, and again in this case, activation of JA signaling pathways altered microbial diversity and composition of root endophytes (133). However, the importance of resistance genes and diversity, as well as the number of pattern recognition receptors, in shaping the plant microbiome remains an open question.

IMPORTANCE OF MICROBIOTA IN SHAPING HOST IMMUNITY

The interaction between the microbiome and the immune system is far from one-sided, as has been elegantly demonstrated in studies from germ-free mice. Microbiome establishment influences levels of circulating myeloid cells, macrophages in tissues, and proper functioning of innate lymphoid cells, all critical for a healthy immune response (134–136). Furthermore, microbiota is critical in development and function of components of adaptive immunity, such as B and T cell diversity and differentiation (119, 137) and there is evidence from germ-free mice supporting a role in natural killer cell priming and function (137, 138). In insects, microbes also play a role in immune system development. For example, tsetse flies lacking their vertically transmitted symbionts are immunocompromised through both altered expression of immunity-related genes and reduced levels of hemocytes, which play an important role in invertebrate immunity (83, 139–141). Altered gene expression and other physiological effects were also found in axenically raised *Drosophila melanogaster* (61). In plants, symbiotic bacteria influence host immunity by priming the plant for future exposure to pathogens through the induction of a systemic response, causing broad-range basal levels of protection. A primed plant can respond more rapidly and strongly to

pathogen invasion through a variety of mechanisms, including quicker closing of stomata, less sensitivity to bacterial manipulation of defenses, upregulation of defense-related genes, and a stronger SA-related immune responses (142). In some cases, the effects of priming can even be trans-generational through chromatin and histone modification, where the subsequent generation of primed plants exhibits enhanced resistance to bacterial, fungal, and herbivorous pathogens (143–146). Immunological priming by microbiota is also observed in arthropods, where it is often described as functional adaptive immunity, as it can occur within one generation or trans-generationally. Its effects have been observed in bumble bees (147, 148), beetles (149), daphnia (150), moths (151), and many more [summarized by Contreras-Garduño et al. (152)].

Host-associated microbiota can also directly influence host resistance against invading pathogens. Common in insects and also plants and mammals, the microbiome can serve a protective role that is independent of the host immune system through antagonism, competitive exclusion, or physical exclusion of pathogens, collectively referred to as defensive symbiosis (153, 154). For example, the mammalian skin microbiota is known to play a large role in pathogen recognition and infection prevention through amplification of immune responses (155) and production of antimicrobials (156). When germ-free mice were inoculated with gut microbiota from a non-mouse host source, they showed a decreased ability to fight infection against *Salmonella*, and particular bacterial strains seem to be required for normal adaptive immune response (157). More recently, it has been shown that a mildly pathogenic bacterium of *Caenorhabditis elegans* can evolve over time to protect its host against the more virulent pathogen, *Staphylococcus aureus* (158). The importance of such pairwise interactions have been demonstrated many times [reviewed in Ref. (159, 160)], and indeed has motivated many current biocontrol strategies, but an open question in the field is how such microbe-mediated protection might scale up to the whole microbiome level. This leads to the idea that by directly protecting their host against pathogens, microbiota could hinder the evolution of host resistance by relaxing selection on host populations and, therefore, increasing host reliance on the microbiome.

MICROBIOME TRANSMISSION AND TIMING OF EXPOSURE

In mammals, it is clear that early exposure to microbes is crucial to the development of both branches of the immune system (161), influencing not only immune development and response against pathogens but also tolerance to commensal or mutualistic microbiota (162). For example, pregnant female mice treated with antibiotics have been shown to have offspring with not only a depauperate microbiome but also decreased levels of blood neutrophils and precursor cells, resulting in higher susceptibility to infection and increased mortality rates as compared to control mice (163). In line with this, there is increasing evidence for a crucial window of opportunity for exposure to microbiota (135). A study in germ-free mice showed that introducing a healthy

microbiome to adult germ-free mice did not restore normal levels of invariant natural killer T cells nor did it lessen the physical effects of induced colitis (164), and altered exposure to bacterial species and their LPS subtypes in human infant guts may have lasting and detrimental effects on development of immunity (165). In the human neonate airway, disruption of microbiome formation as early as the first 2 weeks of life can result in lifelong susceptibility to allergic airway inflammation (166). There are additional documented links between dysbiosis of early-life microbiota and disease or health conditions later in life, reviewed elsewhere (167). Despite the accruing evidence from human and mouse systems, there has been little to no exploration of such a window of opportunity for microbiome–immune system interactions in other systems, such as plants or insects. It also remains unclear whether such early exposure effects should be limited to organisms with adaptive immunity or whether priming of innate immunity at different host developmental stages also affects host–microbiome interactions.

The clear role of early exposure to microbiota, at least in mammals, suggests that it would be advantageous for a community of beneficial microbes to be transmitted vertically from parent to offspring (e.g., through direct contact at birth, seeds, or transovarian) from generation to generation. Vertical transmission in humans may be impacted by delivery mode, as there is good evidence for differences in microbiome composition and diversity between infants delivered *via* vaginal birth versus those delivered *via* cesarean sections (12, 168), but it remains controversial how long-lived such effects are (4). In insects, symbionts are known to be maintained through both vertical transmission [for example, *Buchnera* in aphids; (169)] and other transmission mechanisms such as early social interactions [observed in bees; (36)], proctodeal exchange of fluids [e.g., in termites; (170)], or larval consumption of bacteria-coated egg shells [as observed in *Drosophila*; (59, 70, 171)]. Interestingly, non-social bees (in which early social transmission of symbionts would not occur) do not seem to share the core microbiome that is observed among social bees (33).

Transmission of microbiota in plants can occur vertically through the seeds, or horizontally from the soil and surrounding environment. Plants ranging from trees to grasses are known to harbor bacteria in their seeds, many of which are reported to promote plant health (172–174). Despite this, there is no evidence that plants actively select for transmission of specific microbial communities, and there are no clear examples of adaptations to ensure seed-mediated transmission. Intuitively, vertical transmission of a microbiome or symbionts would allow for maintenance of key members of the microbial community across generations, as beneficial microbes would have primary access to both spatial niches and environmental nutrients provided by seedlings. Interestingly, plants have been shown to have differential onset of resistance to pathogens throughout their life-stages, something described as age-related resistance (ARR) or developmental resistance (175–177). However, much of the work on ARR investigates exposure and resistance to specific pathogens throughout the developmental stage of the plant and does not address if there is a window of opportunity for microbial exposure in general, as observed in mammals.

CONCLUSION AND OPPORTUNITIES FOR ADVANCEMENT IN THE FIELD

Unsurprisingly, that the microbiome is both shaped by and shapes the host immune system is a common feature of eukaryotes. However, the mechanisms underlying such cross talk are highly variable. Although we now have a foundation of knowledge demonstrating the microbiome's role in immune system development and function, key-questions remain unanswered across systems. One specific area for advancement is exploring the importance of both vertical transmission and timing of microbiome exposure across systems with diverse immune mechanisms. For example, despite the known importance of timing of exposure in mouse models and vertical transmission in insects, to our knowledge there are no studies to date that test the importance of timing of non-pathogenic microbial exposure on microbiome establishment or immune function in plants, and few in invertebrates. Would a seedling exposed to beneficial microbes mount as strong of a response as an older plant? And would exposure of otherwise sterile adult plants result in the same successional dynamics of microbiome establishment as has been observed in seedlings of some plant species (93, 178, 179)? Given that we know resistance to pathogens can change throughout the life cycle of a plant, research focused on age-related tolerance and recruitment of beneficial symbionts and plant-growth promoting bacteria has large implications in agricultural practices, such as seed treatment, greenhouse germination, and age-structured planting.

Vertical transmission also ensures stable associations between hosts and their microbiomes over evolutionary time and, therefore, sets the stage for long-term coevolution and even co-speciation. There is good evidence for vertical transmission of microbiota through gametes, secretions, or birth/delivery from across systems, but how this relates to coevolution between microbiota and their hosts remains to be determined. Long-term associations between hosts and microbiota can be uncovered through examination of co-speciation events, and these have been described in insects, such as aphids (51, 52), social bees (31), and termites (40). Furthermore, recent evidence from the hominid phylogeny also strongly supports this phenomenon (180). However, in plant systems, the current evidence is limited to a few pairwise host–symbiont interactions (181, 182). To understand the ways in which microbiota–immunity interactions influence stable association, transmission, and potentially coevolution in organisms such as plants, it may be wise to start by looking for similarities in established examples, such as the reduced genomes of symbionts commonly found in insect symbionts (183), nutritional dependence on symbionts, or physical partitioning of microbiota within the host.

Another area of advancement involves taking into account the whole suite of microbiomes associated with hosts. Despite what we know about spatially distinct microbiota in humans (5) and plants (184, 185), there are still large biases toward the below-ground (rhizosphere) microbiota of plants and the gut microbiota of vertebrates and insects. As more multi-tissue microbiome studies are generated across systems, we will be in a better position to uncover general patterns of potential cross

talk among microbiomes within a host, differences in the types of pathogens being protected against across tissues, and perhaps even the role of distinct microbiomes in shaping tissue tropism of pathogens. Furthermore, parallel studies of spatially distinct microbiomes in insects could offer nice insight into, for example, the roles of internally versus externally colonizing microbiota in shaping disease susceptibility, as well as how the host immune response regulates multiple microbiomes simultaneously.

Finally, the field is still limited by challenges in data interpretation for large, complex, and dynamic microbiome systems, explaining many of the open questions regarding heritability, temporal dynamics, and co-speciation (highlighted in **Table 1**). However, addressing these questions is increasingly feasible through rapidly advancing sequencing and bioinformatics approaches and the compilation of biologically representative synthetic communities. Although we are still some way from having large cross-system comparative microbiome studies, as sequencing costs continue to fall and data standardization across studies becomes more stringent, such meta-analyses will likely uncover larger “rules” of microbiome assembly, diversity, and interplay with host immunity. For example, plant-microbiome literature has forged the way in our understanding of how host genetics versus environment contribute to shaping the adult microbiome [e.g., Ref. (90, 186)], and recent work from humans now raises the question of whether similar rules are true for vertebrates (187). Another, more reductionist, approach for testing fundamental predictions about microbiome establishment genetic underpinning and immune system interactions is using synthetic microbiomes, as has been well-developed in plants (86, 90, 101, 132, 188). For example, a recent study in *D. melanogaster* explored colonization of gnotobiotic flies with specific strains of bacteria to document how host genotype influences microbial abundance levels (65). Though far from painting a complete picture, approaches such as this may also provide a means to study specific microbial adaptations to the immune systems of hosts across environmental conditions and genotypes. In conclusion, as we accumulate more data across systems, we can take more comparative and/or phylogenetic approaches to better understand the evolution of microbiome–immune system interaction mechanisms and to uncover conserved microbiome-mediated immune functions across systems. Such research has broad application to both human and agricultural health and is critical in light of the emergence of antibiotic and chemical-resistant pathogens and the common use of interventions that disrupt host-microbiome associations across systems.

AUTHOR CONTRIBUTIONS

NM and BK both contributed to the development of ideas and writing of this manuscript.

ACKNOWLEDGMENTS

The authors wish to thank the editors for the invitation to contribute to this research topic, the reviewers for their helpful feedback, and Callie Cuff for assistance in background research.

REFERENCES

- Koskella B. Phage-mediated selection on microbiota of a long-lived host. *Curr Biol* (2013) 23:1256–60. doi:10.1016/j.cub.2013.05.038
- Scanlan PD. Bacteria-bacteriophage coevolution in the human gut: implications for microbial diversity and functionality. *Trends Microbiol* (2017) 25:614–23. doi:10.1016/j.tim.2017.02.012
- Barr JJ, Auro R, Furlan M, Whiteson KL, Erb ML, Pogliano J, et al. Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proc Natl Acad Sci U S A* (2013) 110:10771–6. doi:10.1073/pnas.1305923110
- Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med* (2017) 23:314–26. doi:10.1038/nm.4272
- Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science* (2009) 326:1694–7. doi:10.1126/science.1177486
- Gao Z, Tseng C, Pei Z, Blaser MJ. Molecular analysis of human forearm superficial skin bacterial biota. *Proc Natl Acad Sci U S A* (2007) 104:2927–32. doi:10.1073/pnas.0607077104
- Grice EA, Kong HH, Renaud G, Young AC; NISC Comparative Sequencing Program, Bouffard GG, et al. A diversity profile of the human skin microbiota. *Genome Res* (2008) 18:1043–50. doi:10.1101/gr.075549.107
- Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and temporal diversity of the human skin microbiome. *Science* (2009) 324:1190–2. doi:10.1126/science.1171700
- Oh J, Byrd AL, Park M; NISC Comparative Sequencing Program, Kong HH, Segre JA. Temporal stability of the human skin microbiome. *Cell* (2016) 165:854–66. doi:10.1016/j.cell.2016.04.008
- Si J, Lee S, Park JM, Sung J, Ko G. Genetic associations and shared environmental effects on the skin microbiome of Korean twins. *BMC Genomics* (2015) 16:992. doi:10.1186/s12864-015-2131-y
- Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, et al. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* (2010) 107:18933–8. doi:10.1073/pnas.1007028107
- Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* (2010) 107:11971–5. doi:10.1073/pnas.1002601107
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. *Science* (2005) 308:1635–8. doi:10.1126/science.1110591
- Zoetendal EG, Akkermans ADL, Akkermans-van Vliet WM, de Visser JAGM, de Vos WM. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microb Ecol Health Dis* (2001) 13:129–34. doi:10.1080/089106001750462669
- Gill SR, Pop M, DeBoy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. *Science* (2006) 312:1355–9. doi:10.1126/science.1124234
- Greenhalgh K, Meyer KM, Aagaard KM, Wilmes P. The human gut microbiome in health: establishment and resilience of microbiota over a lifetime. *Environ Microbiol* (2016) 18:2103–16. doi:10.1111/1462-2920.13318
- Hopkins MJ, Sharp R, Macfarlane GT. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* (2001) 48:198–205. doi:10.1136/gut.48.2.198
- Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* (2011) 9:279–90. doi:10.1038/nrmicro2540
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* (2012) 486:222–7. doi:10.1038/nature11053
- Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* (2005) 43:5721–32. doi:10.1128/JCM.43.11.5721-5732.2005
- Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, et al. Bacterial diversity in the oral cavity of 10 healthy individuals. *ISME J* (2010) 4:962–74. doi:10.1038/ismej.2010.30
- De La Fuente C, Flores S, Moraga M. DNA from human ancient bacteria: a novel source of genetic evidence from archaeological dental calculus. *Archaeometry* (2013) 55:767–78. doi:10.1111/j.1475-4754.2012.00707.x
- Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, et al. The human oral microbiome. *J Bacteriol* (2010) 192:5002–17. doi:10.1128/JB.00542-10
- Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. *Nature* (2014) 509:357–60. doi:10.1038/nature13178
- Gomez A, Nelson KE. The oral microbiome of children: development, disease, and implications beyond oral health. *Microb Ecol* (2017) 73:492–503. doi:10.1007/s00248-016-0854-1
- Keijser BJE, Zaura E, Huse SM, van der Vossen JMBM, Schuren FHJ, Montijn RC, et al. Pyrosequencing analysis of the oral microflora of healthy adults. *J Dent Res* (2008) 87:1016–20. doi:10.1177/154405910808701104
- Nasidze I, Li J, Quinque D, Tang K, Stoneking M. Global diversity in the human salivary microbiome. *Genome Res* (2009) 19:636–43. doi:10.1101/gr.084616.108
- Stahring SS, Clemente JC, Corley RP, Hewitt J, Knights D, Walters WA, et al. Nurture trumps nature in a longitudinal survey of salivary bacterial communities in twins from early adolescence to early adulthood. *Genome Res* (2012) 22:2146–52. doi:10.1101/gr.140608.112
- Warinner C, Rodrigues JFM, Vyas R, Trachsel C, Shved N, Grossmann J, et al. Pathogens and host immunity in the ancient human oral cavity. *Nat Genet* (2014) 46:336–44. doi:10.1038/ng.2906
- Engel P, Kwong WK, McFrederick Q, Anderson KE, Barribeau SM, Chandler JA, et al. The bee microbiome: impact on bee health and model for evolution and ecology of host-microbe interactions. *MBio* (2016) 7(2):e02164-15.
- Koch H, Abrol DP, Li J, Schmid-Hempel P. Diversity and evolutionary patterns of bacterial gut associates of corbiculate bees. *Mol Ecol* (2013) 22:2028–44. doi:10.1111/mec.12209
- Kwong WK, Mancenido AL, Moran NA. Immune system stimulation by the native gut microbiota of honey bees. *R Soc Open Sci* (2017) 4:170003. doi:10.1098/rsos.170003
- Martinson VG, Danforth BN, Minckley RL, Rueppell O, Tingek S, Moran NA. A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* (2011) 20:619–28. doi:10.1111/j.1365-294X.2010.04959.x
- Mohr KI, Tebbe CC. Diversity and phylotype consistency of bacteria in the guts of three bee species (Apoidea) at an oilseed rape field. *Environ Microbiol* (2006) 8:258–72. doi:10.1111/j.1462-2920.2005.00893.x
- Moran NA. Genomics of the honey bee microbiome. *Curr Opin Insect Sci* (2015) 10:22–8. doi:10.1016/j.cois.2015.04.003
- Powell JE, Martinson VG, Urban-Mead K, Moran NA. Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera*. *Appl Environ Microbiol* (2014) 80:7378–87. doi:10.1128/AEM.01861-14
- Benjamino J, Graf J. Characterization of the core and caste-specific microbiota in the termite, *Reticulitermes flavipes*. *Front Microbiol* (2016) 7:171. doi:10.3389/fmicb.2016.00171
- Dietrich C, Köhler T, Brune A. The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol* (2014) 80:2261–9. doi:10.1128/AEM.04206-13
- Diouf M, Roy V, Mora P, Frechault S, Lefebvre T, Hervé V, et al. Profiling the succession of bacterial communities throughout the life stages of a higher termite *Nasutitermes arborum* (Termitidae, Nasutitermitinae) using 16S rRNA gene pyrosequencing. *PLoS One* (2015) 10:e0140014. doi:10.1371/journal.pone.0140014
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, et al. Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl Environ Microbiol* (2005) 71:6590–9. doi:10.1128/AEM.71.11.6590-6599.2005
- Tinker KA, Ottesen EA. The core gut microbiome of the American cockroach, *Periplaneta americana*, is stable and resilient to dietary shifts. *Appl Environ Microbiol* (2016) 82:6603–10. doi:10.1128/AEM.01837-16
- Wang Y, Su L, Huang S, Bo C, Yang S, Li Y, et al. Diversity and resilience of the wood-feeding higher termite *Mirnasutitermes shangchengensis* gut microbiota in response to temporal and diet variations. *Ecol Evol* (2016) 6:8235–42. doi:10.1002/ece3.2497
- Chen D-Q, Purcell AH. Occurrence and transmission of facultative endosymbionts in aphids. *Curr Microbiol* (1997) 34:220–5. doi:10.1007/s002849900172

44. Chong RA, Moran NA. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. *Proc Natl Acad Sci U S A* (2016) 113:13114–9. doi:10.1073/pnas.1610749113
45. Clark MA, Moran NA, Baumann P, Wernegreen JJ. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* (2000) 54:517–25. doi:10.1111/j.0014-3820.2000.tb00054.x
46. Fukatsu T, Nikoh N, Kawai R, Koga R. The secondary endosymbiotic bacterium of the pea aphid *Acyrtosiphon pisum* (Insecta: Homoptera). *Appl Environ Microbiol* (2000) 66:2748–58. doi:10.1128/AEM.66.7.2748-2758.2000
47. Funk DJ, Helbling L, Wernegreen JJ, Moran NA. Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc R Soc Lond B Biol Sci* (2000) 267:2517–21. doi:10.1098/rspb.2000.1314
48. Gauthier J-P, Outreman Y, Mieuze L, Simon J-C. Bacterial communities associated with host-adapted populations of pea aphids revealed by deep sequencing of 16S ribosomal DNA. *PLoS One* (2015) 10:e0120664. doi:10.1371/journal.pone.0120664
49. Haynes S, Darby AC, Daniell TJ, Webster G, van Veen FJF, Godfray HCJ, et al. Diversity of bacteria associated with natural aphid populations. *Appl Environ Microbiol* (2003) 69:7216–23. doi:10.1128/AEM.69.12.7216-7223.2003
50. Macdonald SJ, Thomas GH, Douglas AE. Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Mol Ecol* (2011) 20:2073–84. doi:10.1111/j.1365-294X.2011.05031.x
51. Moran NA, Munson MA, Baumann P, Ishikawa H. A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc R Soc Lond B Biol Sci* (1993) 253:167–71. doi:10.1098/rspb.1993.0098
52. Munson MA, Baumann P, Clark MA, Baumann L, Moran NA, Voegtlin DJ, et al. Evidence for the establishment of aphid-*Eubacterium* endosymbiosis in an ancestor of four aphid families. *J Bacteriol* (1991) 173:6321–4. doi:10.1128/jb.173.20.6321-6324.1991
53. Peccoud J, Bonhomme J, Mahéo F, de la Huerta M, Cosson O, Simon J-C. Inheritance patterns of secondary symbionts during sexual reproduction of pea aphid biotypes. *Insect Sci* (2014) 21:291–300. doi:10.1111/1744-7917.12083
54. Russell JA, Weldon S, Smith AH, Kim KL, Hu Y, Łukasik P, et al. Uncovering symbiont-driven genetic diversity across North American pea aphids. *Mol Ecol* (2013) 22:2045–59. doi:10.1111/mec.12211
55. Sandström JP, Russell JA, White JP, Moran NA. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Mol Ecol* (2001) 10:217–28. doi:10.1046/j.1365-294X.2001.01189.x
56. Smith AH, Łukasik P, O'Connor MP, Lee A, Mayo G, Drott MT, et al. Patterns, causes and consequences of defensive microbiome dynamics across multiple scales. *Mol Ecol* (2015) 24:1135–49. doi:10.1111/mec.13095
57. Tsuchida T, Koga R, Shibao H, Matsumoto T, Fukatsu T. Diversity and geographic distribution of secondary endosymbiotic bacteria in natural populations of the pea aphid, *Acyrtosiphon pisum*. *Mol Ecol* (2002) 11:2123–35. doi:10.1046/j.1365-294X.2002.01606.x
58. Zytynska SE, Weisser WW. The natural occurrence of secondary bacterial symbionts in aphids. *Ecol Entomol* (2016) 41:13–26. doi:10.1111/een.12281
59. Bakula M. The persistence of a microbial flora during postembryogenesis of *Drosophila melanogaster*. *J Invertebr Pathol* (1969) 14:365–74. doi:10.1016/0022-2011(69)90163-3
60. Broderick NA, Lemaitre B. Gut-associated microbes of *Drosophila melanogaster*. *Gut Microbes* (2012) 3:307–21. doi:10.4161/gmic.19896
61. Broderick NA, Buchon N, Lemaitre B. Microbiota-induced changes in *Drosophila melanogaster* host gene expression and gut morphology. *MBio* (2014) 5:e01117–14. doi:10.1128/mBio.01117-14
62. Chandler JA, Lang JM, Bhatnagar S, Eisen JA, Kopp A. Bacterial communities of diverse *Drosophila* species: ecological context of a host-microbe model system. *PLoS Genet* (2011) 7:e1002272. doi:10.1371/journal.pgen.1002272
63. Chaston JM, Dobson AJ, Newell PD, Douglas AE. Host genetic control of the microbiota mediates the *Drosophila* nutritional phenotype. *Appl Environ Microbiol* (2015) 82:671–9. doi:10.1128/AEM.03301-15
64. Corby-Harris V, Pontaroli AC, Shimkets LJ, Bennetzen JL, Habel KE, Promislow DEL. Geographical distribution and diversity of bacteria associated with natural populations of *Drosophila melanogaster*. *Appl Environ Microbiol* (2007) 73:3470–9. doi:10.1128/AEM.02120-06
65. Early AM, Shanmugarajah N, Buchon N, Clark AG. *Drosophila* genotype influences commensal bacterial levels. *PLoS One* (2017) 12:e0170332. doi:10.1371/journal.pone.0170332
66. Elgart M, Stern S, Salton O, Gnainsky Y, Heifetz Y, Soen Y. Impact of gut microbiota on the fly's germ line. *Nat Commun* (2016) 7:11280. doi:10.1038/ncomms11280
67. Ilinsky Y. Coevolution of *Drosophila melanogaster* mtDNA and *Wolbachia* genotypes. *PLoS One* (2013) 8:e54373. doi:10.1371/journal.pone.0054373
68. Mistry R, Kounatidis I, Ligoxygakis P. Interaction between familial transmission and a constitutively active immune system shapes gut microbiota in *Drosophila melanogaster*. *Genetics* (2017) 206:889–904. doi:10.1534/genetics.116.190215
69. Ren C, Webster P, Finkel SE, Tower J. Increased internal and external bacterial load during *Drosophila* aging without life-span trade-off. *Cell Metab* (2007) 6:144–52. doi:10.1016/j.cmet.2007.06.006
70. Storelli G, Defaye A, Erkosar B, Hols P, Royet J, Leulier F. *Lactobacillus plantarum* promotes *Drosophila* systemic growth by modulating hormonal signals through TOR-dependent nutrient sensing. *Cell Metab* (2011) 14:403–14. doi:10.1016/j.cmet.2011.07.012
71. Wong AC-N, Chaston JM, Douglas AE. The inconstant gut microbiota of *Drosophila* species revealed by 16S rRNA gene analysis. *ISME J* (2013) 7:1922–32. doi:10.1038/ismej.2013.86
72. Wong CNA, Ng P, Douglas AE. Low-diversity bacterial community in the gut of the fruitfly *Drosophila melanogaster*. *Environ Microbiol* (2011) 13:1889–900. doi:10.1111/j.1462-2920.2011.02511.x
73. Aksoy E, Telleria EL, Echodu R, Wu Y, Okedi LM, Weiss BL, et al. Analysis of multiple tsetse fly populations in Uganda reveals limited diversity and species-specific gut microbiota. *Appl Environ Microbiol* (2014) 80:4301–12. doi:10.1128/AEM.00079-14
74. Attardo GM, Lohs C, Heddi A, Alam UH, Yildirim S, Aksoy S. Analysis of milk gland structure and function in *Glossina morsitans*: milk protein production, symbiont populations and fecundity. *J Insect Physiol* (2008) 54:1236–42. doi:10.1016/j.jinsphys.2008.06.008
75. Chen X, Li S, Aksoy S. Concordant evolution of a symbiont with its host insect species: molecular phylogeny of genus *Glossina* and its bacteriome-associated endosymbiont, *Wigglesworthia glossinidia*. *J Mol Evol* (1999) 48:49–58. doi:10.1007/PL00006444
76. Cheng Q, Ruel TD, Zhou W, Moloo SK, Majiwa P, O'Neill SL, et al. Tissue distribution and prevalence of *Wolbachia* infections in tsetse flies, *Glossina* spp. *Med Vet Entomol* (2000) 14:44–50. doi:10.1046/j.1365-2915.2000.00202.x
77. Geiger A, Fardeau M-L, Grebaut P, Vatunga G, Josénando T, Herder S, et al. First isolation of *Enterobacter*, *Enterococcus*, and *Acinetobacter* spp as inhabitants of the tsetse fly (*Glossina palpalis palpalis*) midgut. *Infect Genet Evol* (2009) 9:1364–70. doi:10.1016/j.meegid.2009.09.013
78. Geiger A, Fardeau M-L, Njiokou F, Joseph M, Asonganyi T, Ollivier B, et al. Bacterial diversity associated with populations of *Glossina* spp. from Cameroon and distribution within the campo sleeping sickness focus. *Microb Ecol* (2011) 62:632–43. doi:10.1007/s00248-011-9830-y
79. Geiger A, Ponton F, Simo G. Adult blood-feeding tsetse flies, trypanosomes, microbiota and the fluctuating environment in sub-Saharan Africa. *ISME J* (2015) 9:1496–507. doi:10.1038/ismej.2014.236
80. Hamidou Soumana I, Simo G, Njiokou F, Tchicaya B, Abd-Alla AMM, Cuny G, et al. The bacterial flora of tsetse fly midgut and its effect on trypanosome transmission. *J Invertebr Pathol* (2013) 112:S89–93. doi:10.1016/j.jip.2012.03.029
81. Lindh JM, Lehane MJ. The tsetse fly *Glossina fuscipes fuscipes* (Diptera: Glossina) harbours a surprising diversity of bacteria other than symbionts. *Antonie Van Leeuwenhoek* (2011) 99:711–20. doi:10.1007/s10482-010-9546-x
82. Rio RVM, Wu YN, Filardo G, Aksoy S. Dynamics of multiple symbiont density regulation during host development: tsetse fly and its microbial flora. *Proc Biol Sci* (2006) 273:805–14. doi:10.1098/rspb.2005.3399
83. Wang J, Weiss BL, Aksoy S. Tsetse fly microbiota: form and function. *Front Cell Infect Microbiol* (2013) 3:69. doi:10.3389/fcimb.2013.00069
84. Bulgarelli D, Rott M, Schlaeppli K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, et al. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* (2012) 488:91–5. doi:10.1038/nature11336
85. Chaparro JM, Badri DV, Vivanco JM. Rhizosphere microbiome assemblage is affected by plant development. *ISME J* (2014) 8:790–803. doi:10.1038/ismej.2013.196

86. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, et al. Defining the core *Arabidopsis thaliana* root microbiome. *Nature* (2012) 488:86–90. doi:10.1038/nature11237
87. Micallef SA, Shiaris MP, Colón-Carmona A. Influence of *Arabidopsis thaliana* accessions on rhizobacterial communities and natural variation in root exudates. *J Exp Bot* (2009) 60:1729–42. doi:10.1093/jxb/erp053
88. Micallef SA, Channer S, Shiaris MP, Colón-Carmona A. Plant age and genotype impact the progression of bacterial community succession in the *Arabidopsis* rhizosphere. *Plant Signal Behav* (2009) 4:777–80. doi:10.4161/psb.4.8.9229
89. Schlaeppli K, Dombrowski N, Oter RG, van Themaat EVL, Schulze-Lefert P. Quantitative divergence of the bacterial root microbiota in *Arabidopsis thaliana* relatives. *Proc Natl Acad Sci U S A* (2014) 111:585–92. doi:10.1073/pnas.1321597111
90. Bodenhausen N, Bortfeld-Miller M, Ackermann M, Vorholt JA. A synthetic community approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet* (2014) 10:e1004283. doi:10.1371/journal.pgen.1004283
91. Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, et al. Genome-wide association study of *Arabidopsis thaliana*'s leaf microbial community. *Nat Commun* (2014) 5:5320. doi:10.1038/ncomms6320
92. Kniskern JM, Traw MB, Bergelson J. Salicylic acid and jasmonic acid signaling defense pathways reduce natural bacterial diversity on *Arabidopsis thaliana*. *Mol Plant Microbe Interact* (2007) 20:1512–22. doi:10.1094/MPMI-20-12-1512
93. Maignien L, DeForce EA, Chafee ME, Eren AM, Simmons SL. Ecological succession and stochastic variation in the assembly of *Arabidopsis thaliana* phyllosphere communities. *Mbio* (2014) 5:e00682–13. doi:10.1128/mBio.00682-13
94. Reisberg EE, Hildebrandt U, Riederer M, Hentschel U. Distinct phyllosphere bacterial communities on *Arabidopsis* wax mutant leaves. *PLoS One* (2013) 8:e78613. doi:10.1371/journal.pone.0078613
95. Bouffaud M-L, Kyselková M, Gouesnard B, Grundmann G, Muller D, Moëgne-Loccoz Y. Is diversification history of maize influencing selection of soil bacteria by roots? *Mol Ecol* (2012) 21:195–206. doi:10.1111/j.1365-294X.2011.05359.x
96. Bruseti L, Francia P, Bertolini C, Pagliuca A, Borin S, Sorlini C, et al. Bacterial communities associated with the rhizosphere of transgenic Bt 176 maize (*Zea mays*) and its non transgenic counterpart. *Plant Soil* (2005) 266:11–21. doi:10.1007/s11104-005-5399-x
97. Cavaglieri L, Orlando J, Etcheverry M. Rhizosphere microbial community structure at different maize plant growth stages and root locations. *Microbiol Res* (2009) 164:391–9. doi:10.1016/j.micres.2007.03.006
98. Johnston-Monje D, Raizada MN. Conservation and diversity of seed associated endophytes in *zea* across boundaries of evolution, ethnography and ecology. *PLoS One* (2011) 6:e20396. doi:10.1371/journal.pone.0020396
99. Johnston-Monje D, Lundberg DS, Lazarovits G, Reis VM, Raizada MN. Bacterial populations in juvenile maize rhizospheres originate from both seed and soil. *Plant Soil* (2016) 405:337–55. doi:10.1007/s11104-016-2826-0
100. Li X, Rui J, Mao Y, Yannarell A, Mackie R. Dynamics of the bacterial community structure in the rhizosphere of a maize cultivar. *Soil Biol Biochem* (2014) 68:392–401. doi:10.1016/j.soilbio.2013.10.017
101. Niu B, Paulson JN, Zheng X, Kolter R. Simplified and representative bacterial community of maize roots. *Proc Natl Acad Sci U S A* (2017) 114:E2450–9. doi:10.1073/pnas.1616148114
102. Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangel JL, et al. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc Natl Acad Sci U S A* (2013) 110:6548–53. doi:10.1073/pnas.1302837110
103. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Chapter 24: The adaptive immune system. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science (2002). Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21070/>
104. Beckage N. *Insect Immunology*. Amsterdam: Academic Press (2008).
105. Janeway CA Jr, Travers P, Walport M, Shlomchik MJ. Principles of innate and adaptive immunity. *Immunobiology: The Immune System in Health and Disease*. 5th ed. New York: Garland Science (2001). Available from: <https://www.ncbi.nlm.nih.gov/books/NBK27090/>
106. Jones JDG, Dangl JL. The plant immune system. *Nature* (2006) 444:323–9. doi:10.1038/nature05286
107. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet* (2010) 11:47–59. doi:10.1038/nrg2703
108. Buchmann K. Evolution of innate immunity: clues from invertebrates via fish to mammals. *Front Immunol* (2014) 5:459. doi:10.3389/fimmu.2014.00459
109. Underwood W. The plant cell wall: a dynamic barrier against pathogen invasion. *Front Plant Sci* (2012) 3:85. doi:10.3389/fpls.2012.00085
110. Mansfield JW. Antimicrobial compounds and resistance. In: Slusarenko AJ, Fraser RSS, van Loon LC, editors. *Mechanisms of Resistance to Plant Diseases*. Netherlands: Springer (2000). p. 325–70.
111. Tassetto M, Kunitomi M, Andino R. Circulating immune cells mediate a systemic RNAi-based adaptive antiviral response in *Drosophila*. *Cell* (2017) 169:314–25.e13. doi:10.1016/j.cell.2017.03.033
112. Dhinaut J, Chogne M, Moret Y. Immune priming specificity within and across generations reveals the range of pathogens affecting evolution of immunity in an insect. *J Anim Ecol* (2017). doi:10.1111/1365-2656.12661
113. Cooper D, Eleftherianos I. Memory and specificity in the insect immune system: current perspectives and future challenges. *Front Immunol* (2017) 8:539. doi:10.3389/fimmu.2017.00539
114. Rowley AF, Powell A. Invertebrate immune systems-specific, quasi-specific, or non-specific? *J Immunol* (2007) 179:7209–14. doi:10.4049/jimmunol.179.11.7209
115. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* (2012) 336:1268–73. doi:10.1126/science.1223490
116. Kubinak JL, Round JL. Do antibodies select a healthy microbiota? *Nat Rev Immunol* (2016) 16:767–74. doi:10.1038/nri.2016.114
117. Barroso-Batista J, Demengeot J, Gordo I. Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat Commun* (2015) 6:8945. doi:10.1038/ncomms9945
118. Feng T, Elson CO. Adaptive immunity in the host-microbiota dialog. *Mucosal Immunol* (2011) 4:15–21. doi:10.1038/mi.2010.60
119. Honda K, Littman DR. The microbiota in adaptive immune homeostasis and disease. *Nature* (2016) 535:75–84. doi:10.1038/nature18848
120. Lee YK, Mazmanian SK. Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* (2010) 330:1768–73. doi:10.1126/science.1195568
121. Thaiss CA, Levy M, Suez J, Elinav E. The interplay between the innate immune system and the microbiota. *Curr Opin Immunol* (2014) 26:41–8. doi:10.1016/j.coi.2013.10.016
122. Alam A, Leoni G, Quiros M, Wu H, Desai C, Nishio H, et al. The microenvironment of injured murine gut elicits a local pro-restitutive microbiota. *Nat Microbiol* (2016) 1:15021. doi:10.1038/nmicrobiol.2015.21
123. Alam A, Wu H, Matthews J, Saedi B, Jones R, Neish A. O-010 microbial metabolites establish a gradient of protons, which mediate interkingdom host-microbiome cross-talk and maintain intestinal homeostasis via two proton-sensing GPRs. *Inflamm Bowel Dis* (2017) 23(Suppl 1):S4. doi:10.1097/01.MIB.0000512520.76586.47
124. Palm NW, de Zoete MR, Cullen TW, Barry NA, Stefanowski J, Hao L, et al. Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* (2014) 158:1000–10. doi:10.1016/j.cell.2014.08.006
125. Weiss GA, Hentt T. Mechanisms and consequences of intestinal dysbiosis. *Cell Mol Life Sci* (2017) 74:2959–77. doi:10.1007/s00018-017-2509-x
126. Zeevi D, Korem T, Segal E. Talking about cross-talk: the immune system and the microbiome. *Genome Biol* (2016) 17:50. doi:10.1186/s13059-016-0921-4
127. Eleftherianos I, Atri J, Accetta J, Castillo J. Endosymbiotic bacteria in insects: guardians of the immune system? *Front Physiol* (2013) 4:46. doi:10.3389/fphys.2013.00046
128. Ratzka C, Gross R, Feldhaar H. Endosymbiont tolerance and control within insect hosts. *Insects* (2012) 3:553–72. doi:10.3390/insects3020553
129. Ohkuma M, Brune A. Diversity, structure, and evolution of the termite gut microbial community. In: Bignell DE, Roisin Y, Lo N, editors. *Biology of Termites: A Modern Synthesis*. Netherlands: Springer (2010). p. 413–38.
130. Marianes A, Spradling AC. Physiological and stem cell compartmentalization within the *Drosophila* midgut. *Elife* (2013) 2:e00886. doi:10.7554/eLife.00886
131. Zipfel C, Oldroyd GED. Plant signalling in symbiosis and immunity. *Nature* (2017) 543:328–36. doi:10.1038/nature22009

132. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, et al. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* (2015) 349:860–4. doi:10.1126/science.aaa8764
133. Liu H, Carvalhais LC, Schenk PM, Dennis PG. Effects of jasmonic acid signalling on the wheat microbiome differ between body sites. *Sci Rep* (2017) 7:41766. doi:10.1038/srep41766
134. Erny D, Hrabě de Angelis AL, Jaitin D, Wieghofer P, Staszewski O, David E, et al. Host microbiota constantly control maturation and function of microglia in the CNS. *Nat Neurosci* (2015) 18:965–77. doi:10.1038/nn.4030
135. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. *Science* (2016) 352:539–44. doi:10.1126/science.aad9378
136. Thaïss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. *Nature* (2016) 535:65–74. doi:10.1038/nature18847
137. Sanos SL, Bui VL, Mortha A, Oberle K, Heners C, Johner C, et al. ROR γ t and commensal microflora are required for the differentiation of mucosal interleukin 22-producing NKp46+ cells. *Nat Immunol* (2009) 10:83–91. doi:10.1038/ni.1684
138. Ganal SC, Sanos SL, Kallfass C, Oberle K, Johner C, Kirschning C, et al. Priming of natural killer cells by nonmucosal mononuclear phagocytes requires instructive signals from commensal microbiota. *Immunity* (2012) 37:171–86. doi:10.1016/j.immuni.2012.05.020
139. Benoit JB, Vigneron A, Broderick NA, Wu Y, Sun JS, Carlson JR, et al. Symbiont-induced odorant binding proteins mediate insect host hematopoiesis. *eLife* (2017) 6:e19535. doi:10.7554/eLife.19535
140. Weiss BL, Wang J, Aksoy S. Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol* (2011) 9:e1000619. doi:10.1371/journal.pbio.1000619
141. Weiss BL, Maltz M, Aksoy S. Obligate symbionts activate immune system development in the tsetse fly. *J Immunol* (2012) 190(188):3395–403. doi:10.4049/jimmunol.1103691
142. Pastor V, Luna E, Mauch-Mani B, Ton J, Flors V. Primed plants do not forget. *Environ Exp Bot* (2013) 94:46–56. doi:10.1016/j.envexpbot.2012.02.013
143. Cerf-Bensussan N, Gaboriau-Routhiau V. The immune system and the gut microbiota: friends or foes? *Nat Rev Immunol* (2010) 10:735–44. doi:10.1038/nri2850
144. Luna E, Bruce TJA, Roberts MR, Flors V, Ton J. Next-generation systemic acquired resistance. *Plant Physiol* (2012) 158:844–53. doi:10.1104/pp.111.187468
145. Rasmann S, De Vos M, Casteel CL, Tian D, Halitschke R, Sun JY, et al. Herbivory in the previous generation primes plants for enhanced insect resistance. *Plant Physiol* (2012) 158:854–63. doi:10.1104/pp.111.187831
146. Slaughter A, Daniel X, Flors V, Luna E, Hohn B, Mauch-Mani B. Descendants of primed *Arabidopsis* plants exhibit resistance to biotic stress. *Plant Physiol* (2012) 158:835–43. doi:10.1104/pp.111.191593
147. Moret Y, Schmid-Hempel P. Entomology: immune defence in bumble-bee offspring. *Nature* (2001) 414:506–506. doi:10.1038/35107138
148. Sadd BM, Schmid-Hempel P. Insect immunity shows specificity in protection upon secondary pathogen exposure. *Curr Biol* (2006) 16:1206–10. doi:10.1016/j.cub.2006.04.047
149. Moret Y, Siva-Jothy MT. Adaptive innate immunity? Responsive-mode prophylaxis in the mealworm beetle, *Tenebrio molitor*. *Proc Biol Sci* (2003) 270:2475–80. doi:10.1098/rspb.2003.2511
150. McTaggart SJ, Wilson PJ, Little TJ. *Daphnia magna* shows reduced infection upon secondary exposure to a pathogen. *Biol Lett* (2012) 8:972–5. doi:10.1098/rsbl.2012.0581
151. Mikonranta L, Mappes J, Kaukonen M, Freitak D. Insect immunity: oral exposure to a bacterial pathogen elicits free radical response and protects from a recurring infection. *Front Zool* (2014) 11:23. doi:10.1186/1742-9994-11-23
152. Contreras-Garduño J, Lanz-Mendoza H, Franco B, Nava A, Pedraza-Reyes M, Canales-Lazcano J. Insect immune priming: ecology and experimental evidences. *Ecol Entomol* (2016) 41:351–66. doi:10.1111/een.12300
153. Clay K. Defensive symbiosis: a microbial perspective. *Funct Ecol* (2014) 28:293–8. doi:10.1111/1365-2435.12258
154. Hussa EA, Goodrich-Blair H. It takes a village: ecological and fitness impacts of multipartite mutualism. *Annu Rev Microbiol* (2013) 67:161–78. doi:10.1146/annurev-micro-092412-155723
155. Naik S, Bouladoux N, Wilhelm C, Molloy MJ, Salcedo R, Kastenmuller W, et al. Compartmentalized control of skin immunity by resident commensals. *Science* (2012) 337:1115–9. doi:10.1126/science.1225152
156. Gallo RL, Hooper LV. Epithelial antimicrobial defence of the skin and intestine. *Nat Rev Immunol* (2012) 12:503–16. doi:10.1038/nri3228
157. Chung H, Pamp SJ, Hill JA, Surana NK, Edelman SM, Troy EB, et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* (2012) 149:1578–93. doi:10.1016/j.cell.2012.04.037
158. King KC, Brockhurst MA, Vasieva O, Paterson S, Betts A, Ford SA, et al. Rapid evolution of microbe-mediated protection against pathogens in a worm host. *ISME J* (2016) 10:1915–24. doi:10.1038/ismej.2015.259
159. May G, Nelson P. Defensive mutualisms: do microbial interactions within hosts drive the evolution of defensive traits? *Funct Ecol* (2014) 28:356–63. doi:10.1111/1365-2435.12166
160. Oliver KM, Smith AH, Russell JA. Defensive symbiosis in the real world – advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Funct Ecol* (2014) 28:341–55. doi:10.1111/1365-2435.12133
161. Kaplan JL, Shi HN, Walker WA. The role of microbes in developmental immunologic programming. *Pediatr Res* (2011) 69:465–72. doi:10.1203/PDR.0b013e318217638a
162. Kelly D, King T, Aminov R. Importance of microbial colonization of the gut in early life to the development of immunity. *Mutat Res* (2007) 622:58–69. doi:10.1016/j.mrfmmm.2007.03.011
163. Deshmukh HS, Liu Y, Menkiti OR, Mei J, Dai N, O'Leary CE, et al. The microbiota regulates neutrophil homeostasis and host resistance to *Escherichia coli* K1 sepsis in neonatal mice. *Nat Med* (2014) 20:524–30. doi:10.1038/nm.3542
164. Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, et al. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* (2012) 336:489–93. doi:10.1126/science.1219328
165. Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* (2016) 165:842–53. doi:10.1016/j.cell.2016.04.007
166. Gollwitzer ES, Saglani S, Trompette A, Yadava K, Sherburn R, McCoy KD, et al. Lung microbiota promotes tolerance to allergens in neonates via PD-L1. *Nat Med* (2014) 20:642–7. doi:10.1038/nm.3568
167. Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. *Nat Med* (2016) 22:713–22. doi:10.1038/nm.4142
168. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* (2016) 8:343ra82. doi:10.1126/scitranslmed.aad7121
169. Douglas AE. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annu Rev Entomol* (1998) 43:17–37. doi:10.1146/annurev.ento.43.1.17
170. Nalepa CA. Origin of termite eusociality: trophallaxis integrates the social, nutritional, and microbial environments. *Ecol Entomol* (2015) 40:323–35. doi:10.1111/een.12197
171. Wong AC-N, Luo Y, Jing X, Franzburg S, Bost A, Douglas AE. The host as the driver of the microbiota in the gut and external environment of *Drosophila melanogaster*. *Appl Environ Microbiol* (2015) 81:6232–40. doi:10.1128/AEM.01442-15
172. Barret M, Briand M, Bonneau S, Prévieux A, Valière S, Bouchez O, et al. Emergence shapes the structure of the seed microbiota. *Appl Environ Microbiol* (2015) 81:1257–66. doi:10.1128/AEM.03722-14
173. Shade A, Jacques M-A, Barret M. Ecological patterns of seed microbiome diversity, transmission, and assembly. *Curr Opin Microbiol* (2017) 37:15–22. doi:10.1016/j.mib.2017.03.010
174. Truyens S, Weyens N, Cuypers A, Vangronsveld J. Bacterial seed endophytes: genera, vertical transmission and interaction with plants. *Environ Microbiol Rep* (2015) 7:40–50. doi:10.1111/1758-2229.12181
175. Develley-Rivière M-P, Galiana E. Resistance to pathogens and host developmental stage: a multifaceted relationship within the plant kingdom. *New Phytol* (2007) 175:405–16. doi:10.1111/j.1469-8137.2007.02130.x
176. Panter SN, Jones DA. Age-related resistance to plant pathogens. *Adv Bot Res* (2002) 38:251–80. doi:10.1016/S0065-2296(02)38032-7
177. Whalen MC. Host defence in a developmental context. *Mol Plant Pathol* (2005) 6:347–60. doi:10.1111/j.1364-3703.2005.00286.x

178. Copeland JK, Yuan L, Layeghifard M, Wang PW, Guttman DS. Seasonal community succession of the phyllosphere microbiome. *Mol Plant Microbe Interact* (2015) 28:274–85. doi:10.1094/MPMI-10-14-0331-FI
179. Tkacz A, Cheema J, Chandra G, Grant A, Poole PS. Stability and succession of the rhizosphere microbiota depends upon plant type and soil composition. *ISME J* (2015) 9:2349–59. doi:10.1038/ismej.2015.41
180. Moeller AH, Caro-Quintero A, Mjunga D, Georgiev AV, Lonsdorf EV, Muller MN, et al. Cospeciation of gut microbiota with hominids. *Science* (2016) 353:380–2. doi:10.1126/science.aaf3951
181. Schardl CL, Leuchtmann A, Chung K-R, Penny D, Siegel MR. Coevolution by common descent of fungal symbionts (*Epichloe* spp.) and grass hosts. *Mol Biol Evol* (1997) 14:133–133. doi:10.1093/oxfordjournals.molbev.a025746
182. Brundrett MC. Coevolution of roots and mycorrhizas of land plants. *New Phytol* (2002) 154:275–304. doi:10.1046/j.1469-8137.2002.00397.x
183. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* (2012) 10:13–26. doi:10.1038/nrmicro2670
184. Berg G, Grube M, Schlöter M, Smalla K. Unraveling the plant microbiome: looking back and future perspectives. *Front Microbiol* (2014) 5:148. doi:10.3389/fmicb.2014.00148
185. Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R, et al. Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *ISME J* (2012) 6:1378–90. doi:10.1038/ismej.2011.192
186. Knief C, Ramette A, Frances L, Alonso-Blanco C, Vorholt JA. Site and plant species are important determinants of the *Methylobacterium* community composition in the plant phyllosphere. *ISME J* (2010) 4:719–28. doi:10.1038/ismej.2010.9
187. Rothschild D, Weissbrod O, Barkan E, Korem T, Zeevi D, Costea PI, et al. Environmental factors dominate over host genetics in shaping human gut microbiota composition. *bioRxiv* (2017):150540. doi:10.1101/150540
188. Bai Y, Müller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* (2015) 528:364–9. doi:10.1038/nature16192

Conflict of Interest Statement: 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.

Copyright © 2017 Morella and Koskella. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Neutrophil Evolution and Their Diseases in Humans

Jennifer W. Leiding^{1,2*}

¹ Division of Allergy and Immunology, Department of Pediatrics, University of South Florida, Tampa, FL, United States,

² Johns Hopkins All Children's Hospital, St. Petersburg, FL, United States

Granulocytes have been preserved and have evolved across species, developing into cells that provide one of the first lines of host defense against pathogens. In humans, neutrophils are involved in early recognition and killing of infectious pathogens. Disruption in neutrophil production, emigration, chemotaxis, and function cause a spectrum of primary immune defects characterized by host susceptibility to invasive infections.

Keywords: neutrophil, neutropenia, chemotaxis, immunodeficiency, granulocytes

OPEN ACCESS

Edited by:

Uday Kishore,
Brunel University London,
United Kingdom

Reviewed by:

Abhishek D. Garg,
KU Leuven, Belgium
Ben Croker,
Boston Children's Hospital,
United States
Lubna Kouser,
Imperial College London,
United Kingdom

*Correspondence:

Jennifer W. Leiding
jleiding@health.usf.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 05 May 2017

Accepted: 07 August 2017

Published: 28 August 2017

Citation:

Leiding JW (2017) Neutrophil
Evolution and Their Diseases in
Humans.
Front. Immunol. 8:1009.
doi: 10.3389/fimmu.2017.01009

INTRODUCTION AND NEUTROPHIL EVOLUTION

All vertebrate species possess leukocytes, which divide into several different highly specialized cell lineages involved in immune response and tissue repair. Leukocytes fall into several classes, including granulocytes, macrophages, and lymphocytes. Granulocytes are differentiated from other leukocytes based on their morphology, including a segmented nucleus and staining properties of their cytoplasmic granules (1). Granulocytes are classified according to the morphology of their nucleus and staining properties of their granules (1).

Cells exhibiting some phagocytic activity, termed amebocytes, are seen early in phylogeny in basic invertebrates lacking a true body cavity (celom) or vascular system. Cnidarians, one of the most basic invertebrates contain a gelatinous matrix between an ectoderm and endoderm that contains multiple amebocytes that aid in digestion, are continuously proliferating stem cells and act as phagocytes. Invertebrates that possess a body cavity and vascular system contain a third dermal tissue, the mesoderm that forms mesothelium within the celom. The mesothelial walls are the site of origin of blood progenitor cells, termed hemocytes. Four major hemocyte classifications derive from the mesothelial wall and are carried through phylogeny from invertebrates to vertebrates: prohemocytes that evolve into immature blood precursor cells, hyaline hemocytes that progress to plasmacytes and then to monocytes, eleocytes that develop into other mesodermal derived tissues (i.e., the gastrointestinal tract), and granular hemocytes that develop into granulocytes involved in phagocytosis [reviewed in Ref. (1)].

The bone marrow is the principal hematopoietic organ of all vertebrates with the exception of fish. From the bloodstream, primitive hematopoietic stem cells (HSCs) arrive in the bone marrow in the last embryonic stages (1). Early in embryogenesis, erythrocytes initially are found within the yolk sac in the first 3 weeks of human gestation; the subsequent development of the vascular system allows blood cells to distribute to other embryonic tissues. By 6 weeks, the fetal liver is the major hematopoietic organ; the bone marrow takes over as the major site of hematopoiesis by the end of the second trimester (2). In the developing fetus, neutrophil progenitors are seen as early as the first trimester and increase in quantity nearly fourfold in the second trimester when the bone marrow becomes the major site of hematopoiesis. Circulating neutrophil counts rise abruptly and stabilize in the first 48–72 h of life (3). In preterm infants, the baseline neutrophil count is lower and there is no rise in neutrophil count in the first few days of life (4). In addition to quantitative impairments, neonatal neutrophils also exhibit many qualitative defects. Neutrophil adhesion is impaired by

decreased levels of L-selectin and the $\beta 2$ integrins CD18/CD11b and CD18/CD11a, which are adhesion molecules present on the surface of neutrophils and the endothelial surface and are imperative in neutrophil migration from the vasculature to sites of infection. L-selectin levels continue to decrease in the first 24–72 h, continue to be low in the first few weeks of life, and are even lower in preterm infants (5). Abnormal actin polymerization also is noted in the first few weeks of life causing a substantial decrease in directed migration *in vitro* (6). Although overall killing activity is not impaired, neonatal neutrophils have lower concentration of granular proteins (7). Despite these abnormalities, bone marrow production, neutrophil migration, and neutrophil activity mature rapidly, consistent with their role in serving as first responders to infectious and inflammatory stimuli.

Once developed, neutrophils are the dominant leukocyte population in humans. Neutrophils mature in the bone marrow in an orderly fashion from myeloblast to promyelocyte to myelocyte to metamyelocyte to band form and lastly the mature neutrophil. Only the latter two of these stages, the band form and mature neutrophil are present in peripheral blood. Neutrophils should have a three to four lobed nucleus and a granular cytoplasm (**Figure 1A**). Approximately 100 billion neutrophils enter and leave circulating blood every day (8). Neutrophils originate in the bone marrow and are released to vasculature when they have matured and are stimulated by invasive pathogens and inflammatory signals (**Figure 2**). Chemokines, small signaling molecules are potent chemoattractants for neutrophils to sites of infection or tissue injury. Migration toward the site of infection involves a complex multi-step process, including rolling adhesion of neutrophils on endothelial cells, firm adhesion of neutrophils, extravasation through the endothelium, and chemotactic migration. Upon migration to the site of infection, the neutrophil eliminates the invading pathogen utilizing a combination of NADPH oxidase derived reactive oxygen species, cytotoxic granule components, and neutrophil extracellular traps (8–10).

ZEBRAFISH NEUTROPHIL BIOLOGY

Neutrophils are one of the first cells to respond to sites of acute infection and cell damage, playing key roles in host defense against

infectious pathogens and in the development and resolution of inflammation. In order to understand the complex inflammatory process caused and resolved by neutrophils, models to investigate neutrophil biology have been developed. The short lifespan of human neutrophils is prohibitive in the investigation of neutrophil biology *in vivo* and causes *in vitro* genetic manipulation to be impractical. Because of these restrictions, a zebrafish model of neutrophil biology investigation has become widely accepted. The zebrafish neutrophil mirrors mammalian neutrophils by sharing similar morphology, and biochemical and functional features. It has a polymorphic nucleus, primary and secondary granules, and an NADPH oxidase (11, 12); multiple models of primary immunodeficiency diseases in zebrafish have been developed and studied [reviewed in Ref. (11)].

Using a zebrafish model, the mechanisms of neutrophil recruitment to sites of tissue damage have been elucidated. Damage-associated molecular patterns and chemokines recruit neutrophils effectively. Hydrogen peroxide, released by damaged tissue is one of the earliest attractants for neutrophils to sites of tissue injury. Chemokines, small signaling proteins that attract white blood cells to specific locations in the tissue first evolved ~650 million years ago in fish (13). Neutrophils respond to specific chemokine signatures from dying cells and are able to differentiate pathogen from non-infected cells. Pathogen response-like chemokines, CXCL1, CCL2, and CXCL10, are potent attractors of neutrophils leading to the development of inflammation and elimination of dying cells (14). Chemokine-induced neutrophil recruitment has been conserved across vertebrate species confirming the important process that chemoattraction plays in neutrophil recruitment (13).

Once an infection has resolved and cellular debris cleared, neutrophils must leave the site of tissue injury. High-resolution imaging of transparent zebrafish have uncovered reverse migration as a method of neutrophil resolution of inflammation. Reverse migration is a process whereby neutrophils migrate away from a site of infection or inflammation, a process regulated by pro-inflammatory cytokines (15).

When neutrophils persist in the tissue, inflammation persists and becomes chronic. Chronic neutrophil-driven inflammation has been linked to multiple autoimmune diseases and cancer

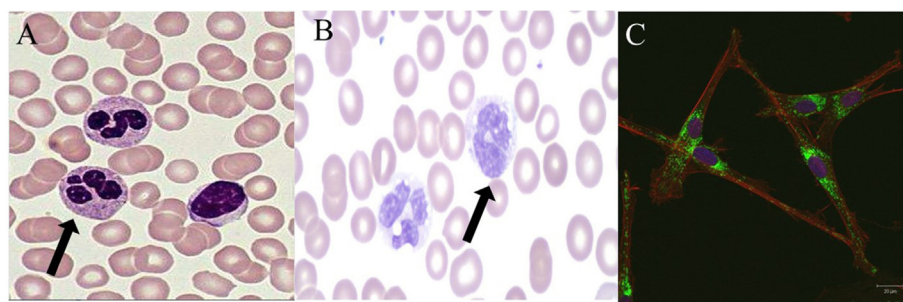
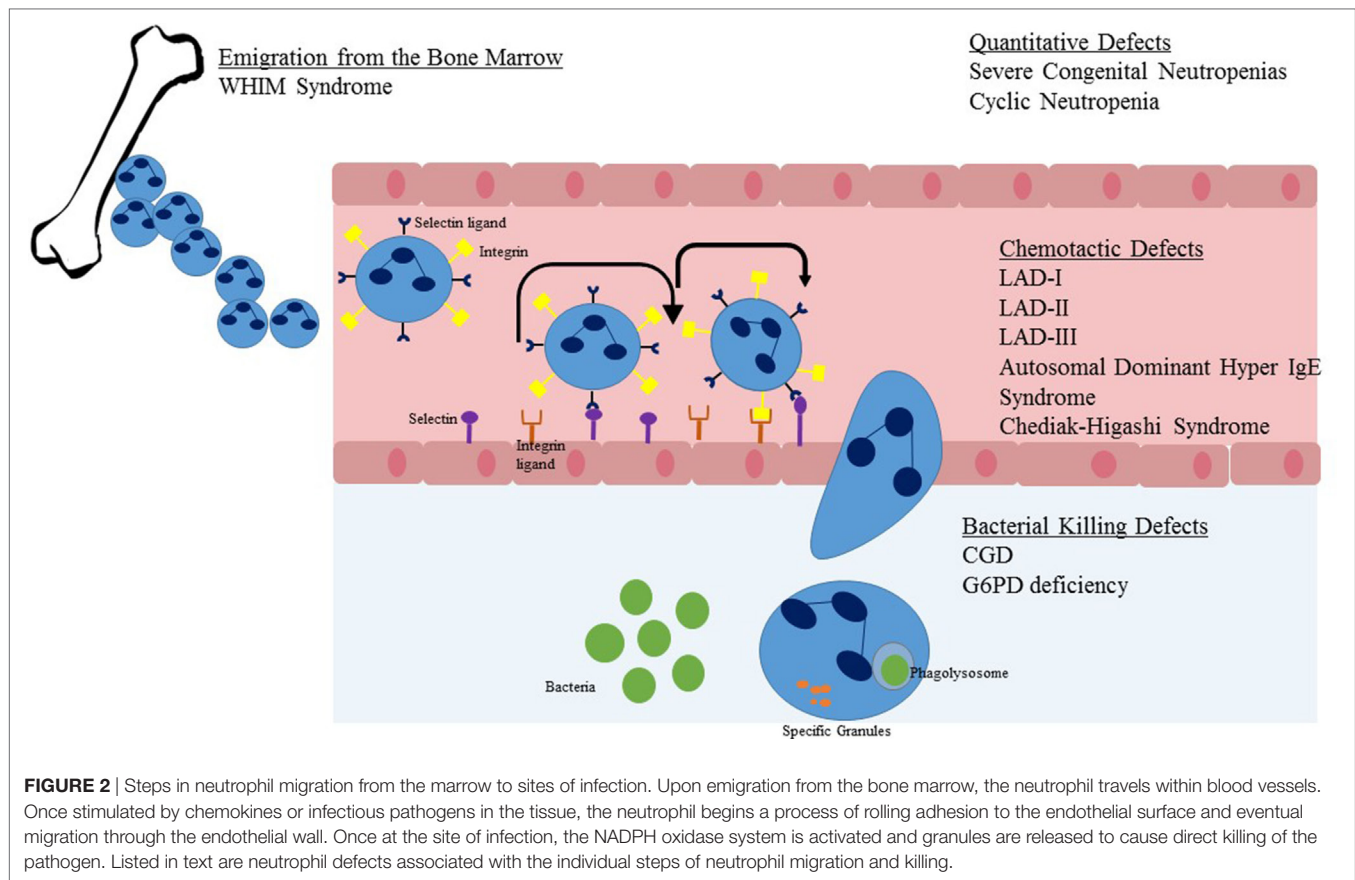


FIGURE 1 | Normal and abnormal neutrophil morphology. **(A)**. Normal human neutrophils (arrow) with characteristic multi-lobed nucleus. Primary and secondary granules are visualized in the cytoplasm. **(B)**. Human neutrophils (arrow) from a patient with a mutation in *CEBP-ε* causing specific granule deficiency. Neutrophils have a characteristic bilobed nucleus and absence of specific granules in the cytoplasm. **(C)**. Human giant fused cytoplasmic granules in a patient with Chediak-Higashi syndrome. Electron microscopy at 20 μ m.



progression. Neutrophils are found within many types of cancers and correlate with more aggressive disease and a poorer prognosis [reviewed in Ref. (16)]. The recruitment of neutrophils to tumor cells occurs in a similar fashion as to that of infected cells; chemokines and hydrogen peroxide produced by tumor cells attract neutrophils to tumor affected cells. Tumor-associated neutrophils are thought to play a role in cancer progression by affecting the extracellular matrix allowing for enhanced cancer cell proliferation and invasion. In addition, neutrophils also suppress anti-tumor immunity from other cell types. Targeting neutrophils has become a desirable therapeutic option for treatment of certain cancers [reviewed in Ref. (16)].

NEUTROPHIL DISEASES IN HUMANS

Immunodeficiency diseases afford novel insight into both normal function and pathophysiology. In terms of abnormal neutrophil function in humans, immunodeficiency that traces to abnormal neutrophil quantity or function is relatively common, occurring in approximately 20% of those with congenital primary immunodeficiency disorders. Disorders of neutrophils can be divided into four types affecting: neutrophil quantity, neutrophil granules, neutrophil chemotaxis, and neutrophil killing. This review focuses on what we have learned about the role of neutrophils in host protection from the four recognized classes of neutrophil disorders (17).

DISORDERS OF NEUTROPHIL QUANTITY

Neutrophils live about 5 days in circulation (18) and approximately 10^{11} neutrophils (8) are made by the bone marrow each day. Neutropenia can be mild [absolute neutrophil count (ANC) 1,000–1,500 cells/ μ L], moderate (ANC 500–1,000 cells/ μ L), or severe (ANC < 500 cells/ μ L). Severe neutropenia is more commonly found acutely rather than chronic. However, when found, cyclic and chronic forms of severe neutropenia cause increased susceptibility to soft tissue and invasive bacterial infections. There often is a characteristic lack of pus at sites of infection (19, 20).

The genetic basis of many of the congenital forms of neutropenia have been well elucidated (Table 1). More than 50% of patients with severe congenital neutropenias (SCNs) and nearly all patients with cyclic neutropenia have autosomal dominant (AD) monoallelic mutations in *ELANE*, the gene that encodes neutrophil elastase (21, 22). Those with cyclic disease typically present in the first year of life with recurring episodes of fever and severe neutropenia in a recurring cycle usually every 21 days. During their nadir, patients are susceptible to mouth sores, soft tissue, and invasive bacterial infections. Diagnosis of cyclic neutropenia includes serial complete blood counts to capture periods of neutropenia, often requiring monitoring of the neutrophil count 2 to 3 times per week for 6–8 weeks (23). Mutations in *ELANE* also cause SCN type 1 in which neutropenia is chronic and not cyclical. *ELANE* is responsible for triggering an aberrant

TABLE 1 | Congenital neutropenia disorders.

Disease	Genetic defect	Inheritance	Immunologic phenotype	Other manifestations	Reference
SCN type 1	ELANE	AD	Chronic or cyclic neutropenia		(22)
SCN2	GFI1	AD	Neutropenia, lymphopenia		(24)
SCN3	HAX1	AR	Neutropenia	Neurologic impairment	(25)
SCN4	G6PC3	AR	Neutropenia Thrombocytopenia	Congenital heart defects Facial dysmorphism increased visibility of superficial veins urogenital malformations endocrine abnormalities hearing loss skin hyperelasticity	(26)
XL congenital neutropenia	WAS	XL	Neutropenia Lymphopenia, myelodysplasia		(27)
Congenital neutropenia and hypopigmentation disorders					
Chediak–Higashi syndrome	LYST	AR	Neutropenia Natural killer (NK) cell dysfunction	Oculocutaneous albinism Neurologic impairment HLH	(28)
Hermansky–Pudlak syndrome type 2	AP3B1	AR	Neutropenia T and NK cell dysfunction	Oculocutaneous albinism	(29, 30)
Griscelli syndrome type 2	RAB27A	AR	Neutropenia NK cell dysfunction	Oculocutaneous albinism HLH	(31)
Other syndromes with neutropenia as a key feature					
Reticular dysgenesis	AK2	AR	Neutropenia Severe lymphopenia	Sensorineural hearing loss	(32)
Shwachman–Diamond syndrome	SBDS	AR	Neutropenia	Exocrine pancreatic insufficiency Skeletal dysplasia Liver and heart disease	(33)
Poikiloderma with neutropenia	C16ORF57	AR	Neutropenia	Poikiloderma, increased photosensitivity	(34)
Cartilage–Hair hypoplasia	RMRP	AR	Neutropenia T and NK cell lymphopenia	Autoimmune cytopenias Skeletal dysplasia Dwarfism	(35)
XL hyper IgM syndrome	CD40L	XL	Intermittent neutropenia Defective B cell class switching T and B cell defects		(36)
XL agammaglobulinemia	BTK	XL	Low to absent B cells Hypogammaglobulinemia		(37)
Barth syndrome	G4.5/TAZ	XL	Neutropenia	Cardioskeletal abnormalities Myopathy Growth retardation	(38)
Cohen syndrome	VPS13B/ COH	AR	Intermittent neutropenia	Psychomotor retardation Skeletal dysplasia hypotonia	(39)
Pearson syndrome		Mitochondrial DNA	Neutropenia	Bone marrow failure Exocrine pancreatic insufficiency Endocrine abnormalities Neuromuscular degeneration	(40)

AD, autosomal dominant; AR, autosomal recessive; G6PC3, glucose-6-phosphatase catalytic subunit 3; WAS, Wiskott–Aldrich syndrome; XL, X-linked; BTK, Bruton's tyrosine kinase.

stress response in the neutrophil and when mutated leads to premature apoptosis of the neutrophil.

Severe congenital neutropenia 2 is caused by mutations in *GFI1* a transcription factor that regulates normal neutrophil hematopoiesis. In addition to its effects on neutrophils, mutations in *GFI1* are associated with defects in lymphoid and myeloid cell lines (24).

Approximately 15% of SCNs are caused by autosomal recessive (AR) mutations in *HAX1* (SCN type 3). Patients with *HAX1* deficiency present with marked neutropenia and may have life

threatening bacterial infections as early as the newborn period. Although the exact role that *HAX1* plays in neutrophil ontogeny is unknown; one suggested mechanism is that *HAX1* is a major inhibitor of neutrophil apoptosis in myeloid cells and the neutropenia described in *HAX1*-deficient patients is due to the lack of anti-apoptotic effect (25).

Defects in glucose-6-phosphatase catalytic subunit 3 (*G6PC3*) cause SCN4. Patients with mutations in *G6PC3* suffer from myeloid maturation arrest leading to congenital neutropenia. They also suffer from various other congenital defects, including

cardiac and urogenital defects and facial dysmorphism, increased visibility of superficial veins, inner ear hearing loss, endocrine abnormalities, or myopathy (26).

Wiskott–Aldrich syndrome (WAS) is an X-linked (XL) disorder caused by deleterious loss of function mutations in *WAS* and its cognate protein Wiskott–Aldrich syndrome protein and is characterized by susceptibility to infections, thrombocytopenia with bleeding diathesis, and eczema (41). Rare activating mutations in *WAS* cause a constitutive activation with increase in actin polymerization (27), and instead of classic WAS, these patients present with X-linked congenital neutropenia associated with myelodysplasia, lymphoid abnormalities, and increased myeloid apoptosis (42).

In contrast to SCNs in which myeloid arrest or increased apoptosis cause neutropenia, myelokathexis, or inability of neutrophils to immigrate from the bone marrow can cause severe congenital neutropenia. Warts, hypogammaglobulinemia, infections, myelokathexis syndrome, in which the clinical manifestations include neutropenia, hypogammaglobulinemia, and mild to extensive warts is an AD immunodeficiency caused by gain of function mutations in the chemokine receptor CXCR4. Stromal cell-derived growth factor-1 (SDF1, also known as CXCL12) is found in the bone marrow stroma and is the ligand for CXCR4 found on neutrophils; both are important bone marrow retention factors for neutrophils. Myelokathexis, hyperplasia with an accumulation of apoptotic neutrophils in the bone marrow and neutropenia in the periphery, is the hallmark of this disorder (43, 44).

In addition to congenital neutropenia disorders described thus far, several disorders with neutropenia and hypopigmentation also have been described (Table 1). Neutropenia may be constant in some or intermittent in others. Lastly, neutropenia leading to susceptibility to invasive bacterial infections can be a clinical manifestation in other immunodeficiency syndromes, such as XL hyper IgM syndrome (36) and XL agammaglobulinemia (37).

Patients with SCN typically present in infancy with recurrent mouth sores, pharyngitis, otitis media, respiratory infections, skin infections, and neutropenia (ANC < 200/ μ L). Evaluation of the bone marrow may be helpful in narrowing the differential diagnosis of congenital neutropenia. In SCN syndromes, there is a characteristic normal or decreased cellularity with early myeloid arrest at the pro-myelocyte or myelocyte stages often with atypical nuclei and cytoplasmic vacuolization (45).

Treatment of SCN includes daily subcutaneous injections of recombinant granulocyte colony stimulating factor (G-CSF). Most patients with SCN respond to G-CSF; however, patients continue to be at risk for myelodysplasia, acute leukemias, and severe infections. Because of these risks and negative impact of disease on quality of life, patients with SCNs should be considered for curative therapy with HSC transplantation (20).

DISORDERS OF NEUTROPHIL CHEMOTAXIS

For efficient neutrophil killing, neutrophils must first leave the vasculature and reach a site of infection. Recruitment of

neutrophils to leave the blood stream consists of three major steps: initiation of adherence of activated endothelial cells and rolling, firm attachment of neutrophils to the endothelium, and migrating of the neutrophil across the endothelial barrier (Figure 2). The initial steps occur due to interaction between P-selectin glycoprotein ligand-1 of neutrophils and P-selectin or E-selectin of endothelial cells. Firm attachment of neutrophils to the endothelium is dependent on β 2 integrins (LFA-1 and Mac-1) present on the surface of neutrophils interacting with intracellular adhesion molecule-1 on endothelial cells. Final migration is triggered by local chemokines and bacterial products at the site of infection.

Defects in a number of these adhesion molecules results in clinical syndromes. Leukocyte adhesion deficiency (LAD)-I is an AR syndrome due to defects in CD18, the common β chain of the β 2 integrin family. The β 2 integrin is required for stable expression of three distinct β 2 integrins: CD11a/CD18 (LFA-1), CD11b/CD18 (Mac-1), and CD11c/CD18 (p150,95). Patients with LAD-I typically present with early onset of soft tissue and invasive bacterial infections, delayed separation of the umbilical cord, poor wound healing, omphalitis, periodontal disease, and neutrophilia in the serum. Diagnosis of LAD-I is confirmed by absence of CD18 and the associated alpha subunits CD11a, CD11b, and CD11c or by sequencing of the β 2 integrin. Treatment includes use of prophylactic antibiotics and hematopoietic stem cell transplant (HSCT) for those with a severe phenotype (46).

Leukocyte adhesion deficiency-II is a very rare AR syndrome that results from defects in the guanosine diphosphate fucose transporter gene (*SLC35C1*) leading to abnormal fucosylation on the neutrophil surface that results in defective rolling of leukocytes (46, 47). Fucosylated proteins such as sialyl Lewis X (CD15s) are ligands for endothelial selectins and are important for the early phases of adhesion. However, neutrophils are able to adhere and transmigrate *via* β 2 integrins, allowing for some level of neutrophil defense against bacterial infections. Clinical manifestations include susceptibility to pyogenic infections although less severe than in LAD-I. Patients also have intellectual disability, short stature, depressed nasal bridge, microcephaly, and cortical atrophy, and the rare Bombay (hh) blood phenotype with lack of A, B, and H antigens. Absence of SLeX (CD15a) shown by analysis of peripheral leukocytes is diagnostic. Treatment includes use of prophylactic antibiotics (46, 47). Trials of fucose supplementation have been beneficial in some (47).

Leukocyte adhesion deficiency-III is a rare AR syndrome caused by mutations in *Kindlin 3*, an integrin cytoplasmic tail binding adaptor that is essential for integrin activation. Patients with LAD-III have similar manifestations as those with LAD-I but with milder symptoms. Unlike LAD-I, increased bleeding tendency is the major source of morbidity. Platelet aggregation requires both β 1 and β 2 integrin activation, and because of the integrin activation defect in these patients, bleeding severity is increased (48).

Autosomal dominant Hyper IgE syndrome (AD-HIES) is a multi-system disorder characterized by elevated serum levels of IgE, recurrent cutaneous and pulmonary bacterial and fungal infections, development of pneumatoceles, chronic skin dermatitis, and many skeletal and dental abnormalities (49).

Staphylococcal infections of the skin and lung are often indolent and lack characteristic inflammatory characteristics (cold abscesses). Loss of function mutations in signal transducer activator of transcription 3 (50) lead to loss in production of Th17 cells and are causative of AD-HIES (51). Neutrophils in patients with AD-HIES have a profound defect in chemotaxis. Diagnosis is based on recognition of the constellation of symptoms along with often profound elevation in serum IgE levels. Treatment consists of antibiotic prophylaxis.

DISORDERS OF NEUTROPHIL INGESTION AND DEGRANULATION GRANULES

Following phagocytosis, phagosome membranes fuse with neutrophil granules and granular contents are released into the phagosome lumen where direct microbial killing occurs. These microbicidal products are contained within four types of secretory granules: azurophilic (primary), specific (secondary), gelatinase (tertiary), and secretory vesicles (52). Defensins, neutrophil elastase, lactoferrin, and gelatinase are released upon stimulation of the neutrophil from certain infections. Granules can be easily visualized within neutrophils *via* light and electron microscopy.

Chediak-Higashi syndrome (CHS) is an AR disorder caused by defects in *LYST* leading to defects in granule morphogenesis (Figure 1C) with delayed and incomplete degranulation (28, 53). Clinical manifestations include oculocutaneous albinism, neurologic disease, immunodeficiency, and mild bleeding tendency. Natural killer cells are present but function abnormally, as do neutrophils with abnormal chemotaxis and killing both causing an increased risk of bacterial infections. Platelets have irregular morphology; mild bleeding is a common feature of CHS. Neurologic features include cognitive impairment, peripheral neuropathy, ataxia, and parkinsonism. Giant peroxidase positive granules that coalesce azurophilic and specific granules are present within the peripheral neutrophils and are even more prominent within bone marrow-derived neutrophils of CHS patients. Pigment clumping also can be found on hair from CHS patients. About 85% of CHS patients enter the accelerated phase of disease with lymphoproliferative infiltration of the bone marrow and other reticuloendothelial system organs. Treatment consists of chemotherapy followed by HSCT for the accelerated phase (28, 53).

Neutrophil-specific granule deficiency (SGD) is a rare neutrophil defect in which neutrophils lack specific granules and, therefore, have virtually absent lactoferrin production. Clinical manifestations include susceptibility to severe invasive pyogenic infections with *Staphylococcus aureus*, *Pseudomonas aeruginosa*, and *Candida albicans* (54). Most patients present in the first few years of life with severe infection. SGD is caused by AR mutations in CCAT/enhancer binding protein epsilon (C/EBP- ϵ) (55). This defect in C/EBP- ϵ blocks the transition of neutrophil development from the promyelocyte to myelocyte stage. The pathognomonic feature of SGD is a paucity of specific granules and predominantly bilobed nuclei that can be visualized on a peripheral smear

(Figure 1B). Neutrophils from SGD patients also show abnormal chemotaxis but with normal aggregation, impaired disaggregation, and decreased bactericidal activity (55, 56). Diagnosis of SGD is made by careful examination of a peripheral smear and confirmed with molecular testing. Treatment consists primarily of use of anti-bacterial prophylaxis and possibly HSCT (56).

DISORDERS OF NEUTROPHIL KILLING

Prior to exposure to microbes, the neutrophil NADPH oxidase is inactive with its subunits residing in different cell compartments. Some are membrane bound (gp91^{phox} and p22^{phox}) and others are cytoplasmic (p47^{phox}, p67^{phox}, and p40^{phox}). After intracellular ingestion of bacteria and fungi, the components of the NADPH oxidase come together in an oxidative burst shuttling electrons across the phagosomal membrane from cytoplasmic NADPH to molecular oxygen. These reactive oxygen species then directly kill ingested microbes (57).

Mutations in all five structural genes that comprise the NADPH oxidase cause chronic granulomatous disease (CGD) (Table 2) and occurs in approximately 1:200,000 (58). The majority of patients with CGD present before age 5 with a severe or recurrent infections. The skin, lungs, lymph nodes, and liver are the most common sites of infection with a narrow spectrum of catalase-positive organisms. Infections from *Staphylococcus aureus*, *Burkholderia cepacia*, *Serratia marcescens*, *Nocardia* species, and *Aspergillus* species are the most common in North America. Formation of granulomata and a dysregulated inflammatory response to infection are a leading cause of morbidity in CGD patients. Diagnosis of CGD relies on direct measurement of superoxide production; the dihydrorhodamine (DHR) assay is the most commonly used and accepted test to diagnose CGD. The DHR assay uses flow cytometry to measure the production of hydrogen peroxide in the presence of peroxidase and directly correlates with superoxide production by the NADPH oxidase (59). Management of CGD patients relies on life long anti-bacterial and anti-fungal prophylaxis and interferon gamma. Treatment of the immune dysregulation of CGD is often accomplished by the use of corticosteroids or other immunosuppressants. Allogeneic HSCT can cure CGD, and new gene therapy protocols offer a potential cure as well (57).

Myeloperoxidase (MPO) deficiency is a common AR disorder caused by mutations in the *MPO* gene. MPO deficiency inhibits formation of hypochlorous acid from chloride and hydrogen peroxide. Despite the significant *in vitro* killing defects, there is a lack of clinical symptoms present in patients with MPO

TABLE 2 | Molecular defects of the NADPH oxidase causing CGD (57).

Gene	Protein	Inheritance pattern	Percentage
CYBA	p22 ^{phox}	AR	6%
NCF1	p47 ^{phox}	AR	20%
NCF2	p67 ^{phox}	AR	6%
NCF4	p40 ^{phox}	AR	1 individual
CYBB	gp91 ^{phox}	XL	70%

CGD, chronic granulomatous disease; AR, autosomal recessive; XL, X-linked.

deficiency. No specific treatment, including the use of prophylactic antibiotics, is recommended (60).

Glucose-6-phosphate dehydrogenase (G6PD) catalyzes the two reactions of the hexose monophosphate shunt pathway responsible for forming NADPH. Mutations in G6PD cause a gradual decay in G6PD which have little effect on the short life span of neutrophils. The majority of patients with G6PD deficiency develop red cell hemolysis triggered by oxidative stress. However, a few G6PD mutations have led to very low levels of G6PD leading to severe hemolytic anemia and NADPH oxidase deficiency that clinically resembles CGD (61).

CONCLUSION

Across species, neutrophils are critical for host defense against invasive bacteria and fungi. Evolution of neutrophils in humans has developed into an eloquent process of neutrophil ontogeny, trafficking, and killing to become a major first line defense against infection. Defects in neutrophil quantity, adherence, chemotaxis, and killing all lead to severe and potentially life-threatening disease in humans, underscoring the important role of the neutrophil

in the immune system. Dissecting the molecular pathology of disorders of neutrophil function has given us unique insight into the primary means by which the innate immune system confronts pathogen challenges. Further investigations of similarities and differences between species in how neutrophils function has considerable potential for revealing the inner workings of a complex mechanism of host defense.

AUTHOR CONTRIBUTIONS

JWL developed and wrote this review.

ACKNOWLEDGMENTS

The author wishes to thank Dr. Wil Chamizo and Dr. Aleksandra Petrovic for neutrophil figures.

FUNDING

Funding for this review was provided by the University of South Florida Morsani College of Medicine.

REFERENCES

- Hartenstein V. Blood cells and blood cell development in the animal kingdom. *Annu Rev Cell Dev Biol* (2006) 22:677–712. doi:10.1146/annurev.cellbio.22.010605.093317
- Palis J, Yoder MC. Yolk-sac hematopoiesis: the first blood cells of mouse and man. *Exp Hematol* (2001) 29(8):927–36. doi:10.1016/S0301-472X(01)00669-5
- Manroe BL, Weinberg AG, Rosenfeld CR, Browne R. The neonatal blood count in health and disease. I. Reference values for neutrophilic cells. *J Pediatr* (1979) 95(1):89–98. doi:10.1016/S0022-3476(79)80096-7
- Mouzinho A, Rosenfeld CR, Sanchez PJ, Risser R. Revised reference ranges for circulating neutrophils in very-low-birth-weight neonates. *Pediatrics* (1994) 94(1):76–82.
- Anderson DC, Abbassi O, Kishimoto TK, Koenig JM, McIntire LV, Smith CW. Diminished lectin-, epidermal growth factor-, complement binding domain-cell adhesion molecule-1 on neonatal neutrophils underlies their impaired CD18-independent adhesion to endothelial cells in vitro. *J Immunol* (1991) 146(10):3372–9.
- Harris MC, Shalit M, Southwick FS. Diminished actin polymerization by neutrophils from newborn infants. *Pediatr Res* (1993) 33(1):27–31. doi:10.1203/00006450-199301000-00006
- Levy O, Martin S, Eichenwald E, Ganz T, Valore E, Carroll SF, et al. Impaired innate immunity in the newborn: newborn neutrophils are deficient in bactericidal/permeability-increasing protein. *Pediatrics* (1999) 104(6):1327–33. doi:10.1542/peds.104.6.1327
- Borregaard N. Neutrophils, from marrow to microbes. *Immunity* (2010) 33(5):657–70. doi:10.1016/j.immuni.2010.11.011
- Keszei M, Westerberg LS. Congenital defects in neutrophil dynamics. *J Immunol Res* (2014) 2014:303782. doi:10.1155/2014/303782
- Teng TS, Ji AL, Ji XY, Li YZ. Neutrophils and immunity: from bactericidal action to being conquered. *J Immunol Res* (2017) 2017:9671604. doi:10.1155/2017/9671604
- Harvie EA, Huttenlocher A. Neutrophils in host defense: new insights from zebrafish. *J Leukoc Biol* (2015) 98(4):523–37. doi:10.1189/jlb.4MR1114-524R
- Henry KM, Loynes CA, Whyte MK, Renshaw SA. Zebrafish as a model for the study of neutrophil biology. *J Leukoc Biol* (2013) 94(4):633–42. doi:10.1189/jlb.1112594
- DeVries ME, Kelvin AA, Xu L, Ran L, Robinson J, Kelvin DJ. Defining the origins and evolution of the chemokine/chemokine receptor system. *J Immunol* (2006) 176(1):401–15. doi:10.4049/jimmunol.176.1.401
- Garg AD, Vandenberk L, Fang S, Fasche T, Van Eygen S, Maes J, et al. Pathogen response-like recruitment and activation of neutrophils by sterile immunogenic dying cells drives neutrophil-mediated residual cell killing. *Cell Death Differ* (2017) 24(5):832–43. doi:10.1038/cdd.2017.15
- Starnes TW, Huttenlocher A. Neutrophil reverse migration becomes transparent with zebrafish. *Adv Hematol* (2012) 2012:398640. doi:10.1155/2012/398640
- Powell DR, Huttenlocher A. Neutrophils in the tumor microenvironment. *Trends Immunol* (2016) 37(1):41–52. doi:10.1016/j.it.2015.11.008
- Bousfiha A, Jeddane L, Al-Herz W, Ailal F, Casanova JL, Chatila T, et al. The 2015 IUIS phenotypic classification for primary immunodeficiencies. *J Clin Immunol* (2015) 35(8):727–38. doi:10.1007/s10875-015-0198-5
- Pillay J, den Braber I, Vrsekooop N, Kwast LM, de Boer RJ, Borghans JA, et al. In vivo labeling with $^2\text{H}_2\text{O}$ reveals a human neutrophil lifespan of 5.4 days. *Blood* (2010) 116(4):625–7. doi:10.1182/blood-2010-01-259028
- Boztug K, Klein C. Genetic etiologies of severe congenital neutropenia. *Curr Opin Pediatr* (2011) 23(1):21–6. doi:10.1097/MOP.0b013e32834262f8
- Boztug K, Klein C. Genetics and pathophysiology of severe congenital neutropenia syndromes unrelated to neutrophil elastase. *Hematol Oncol Clin North Am* (2013) 27(1):43–60, vii. doi:10.1016/j.hoc.2012.11.004
- Dale DC, Person RE, Bolyard AA, Aprikan AG, Bos C, Bonilla MA, et al. Mutations in the gene encoding neutrophil elastase in congenital and cyclic neutropenia. *Blood* (2000) 96(7):2317–22.
- Horwitz MS, Duan Z, Korkmaz B, Lee HH, Mealiffe ME, Salipante SJ. Neutrophil elastase in cyclic and severe congenital neutropenia. *Blood* (2007) 109(5):1817–24. doi:10.1182/blood-2006-08-019166
- Horwitz M, Benson KE, Person RE, Aprikan AG, Dale DC. Mutations in ELA2, encoding neutrophil elastase, define a 21-day biological clock in cyclic haematopoiesis. *Nat Genet* (1999) 23(4):433–6. doi:10.1038/70544
- Hock H, Hamblen MJ, Rooke HM, Traver D, Bronson RT, Cameron S, et al. Intrinsic requirement for zinc finger transcription factor Gfi-1 in neutrophil differentiation. *Immunity* (2003) 18(1):109–20. doi:10.1016/S1074-7613(02)00501-0
- Klein C, Grudzien M, Appaswamy G, Germeshausen M, Sandrock I, Schaffer AA, et al. HAX1 deficiency causes autosomal recessive severe congenital neutropenia (Kostmann disease). *Nat Genet* (2007) 39(1):86–92. doi:10.1038/ng1940
- Boztug K, Rosenberg PS, Dorda M, Banka S, Moulton T, Curtin J, et al. Extended spectrum of human glucose-6-phosphatase catalytic subunit 3 deficiency: novel genotypes and phenotypic variability in severe congenital neutropenia. *J Pediatr* (2012) 160(4):679–83.e2. doi:10.1016/j.jpeds.2011.09.019

27. Ancliff PJ, Blundell MP, Cory GO, Calle Y, Worth A, Kempinski H, et al. Two novel activating mutations in the Wiskott-Aldrich syndrome protein result in congenital neutropenia. *Blood* (2006) 108(7):2182–9. doi:10.1182/blood-2006-01-010249
28. Kaplan J, De Domenico I, Ward DM. Chediak-Higashi syndrome. *Curr Opin Hematol* (2008) 15(1):22–9. doi:10.1097/MOH.0b013e3282f2bcce
29. Fontana S, Parolini S, Vermi W, Booth S, Gallo F, Donini M, et al. Innate immunity defects in Hermansky-Pudlak type 2 syndrome. *Blood* (2006) 107(12):4857–64. doi:10.1182/blood-2005-11-4398
30. Jung J, Bohn G, Allroth A, Boztug K, Brandes G, Sandrock I, et al. Identification of a homozygous deletion in the AP3B1 gene causing Hermansky-Pudlak syndrome, type 2. *Blood* (2006) 108(1):362–9. doi:10.1182/blood-2005-11-4377
31. Menasche G, Pastural E, Feldmann J, Certain S, Ersoy F, Dupuis S, et al. Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome. *Nat Genet* (2000) 25(2):173–6. doi:10.1038/76024
32. Lagresle-Peyrou C, Six EM, Picard C, Rieux-Laucat F, Michel V, Ditadi A, et al. Human adenylate kinase 2 deficiency causes a profound hematopoietic defect associated with sensorineural deafness. *Nat Genet* (2009) 41(1):106–11. doi:10.1038/ng.278
33. Boockvar GR, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR, et al. Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet* (2003) 33(1):97–101. doi:10.1038/ng1062
34. Clericuzio C, Harutyunyan K, Jin W, Erickson RP, Irvine AD, McLean WH, et al. Identification of a novel C16orf57 mutation in Athabaskan patients with poikiloderma with neutropenia. *Am J Med Genet A* (2011) 155A(2):337–42. doi:10.1002/ajmg.a.33807
35. Notarangelo LD, Roifman CM, Giliani S. Cartilage-hair hypoplasia: molecular basis and heterogeneity of the immunological phenotype. *Curr Opin Allergy Clin Immunol* (2008) 8(6):534–9. doi:10.1097/ACI.0b013e328310fe7d
36. Leven EA, Maffucci P, Ochs HD, Scholl PR, Buckley RH, Fuleihan RL, et al. Hyper IgM syndrome: a report from the USIDNET registry. *J Clin Immunol* (2016) 36(5):490–501. doi:10.1007/s10875-016-0291-4
37. Farrar JE, Rohrer J, Conley ME. Neutropenia in X-linked agammaglobulinemia. *Clin Immunol Immunopathol* (1996) 81(3):271–6. doi:10.1006/clin.1996.0188
38. Barth PG, Scholte HR, Berden JA, Van der Klei-Van Moorsel JM, Luyt-Houwen IE, Van't Veer-Korthof ET, et al. An X-linked mitochondrial disease affecting cardiac muscle, skeletal muscle and neutrophil leucocytes. *J Neurol Sci* (1983) 62(1–3):327–55. doi:10.1016/0022-510X(83)90209-5
39. Kolehmainen J, Black GC, Saarinen A, Chandler K, Clayton-Smith J, Traskelin AL, et al. Cohen syndrome is caused by mutations in a novel gene, COH1, encoding a transmembrane protein with a presumed role in vesicle-mediated sorting and intracellular protein transport. *Am J Hum Genet* (2003) 72(6):1359–69. doi:10.1086/375454
40. Pearson HA, Lobel JS, Kocoshis SA, Naiman JL, Windmiller J, Lammi AT, et al. A new syndrome of refractory sideroblastic anemia with vacuolization of marrow precursors and exocrine pancreatic dysfunction. *J Pediatr* (1979) 95(6):976–84. doi:10.1016/S0022-3476(79)80286-3
41. Buchbinder D, Nugent DJ, Fillipovich AH. Wiskott-Aldrich syndrome: diagnosis, current management, and emerging treatments. *Appl Clin Genet* (2014) 7:55–66. doi:10.2147/TACG.S58444
42. Beel K, Cotter MM, Blatny J, Bond J, Lucas G, Green F, et al. A large kindred with X-linked neutropenia with an I294T mutation of the Wiskott-Aldrich syndrome gene. *Br J Haematol* (2009) 144(1):120–6. doi:10.1111/j.1365-2141.2008.07416.x
43. Kawai T, Malech HL. WHIM syndrome: congenital immune deficiency disease. *Curr Opin Hematol* (2009) 16(1):20–6. doi:10.1097/MOH.0b013e32831ac557
44. Martin C, Burdon PC, Bridger G, Gutierrez-Ramos JC, Williams TJ, Rankin SM. Chemokines acting via CXCR2 and CXCR4 control the release of neutrophils from the bone marrow and their return following senescence. *Immunity* (2003) 19(4):583–93. doi:10.1016/S1074-7613(03)00263-2
45. Donadieu J, Fenneteau O, Beaupain B, Mahlaoui N, Chantelot CB. Congenital neutropenia: diagnosis, molecular bases and patient management. *Orphanet J Rare Dis* (2011) 6:26. doi:10.1186/1750-1172-6-26
46. Hanna S, Etzioni A. Leukocyte adhesion deficiencies. *Ann N Y Acad Sci* (2012) 1250:50–5. doi:10.1111/j.1749-6632.2011.06389.x
47. Etzioni A, Tonetti M. Fucose supplementation in leukocyte adhesion deficiency type II. *Blood* (2000) 95(11):3641–3.
48. Etzioni A. Leukocyte adhesion deficiency III – when integrins activation fails. *J Clin Immunol* (2014) 34(8):900–3. doi:10.1007/s10875-014-0094-4
49. Sowerwine KJ, Holland SM, Freeman AF. Hyper-IgE syndrome update. *Ann N Y Acad Sci* (2012) 1250:25–32. doi:10.1111/j.1749-6632.2011.06387.x
50. Holland SM, DeLeo FR, Elloumi HZ, Hsu AP, Uzel G, Brodsky N, et al. STAT3 mutations in the hyper-IgE syndrome. *N Engl J Med* (2007) 357(16):1608–19. doi:10.1056/NEJMoa073687
51. Milner JD, Brenchley JM, Laurence A, Freeman AF, Hill BJ, Elias KM, et al. Impaired T(H)17 cell differentiation in subjects with autosomal dominant hyper-IgE syndrome. *Nature* (2008) 452(7188):773–6. doi:10.1038/nature06764
52. Borregaard N, Cowland JB. Granules of the human neutrophilic polymorphonuclear leukocyte. *Blood* (1997) 89(10):3503–21.
53. Introne WJ, Westbrook W, Golas GA, Adams D. Chediak-Higashi syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al., editors. *GeneReviews(R)*. Seattle, WA: University of Washington, Seattle (1993).
54. McIlwaine L, Parker A, Sandilands G, Gallipoli P, Leach M. Neutrophil-specific granule deficiency. *Br J Haematol* (2013) 160(6):735. doi:10.1111/bjh.12207
55. Lekstrom-Himes JA, Dorman SE, Kopar P, Holland SM, Gallin JI. Neutrophil-specific granule deficiency results from a novel mutation with loss of function of the transcription factor CCAAT/enhancer binding protein epsilon. *J Exp Med* (1999) 189(11):1847–52. doi:10.1084/jem.189.11.1847
56. Gallin JI. Neutrophil specific granule deficiency. *Annu Rev Med* (1985) 36:263–74. doi:10.1146/annurev.me.36.020185.001403
57. Leiding JW, Holland SM. Chronic granulomatous disease. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, et al., editors. *GeneReviews(R)*. Seattle, WA: University of Washington, Seattle (1993).
58. Winkelstein JA, Marino MC, Johnston RB Jr, Boyle J, Curnutte J, Gallin JI, et al. Chronic granulomatous disease. Report on a national registry of 368 patients. *Medicine* (2000) 79(3):155–69. doi:10.1097/00005792-200005000-00003
59. Kuhns DB, Alvord WG, Heller T, Feld JJ, Pike KM, Marciano BE, et al. Residual NADPH oxidase and survival in chronic granulomatous disease. *N Engl J Med* (2010) 363(27):2600–10. doi:10.1056/NEJMoa1007097
60. Klebanoff SJ, Kettle AJ, Rosen H, Winterbourn CC, Nauseef WM. Myeloperoxidase: a front-line defender against phagocytosed microorganisms. *J Leukoc Biol* (2013) 93(2):185–98. doi:10.1189/jlb.0712349
61. Baehner RL, Nathan DG, Castle WB. Oxidant injury of caucasian glucose-6-phosphate dehydrogenase-deficient red blood cells by phagocytosing leukocytes during infection. *J Clin Invest* (1971) 50(12):2466–73. doi:10.1172/JCI106747

Conflict of Interest Statement: The 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.

Copyright © 2017 Leiding. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Alkaline Phosphatase, an Unconventional Immune Protein

Bethany A. Rader*

Department of Microbiology, Southern Illinois University, Carbondale, IL, United States

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Alain Couvineau,
Institut national de la santé et de la
recherche médicale, France
Elmar Pieterse,
Radboud University Nijmegen
Medical Center, Netherlands

*Correspondence:

Bethany A. Rader
bethany.rader@siu.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 24 May 2017

Accepted: 13 July 2017

Published: 03 August 2017

Citation:

Rader BA (2017) Alkaline
Phosphatase, an Unconventional
Immune Protein.
Front. Immunol. 8:897.
doi: 10.3389/fimmu.2017.00897

Recent years have seen an increase in the number of studies focusing on alkaline phosphatases (APs), revealing an expanding complexity of function of these enzymes. Of the four human AP (hAP) proteins, most is known about tissue non-specific AP (TNAP) and intestinal AP (IAP). This review highlights current understanding of TNAP and IAP in relation to human health and disease. TNAP plays a role in multiple processes, including bone mineralization, vitamin B6 metabolism, and neurogenesis, is the genetic cause of hypophosphatasia, influences inflammation through regulation of purinergic signaling, and has been implicated in Alzheimer's disease. IAP regulates fatty acid absorption and has been implicated in the regulation of diet-induced obesity and metabolic syndrome. IAP and TNAP can dephosphorylate bacterial-derived lipopolysaccharide, and IAP has been identified as a potential regulator of the composition of the intestinal microbiome, an evolutionarily conserved function. Endogenous and recombinant bovine APs and recombinant hAPs are currently being explored for their potential as pharmacological agents to treat AP-associated diseases and mitigate multiple sources of inflammation. Continued research on these versatile proteins will undoubtedly provide insight into human pathophysiology, biochemistry, and the human holobiont.

Keywords: alkaline phosphatase, hypophosphatasia, tissue non-specific AP, intestinal AP, lipopolysaccharide, microbiome

INTRODUCTION

Alkaline phosphatases (APs) belong to a superfamily of proteins (EC 3.1.3.1) sharing conservation of metal binding sites, amino acids required for activity, and predicted fold structure (1). APs are used extensively in life sciences education, as a tool in molecular biology research and as a blood serum marker for liver and bone health, and yet we know surprisingly little about the potential these proteins have to influence our health. In general, APs are anchored to outside surface of the plasma membrane and catalyze the hydrolysis of phosphate groups from a variety of different substrates (dephosphorylation) in an alkaline environment, freeing inorganic phosphate (Pi) (2–4). APs are ubiquitous, with members of the AP super family of proteins extending from the archaea (5) to humans (2). Their ubiquity across life and their expansion and subsequent dynamic evolution in vertebrates implies both variety and conservation of function (6, 7). There are four genes encoding APs in humans. Three genes, *ALPI*, *ALPP*, and *ALPPL2*, display tissue-specific expression (TSAP proteins), whereas the fourth, *ALPL* is tissue non-specific in expression [tissue non-specific AP (TNAP) proteins] (Table 1). Unlike tissue distribution, surprisingly less is known about the function of these proteins, especially *ALPP* and *ALPPL2* (Table 1). This mini-review will briefly highlight current knowledge of TNAP and intestinal AP (IAP) function in human health and disease (see Figure 1 for summary).

TISSUE NON-SPECIFIC AP

The most direct link between APs and human disease is hypophosphatasia (HPP), a disease characterized by mutations in TNAP associated with decreased enzyme activity in specific organs (10, 11) (**Figure 1B**). This decrease in AP activity results in variable symptoms that range from perinatal HPP that can result in still birth from profound skeletal hypomineralization

(11, 12), potentially lethal seizures in infantile HPP (13–15), to milder phenotypes such as bone fractures and periodontal disease in juvenile HPP and adult HPP (16, 17). A relatively recent mouse model for HPP, in conjunction with medical data and genetic analysis has provided insight into the mechanism of HPP pathophysiology regarding at least two TNAP substrates, extracellular pyrophosphate (PPi), and pyridoxal-5-phosphate (PLP) (7).

HYPOPHOSPHATASIA

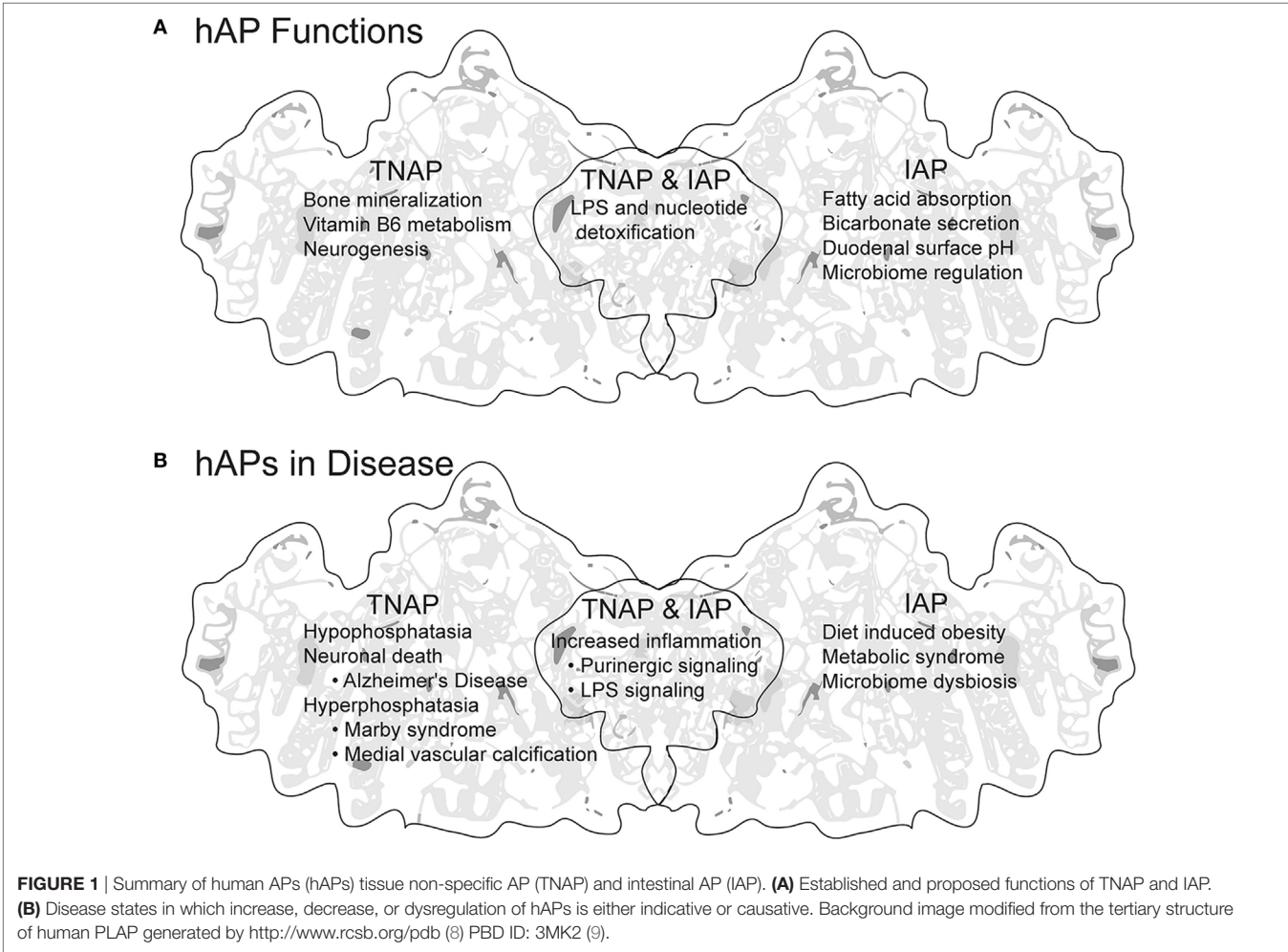
Tissue non-specific AP is anchored to the cell membranes of osteoblasts and chondrocytes and to matrix vesicles released by those cells, where it degrades PPi to Pi. PPi is an inhibitor of mineralization (18) and regulation by TNAP controls propagation of extracellular mineralization of apatite crystals. TNAP deficiency increases the amount of inhibitory PPi thus decreasing extracellular mineralization, and humans with HPP show a loss of mineralization fronts (19). This has been recapitulated in a TNAP knockout mouse model for infantile HPP (20–22). The loss of mineralization results in various symptoms including softening of bone, bowing and spontaneous breakage of bones, rickets, and tooth (dentin/cementum/enamel) defects (23).

TABLE 1 | Description of human alkaline phosphatases (APs).^a

AP gene	AP protein	Tissue distribution	Known function
ALPL	Tissue non-specific AP	Liver, kidney, skeletal tissue, nervous system	Bone and tooth deposition
ALPP	PLAP ^b	Syncytiotrophoblasts, reproductive tumors	Unknown
ALPPL2	GCAP ^b	Testis, reproductive tumors	Unknown
ALPI	IAP ^b	Intestine, enterocyte	Fatty acid absorption, lipopolysaccharide detoxification

^aInformation from Ref. (2, 7).

^bTSAPs.



Pyridoxal-5-phosphate, the active form of vitamin B6 (24), is elevated in the serum of HPP patients (25, 26). Hydrolysis of PLP to pyridoxal (PL) by TNAP facilitates diffusion of PL across cell membranes, where it is then re-phosphorylated into PLP. PLP is a versatile cofactor for an estimated 4% of enzymatic reactions and is used by over 110 enzymes to produce or metabolize various molecules (27). PLP-dependent enzymes in the brain are responsible for the production of important neurochemicals including serotonin, dopamine, and gamma-aminobutyric acid (28). The decrease in PLP and resulting decrease in PLP-dependent metabolism in the brain in perinatal HPP patients has been implicated as the cause of neonatal seizures (29, 30).

NON-HPP TNAP PATHOPHYSIOLOGY

Tissue non-specific AP has been implicated in non-HPP related medical conditions (**Figure 1B**). TNAP is expressed during embryonic neural and spinal chord development, and promotes axonal growth *in vitro* and neurogenesis in adults (31), suggesting an importance in proper neural function. Indeed, increased TNAP activity in the brain has been demonstrated in postmortem hippocampus and serum samples from Alzheimer's disease patients and has been implicated in neuronal death through increased dephosphorylation of tau (32). Increased serum levels of AP (TNAP and/or TSAPs) due to mutations in GPI anchor synthesis, termed hyperphosphatasia, results most notably in Marby syndrome characterized by seizures, intellectual disability, and facial dysmorphism (33). TNAP upregulation in the vasculature contributes to medial vascular calcification causing vascular stiffening and eventually heart failure (34, 35). An emerging function for TNAP is regulation of purinergic signaling. Extracellular ATP and ADP, through the binding of nucleotide receptors, act as signals inducing inflammation after an acute event such as necrosis induced by damage or infection that releases intracellular nucleotides. In contrast, degradation of extracellular ATP and ADP to AMP and adenine causes cessation of inflammatory signaling, and induction through adenine receptors of an anti-inflammation response (36, 37). TNAP has been implicated in protection against inflammation in multiple diseases and promotion of intestinal microbial populations through hydrolysis of extracellular ATP/ADP to AMP and adenosine (38–40).

INTESTINAL AP

Intestinal AP is expressed in villus-associated enterocytes where it regulates fatty acid absorption through secretion of vesicles at both the luminal and basolateral surfaces (41, 42), regulates bicarbonate secretion and duodenal surface pH (43), and has been implicated in the regulation of diet-induced obesity (44, 45) and metabolic syndrome (46, 47) (**Figure 1A**). But perhaps, the most remarkable function of IAP centers on its protective interactions with the bacterial symbionts that inhabit or invade our enteric system. IAP has been shown to dephosphorylate (detoxify) the lipid A moiety of lipopolysaccharide (LPS), the outer lipid layer of the outer membrane of Gram-negative

bacteria (48). In vertebrates, these phosphates are important for binding of LPS to the toll-like receptor 4/MD-2 innate immune receptor complex (49), initiation of NF- κ B signaling, and immune response induction (50–52).

Intestinal AP deficiency has been associated with inflammation in the human intestine (53) and in the intestines of vertebrate models in which AP levels are decreased (54). Supplementation of IAP to animals where intestinal inflammation is induced directly or indirectly (with antibiotic use for example) reduces inflammation (53, 55, 56). In addition, a protective role has been ascribed to IAP in mouse models of necrotizing enterocolitis (57–59). This protective role may include IAP-dependent shaping (60) and homeostasis (61) of the microbiome. Along with direct regulation of intestinal homeostasis, IAPs and LPS detoxification have been implicated in other immune-related processes including prevention of bacterial translocation by endogenous or pharmacologically administered IAPs (62–64), and resolution of intestinal inflammation and tissue regeneration (65–67). It should also be noted that in addition to vertebrate IAP, TNAP has been shown to dephosphorylate LPS when it is applied to tissue sections from rat livers (68) and in the mouse uterus (69). With the current and increasing interest in the microbiome, IAP function as it relates to interaction with the endogenous microbes and its influence on human health will undoubtedly be clarified in the coming years.

CLINICAL USE OF APs

Although there are a multitude of AP studies focusing on vertebrate models of disease, there are relatively few publications to date reporting pharmacological use of APs as a treatment in humans. At the time this article was written, a search of <http://clinicaltrials.gov> using AP as a search term produced over several hundred responses, however, the vast majority assay for AP levels in serum (a constant hazard when searching any science or medical database using “alkaline phosphatase” as a search term). However, there were at least 11 clinical trials concerning AP treatment of HPP, 3 concerning AP treatment of sepsis with renal injury or failure, 2 concerning AP treatment during or after cardiac surgery, and at least 1 each concerning AP treatment of rheumatoid arthritis, and ulcerative colitis (UC). Interestingly, these studies use several AP sources such as isolated bovine IAP (bIAP), recombinant bIAP, and recombinant human APs (hAPs). AP enzyme replacement therapy is also currently available to treat HPP. A recombinant soluble human TNAP has been approved for use in perinatal, infantile, and juvenile-onset HPP (70, 71) and has proven successful in symptom improvement and survival in perinatal and infantile HPP (72, 73). In addition to HPP, use of AP as treatment increased renal function in sepsis-induced acute kidney injury (74, 75) and showed short-term improvement of severity of UC in patients with moderate-to-severe UC (76). These studies are a first glimpse into AP use as a treatment for disease, with very positive results. Given the jack of all trades nature of APs and the potential for APs as pharmacological agents in various diseases, studies like these should increase in the coming years.

PERSPECTIVE

The ability of APs to detoxify LPS appears to be an evolutionarily conserved function as it was recently implicated in symbiont recognition and homeostasis in the invertebrate squid-*Vibrio* symbiosis model (77). As it is becoming clear that metazoans developed in a microbial world (78), it seems likely that APs have been and may continue to be an evolutionary force shaping the diversity and function of our endogenous microbial populations. Indeed, alterations in IAP have been shown to influence the composition of the intestinal microbiome (60). We can even expand this thinking—if hAPs evolved from an ancient ancestral bacterial AP, then APs may have had a prominent role in shaping basic human biochemistry in addition to our interactions with microbes, and thus exerted a profound influence on human health.

The reader of this review will notice that many of the articles cited might be considered old, with contributions from the 1960s,

1970s, and 1980s. In fact, the study of APs goes back close to 100 years when a bone enzyme freeing phosphate was first mentioned by Robison and Soames (79). That begs the question: how is it, after 90+ years, we still know relatively little about the overall functions of APs? The recent resurgence of interest in APs, should it continue, will hopefully provide more insight into all aspects of AP biology, especially as it relates to health. The ubiquity and functions of AP distinguish them as unconventional immune proteins, and to this writer, APs are unendingly fascinating.

AUTHOR CONTRIBUTIONS

BR solely contributed to the production of this manuscript.

ACKNOWLEDGMENT

This work was supported by NIH grant 1R15GM119100 to BR.

REFERENCES

- Galperin MY, Koonin EV, Bairoch A. A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. *Protein Sci* (1998) 7:1829–35. doi:10.1002/pro.5560070819
- Millán JL. *Mammalian Alkaline Phosphatases: From Biology to Applications in Medicine and Biotechnology*. Weinheim: Wiley-VCH (2006).
- Lallès JP. Intestinal alkaline phosphatase: multiple biological roles in maintenance of intestinal homeostasis and modulation by diet. *Nutr Rev* (2010) 68:323–32. doi:10.1111/j.1753-4887.2010.00292.x
- Lallès J-P. Intestinal alkaline phosphatase: novel functions and protective effects. *Nutr Rev* (2014) 72:82–94. doi:10.1111/nure.12082
- Zimmerman AE, Martiny AC, Allison SD. Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J* (2013) 7:1187–99. doi:10.1038/ismej.2012.176
- Yang Y, Wandler AM, Postlethwait JH, Guillemin K. Dynamic evolution of the LPS-detoxifying enzyme intestinal alkaline phosphatase in zebrafish and other vertebrates. *Front Immunol* (2012) 3:314. doi:10.3389/fimmu.2012.00314
- Buchet R, Millán JL, Magne D. Multisystemic functions of alkaline phosphatases. *Methods Mol Biol* (2013) 1053:27–51. doi:10.1007/978-1-62703-562-0_3
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein databank. *Nucleic Acids Res* (2000) 28:235–42. doi:10.1002/0470020571.ch10
- Stec B, Cheltsov A, Millán JL. Refined structures of placental alkaline phosphatase show a consistent pattern of interactions at the peripheral site. *Acta Crystallogr Sect F Struct Biol Cryst Commun* (2010) 66:866–70. doi:10.1107/S1744309110019767
- Linglart A, Bioso-Duplan M. Hypophosphatasia. *Curr Osteoporos Rep* (2016) 14:95–105. doi:10.1007/s11914-016-0309-0
- Mornet E. Hypophosphatasia: the mutations in the tissue-nonspecific alkaline phosphatase gene. *Hum Mutat* (2000) 15:309–15. doi:10.1002/(SICI)1098-1004(200004)15:4<309:AID-HUMU2>3.0.CO;2-C
- Olech EM, Zemojtel T, Sowińska-Seidler A, Robinson PN, Mundlos S, Karczewski M, et al. Identification of a molecular defect in a stillborn fetus with perinatal lethal hypophosphatasia using a disease-associated genome sequencing approach. *Pol J Pathol* (2016) 67:78–83. doi:10.5114/pjp.2016.59480
- Baumgartner-Sigl S, Haberlandt E, Mumm S, Scholl-Bürgi S, Sergi C, Ryan L, et al. Pyridoxine-responsive seizures as the first symptom of infantile hypophosphatasia caused by two novel missense mutations (c.677T > C, p.M226T; c.1112C > T, p.T371I) of the tissue-nonspecific alkaline phosphatase gene. *Bone* (2007) 40:1655–61. doi:10.1016/j.bone.2007.01.020
- Nunes ML, Mugnol F, Bica I, Fiori RM. Pyridoxine-dependent seizures associated with hypophosphatasia in a newborn. *J Child Neurol* (2002) 17:222–4. doi:10.1177/088307380201700314
- De Roo MG, Abeling NG, Majoie CB, Bosch AM, Koelman JH, Cobben JM, et al. Infantile hypophosphatasia without bone deformities presenting with severe pyridoxine-resistant seizures. *Mol Genet Metab* (2014) 111:404–7. doi:10.1016/j.ymgme.2013.09.014
- Moulin P, Vaysse F, Bieth E, Mornet E, Gennero I, Dalicieux-Laurencin S, et al. Hypophosphatasia may lead to bone fragility: don't miss it. *Eur J Pediatr* (2009) 168:783–8. doi:10.1007/s00431-008-0835-6
- Weber TJ, Sawyer EK, Moseley S, Odrjlin T, Kishnani PS. Burden of disease in adult patients with hypophosphatasia: results from patient-reported outcome surveys. *Metabolism* (2014) 65:1522–30. doi:10.1016/j.metabol.2016.07.006
- Fleisch H, Bisaz S. Mechanism of calcification: inhibitory role of pyrophosphate. *Nature* (1962) 195:911. doi:10.1038/195911a0
- Anderson HC, Hsu HH, Morris DC, Fedde KN, Whyte MP. Matrix vesicles in osteomalacic hypophosphatasia bone contain apatite-like mineral crystals. *Am J Pathol* (1997) 151:1555–61.
- Goldberg RE, Austen WG, Zhang X, Munene G, Mostafa G, Biswas S, et al. Intestinal alkaline phosphatase is a gut mucosal defense factor maintained by enteral nutrition. *Proc Natl Acad Sci U S A* (2008) 105:3551–6. doi:10.1073/pnas.0712140105
- Fedde KN, Blair L, Silverstein J, Coburn SP, Ryan LM, Weinstein RS, et al. Alkaline phosphatase knock-out mice recapitulate the metabolic and skeletal defects of infantile hypophosphatasia. *J Bone Miner Res* (1999) 14:2015–26. doi:10.1359/jbmr.1999.14.12.2015
- Millán JL, Yadav M, Simao A, Narisawa S, Huesa C, McKee MD, et al. Loss of bone mineralization by the simultaneous ablation of PHOSPHO1 and alkaline phosphatase function. *Bone* (2010) 46:S78. doi:10.1016/j.bone.2010.01.191
- Millán JL. The role of phosphatases in the initiation of skeletal mineralization. *Calcif Tissue Int* (2013) 93:299–306. doi:10.1007/s00223-012-9672-8
- Shideler CE. Vitamin B6: an overview. *Am J Med Technol* (1983) 49:17–22.
- Whyte MP, Mahuren JD, Vrabel LA, Coburn SP. Markedly increased circulating pyridoxal-5'-phosphate levels in hypophosphatasia. Alkaline phosphatase acts in vitamin B6 metabolism. *J Clin Invest* (1985) 76:752–6. doi:10.1172/JCI112031
- Whyte MP, Mahuren JD, Fedde KN, Cole FS, McCabe ERB, Coburn SP. Perinatal hypophosphatasia: tissue levels of vitamin B6 are unremarkable despite markedly increased circulating concentrations of pyridoxal-5'-phosphate. Evidence for an ectoenzyme role for tissue-nonspecific alkaline phosphatase. *J Clin Invest* (1988) 81:1234–9. doi:10.1172/JCI113440
- Percudani R, Peracchi A. The B6 database: a tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families. *BMC Bioinformatics* (2009) 10:273. doi:10.1186/1471-2105-10-273
- Cruz T, Gleizes M, Balayssac S, Mornet E, Marsal G, Millán JL, et al. Identification of altered brain metabolites associated with TNAP activity in a mouse model of hypophosphatasia using untargeted NMR-based metabolomics analysis. *J Neurochem* (2017) 140:919–40. doi:10.1111/jnc.13950

29. Sebastian-Serrano A, Engel T, de Diego-Garcia L, Olivos-Ore LA, Arribas-Blazquez M, Martinez-Frailes C, et al. Neurodevelopmental alterations and seizures developed by mouse model of infantile hypophosphatasia are associated with purinergic signalling deregulation. *Hum Mol Genet* (2016) 1–14. doi:10.1093/hmg/ddw248
30. Balasubramaniam S, Bowling F, Carpenter K, Earl J, Chaitow J, Pitt J, et al. Perinatal hypophosphatasia presenting as neonatal epileptic encephalopathy with abnormal neurotransmitter metabolism secondary to reduced co-factor pyridoxal-5'-phosphate availability. *J Inherit Metab Dis* (2010) 33:S25–33. doi:10.1007/s10545-009-9012-y
31. Zimmermann H, Langer D. Tissue-nonspecific alkaline phosphatase in the developing brain and in adult neurogenesis. *Subcell Biochem* (2015) 76:61–84. doi:10.1007/978-94-017-7197-9_4
32. Kellett KAB, Hooper NM. The role of tissue non-specific alkaline phosphatase (TNAP) in neurodegenerative diseases: Alzheimer's disease in the focus. *Subcell Biochem* (2015) 76:363–74. doi:10.1007/978-94-017-7197-9_17
33. Cole DEC, Thompson MD. Neurogenetic aspects of hyperphosphatasia in Mabry syndrome. *Subcell Biochem* (2015) 76:343–61. doi:10.1007/978-94-017-7197-9_16
34. Sheen CR, Kuss P, Narisawa S, Yadav MC, Nigro J, Wang W, et al. Pathophysiological role of vascular smooth muscle alkaline phosphatase in medial artery calcification. *J Bone Miner Res* (2015) 30:824–36. doi:10.1002/jbmr.2420
35. Savinov AY, Salehi M, Yadav MC, Radichev I, Millán JL, Savinova OV. Transgenic overexpression of tissue-nonspecific alkaline phosphatase (TNAP) in vascular endothelium results in generalized arterial calcification. *J Am Heart Assoc* (2015) 4. doi:10.1161/JAHA.115.002499
36. Idzko M, Ferrari D, Eltzschig HK. Nucleotide signalling during inflammation. *Nature* (2014) 509:310–7. doi:10.1038/nature13085
37. Cauwels A, Rogge E, Vandendriessche B, Shiva S, Brouckaert P. Extracellular ATP drives systemic inflammation, tissue damage and mortality. *Cell Death Dis* (2014) 5:e1102. doi:10.1038/cddis.2014.70
38. Malo MS, Moaven O, Muhammad N, Biswas B, Alam SN, Economopoulos KP, et al. Intestinal alkaline phosphatase promotes gut bacterial growth by reducing the concentration of luminal nucleotide triphosphates. *Am J Physiol Gastrointest Liver Physiol* (2014) 306:G826–38. doi:10.1152/ajpgi.00357.2013
39. Peters E, Geraci S, Heemskerck S, Wilmer MJ, Bilos A, Kraenzlin B, et al. Alkaline phosphatase protects against renal inflammation through dephosphorylation of lipopolysaccharide and adenosine triphosphate. *Br J Pharmacol* (2015) 172:4932–45. doi:10.1111/bph.13261
40. Davidson JA, Urban T, Tong S, Twite M, Woodruff A, Wischmeyer PE, et al. Alkaline phosphatase, soluble extracellular adenosine nucleotides, and adenosine production after infant cardiopulmonary bypass. *PLoS One* (2016) 11:e0158981. doi:10.1371/journal.pone.0158981
41. Mahmood A, Engle MJ, Alpers DH. Secreted intestinal surfactant-like particles interact with cell membranes and extracellular matrix proteins in rats. *J Physiol* (2002) 542:237–44. doi:10.1113/jphysiol.2002.017087
42. McConnell RE, Higginbotham JN, Shifrin DA, Tabb DL, Coffey RJ, Tyska MJ. The enterocyte microvillus is a vesicle-generating organelle. *J Cell Biol* (2009) 185:1285–98. doi:10.1083/jcb.200902147
43. Akiba Y, Mizumori M, Guth PH, Engel E, Kaunitz JD. Duodenal brush border intestinal alkaline phosphatase activity affects bicarbonate secretion in rats. *Am J Physiol Gastrointest Liver Physiol* (2007) 293:G1223–33. doi:10.1152/ajpgi.00313.2007
44. Šefčíková Z, Hájek T, Lenhardt L, Racek L, Mozes S. Different functional responsibility of the small intestine to high-fat/high-energy diet determined the expression of obesity-prone and obesity-resistant phenotypes in rats. *Physiol Res* (2008) 57:467–74.
45. Barbier de La Serre C, Ellis CL, Lee J, Hartman AL, Rutledge JC, Raybould HE. Propensity to high-fat diet-induced obesity in rats is associated with changes in the gut microbiota and gut inflammation. *Am J Physiol Gastrointest Liver Physiol* (2010) 299(2):G440–8. doi:10.1152/ajpgi.00098.2010
46. Malo MS. A high level of intestinal alkaline phosphatase is protective against type 2 diabetes mellitus irrespective of obesity. *EBioMedicine* (2015) 2:2016–23. doi:10.1016/j.ebiom.2015.11.027
47. Kaliannan K, Hamarneh SR, Economopoulos KP, Nasrin Alam S, Moaven O, Patel P, et al. Intestinal alkaline phosphatase prevents metabolic syndrome in mice. *Proc Natl Acad Sci U S A* (2013) 110:7003–8. doi:10.1073/pnas.1220180110
48. Bentala H, Verweij WR, Huizinga-Van der Vlag A, van Loenen-Weemaes AM, Meijer DKE, Poelstra K. Removal of phosphate from lipid A as a strategy to detoxify lipopolysaccharide. *Shock* (2002) 18:561–6. doi:10.1097/01.shk.0000043623.17707.47
49. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* (2009) 458:1191–5. doi:10.1038/nature07830
50. Hoshino K, Takeuchi O, Kawai T, Sanjo H, Ogawa T, Takeda Y, et al. Cutting edge: toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J Immunol* (1999) 162:3749–52. doi:10.1038/nri2275
51. Akira S, Yamamoto M. Lipid A receptor TLR4-mediated signaling pathways. *Adv Exp Med Biol* (2009) 667:59–68. doi:10.1007/978-1-4419-1603-7_6
52. Beutler B, Rietschel ET. Innate immune sensing and its roots: the story of endotoxin. *Nat Rev Immunol* (2003) 3:169–76. doi:10.1038/nri1004
53. Tuin A, Poelstra K, de Jager-Krieken A, Bok L, Raaben W, Velders MP, et al. Role of alkaline phosphatase in colitis in man and rats. *Gut* (2009) 58:379–87. doi:10.1136/gut.2007.128868
54. Bates JM, Akerlund J, Mittge E, Guillemin K. Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota. *Cell Host Microbe* (2007) 2:371–82. doi:10.1016/j.chom.2007.10.010
55. Ramasamy S, Nguyen DD, Eston MA, Nasrin Alam S, Moss AK, Ebrahimi F, et al. Intestinal alkaline phosphatase has beneficial effects in mouse models of chronic colitis. *Inflamm Bowel Dis* (2011) 17:532–42. doi:10.1002/ibd.21377
56. Alam SN, Yammine H, Moaven O, Ahmed R, Moss AK, Biswas B, et al. Intestinal alkaline phosphatase prevents antibiotic-induced susceptibility to enteric pathogens. *Ann Surg* (2014) 259:715–22. doi:10.1097/SLA.0b013e31828fae14
57. Heinzerling NP, Liedel JL, Welak SR, Fredrich K, Biesterveld BE, Pritchard KA, et al. Intestinal alkaline phosphatase is protective to the preterm rat pup intestine. *J Pediatr Surg* (2014) 49:954–60; discussion 960. doi:10.1016/j.jpedsurg.2014.01.031
58. Rentea RM, Liedel JL, Welak SR, Cassidy LD, Mayer AN, Pritchard KA, et al. Intestinal alkaline phosphatase administration in newborns is protective of gut barrier function in a neonatal necrotizing enterocolitis rat model. *J Pediatr Surg* (2012) 47(6):1135–41. doi:10.1016/j.jpedsurg.2012.03.018
59. Biesterveld BE, Koehler SM, Heinzerling NP, Rentea RM, Fredrich K, Welak SR, et al. Intestinal alkaline phosphatase to treat necrotizing enterocolitis. *J Surg Res* (2015) 196:235–40. doi:10.1016/j.jss.2015.02.030
60. Fawley J, Koehler S, Cabrera S, Lam V, Fredrich K, Hessner M, et al. Intestinal alkaline phosphatase deficiency leads to dysbiosis and bacterial translocation in the newborn intestine. *J Surg Res* (2017) 218:35–42. doi:10.1016/j.jss.2017.03.049
61. Malo MS, Alam SN, Mostafa G, Zeller SJ, Johnson PV, Mohammad N, et al. Intestinal alkaline phosphatase preserves the normal homeostasis of gut microbiota. *Gut* (2010) 59:1476–84. doi:10.1136/gut.2010.211706
62. Shifrin DA, McConnell RE, Nambiar R, Higginbotham JN, Coffey RJ, Tyska MJ. Enterocyte microvillus-derived vesicles detoxify bacterial products and regulate epithelial-microbial interactions. *Curr Biol* (2012) 22:627–31. doi:10.1016/j.cub.2012.02.022
63. Martínez-Moya P, Ortega-González M, González R, Anzola A, Ocón B, Hernández-Chirlaque C, et al. Exogenous alkaline phosphatase treatment complements endogenous enzyme protection in colonic inflammation and reduces bacterial translocation in rats. *Pharmacol Res* (2012) 66:144–53. doi:10.1016/j.phrs.2012.04.006
64. Wang W, Chen SW, Zhu J, Zuo S, Ma YY, Chen ZY, et al. Intestinal alkaline phosphatase inhibits the translocation of bacteria of gut-origin in mice with peritonitis: mechanism of action. *PLoS One* (2015) 10:e0124835. doi:10.1371/journal.pone.0124835
65. Yagi M, Sakamoto K, Inoue T, Fukushima W, Hashimoto T, Shimizu K, et al. Effect of glutamine-enriched, elemental diet on regeneration of residual small bowel mucosa and hepatic steatosis following massive bowel resection. *J Clin Biochem Nutr* (1993) 15:219–25. doi:10.3164/jcbs.15.219

66. Campbell EL, Macmanus CF, Kominsky DJ, Keely S, Glover LE, Bowers BE, et al. Resolvin E1-induced intestinal alkaline phosphatase promotes resolution of inflammation through LPS detoxification. *Proc Natl Acad Sci U S A* (2010) 107:14303. doi:10.1073/pnas.0914730107
67. Liu W, Hu D, Huo H, Zhang W, Adiliaghdam F, Morrison S, et al. Intestinal alkaline phosphatase regulates tight junction protein levels. *J Am Coll Surg* (2016) 222:1009–17. doi:10.1016/j.jamcollsurg.2015.12.006
68. Poelstra K, Bakker WW, Klok PA, Kamps JA, Hardonk MJ, Meijer DK. Dephosphorylation of endotoxin by alkaline phosphatase in vivo. *Am J Pathol* (1997) 151:1163–9.
69. Lei W, Ni H, Herington J, Reese J, Paria BC. Alkaline phosphatase protects lipopolysaccharide-induced early pregnancy defects in mice. *PLoS One* (2015) 10:e0123243. doi:10.1371/journal.pone.0123243
70. Bloch-Zupan A. Hypophosphatasia: diagnosis and clinical signs – a dental surgeon perspective. *Int J Paediatr Dent* (2016) 26(6):426–38. doi:10.1111/ipd.12232
71. Orimo H. Pathophysiology of hypophosphatasia and the potential role of asfotase alfa. *Ther Clin Risk Manag* (2016) 12:777–86. doi:10.2147/TCRM.S87956
72. Whyte MP, Rockman-Greenberg C, Ozono K, Riese R, Moseley S, Melian A, et al. Asfotase alfa treatment improves survival for perinatal and infantile hypophosphatasia. *J Clin Endocrinol Metab* (2016) 101:334–42. doi:10.1210/jc.2015-3462
73. Hofmann C, Seefried L, Jakob F. Asfotase alfa: enzyme replacement for the treatment of bone disease in hypophosphatasia. *Drugs Today (Barc)* (2016) 52:271–85. doi:10.1358/dot.2016.52.5.2482878
74. Peters E, van Elsas A, Heemskerk S, Jonk L, van der Hoeven J, Arend J, et al. Alkaline phosphatase as a treatment of sepsis-associated acute kidney injury. *J Pharmacol Exp Ther* (2013) 344:2–7. doi:10.1124/jpet.112.198226
75. Heemskerk S, Masereeuw R, Moesker O, Bouw MP, van der Hoeven JG, Peters WH, et al. Alkaline phosphatase treatment improves renal function in severe sepsis or septic shock patients. *Crit Care Med* (2009) 37:417–423, e1. doi:10.1097/CCM.0b013e31819598af
76. Lukas M, Drastich P, Konecny M, Gionchetti P, Urban O, Cantoni F, et al. Exogenous alkaline phosphatase for the treatment of patients with moderate to severe ulcerative colitis. *Inflamm Bowel Dis* (2010) 16:1180–6. doi:10.1002/ibd.21161
77. Rader BA, Kremer N, Apicella MA, Goldman WE, McFall-Ngai MJ. Modulation of symbiont lipid signaling by host alkaline phosphatases in the squid-vibrio symbiosis. *MBio* (2012) 3. doi:10.1128/mBio.00093-12
78. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A* (2013) 110:3229–36. doi:10.1073/pnas.1218525110
79. Robison R, Soames KM. The possible significance of hexosephosphoric esters in ossification. *Biochem J* (1924) 18:740–54. doi:10.1097/00003086-199106000-00001

Conflict of Interest Statement: The 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.

Copyright © 2017 Rader. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Patterns of Early-Life Gut Microbial Colonization during Human Immune Development: An Ecological Perspective

Isabelle Laforest-Lapointe^{1,2} and Marie-Claire Arrieta^{1,2*}

¹ Department of Physiology and Pharmacology, University of Calgary, Calgary, AB, Canada, ² Department of Pediatrics, University of Calgary, Calgary, AB, Canada

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, United States

Reviewed by:

Leticia A. Carneiro,
Federal University of Rio
de Janeiro, Brazil
Shai Bel,
University of Texas
Southwestern Medical
Center, United States

*Correspondence:

Marie-Claire Arrieta
marie.arrieta@ucalgary.ca

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 30 April 2017

Accepted: 22 June 2017

Published: 10 July 2017

Citation:

Laforest-Lapointe I and Arrieta M-C
(2017) Patterns of Early-Life
Gut Microbial Colonization during
Human Immune Development:
An Ecological Perspective.
Front. Immunol. 8:788.
doi: 10.3389/fimmu.2017.00788

Alterations in gut microbial colonization during early life have been reported in infants that later developed asthma, allergies, type 1 diabetes, as well as in inflammatory bowel disease patients, previous to disease flares. Mechanistic studies in animal models have established that microbial alterations influence disease pathogenesis *via* changes in immune system maturation. Strong evidence points to the presence of a window of opportunity in early life, during which changes in gut microbial colonization can result in immune dysregulation that predisposes susceptible hosts to disease. Although the ecological patterns of microbial succession in the first year of life have been partly defined in specific human cohorts, the taxonomic and functional features, and diversity thresholds that characterize these microbial alterations are, for the most part, unknown. In this review, we summarize the most important links between the temporal mosaics of gut microbial colonization and the age-dependent immune functions that rely on them. We also highlight the importance of applying ecology theory to design studies that explore the interactions between this complex ecosystem and the host immune system. Focusing research efforts on understanding the importance of temporally structured patterns of diversity, keystone groups, and inter-kingdom microbial interactions for ecosystem functions has great potential to enable the development of biologically sound interventions aimed at maintaining and/or improving immune system development and preventing disease.

Keywords: microbiome, early-life events, immune development, microbial ecology, diversity, keystone taxa

INTRODUCTION

Recent advances in immune-mediated disease research have provided a considerable body of proof revealing the importance of the early gut microbiome for neonatal immune system development and disease pathogenesis [see Ref. (1) for a review]. The drastic increase of allergies and other immune-mediated diseases in industrialized countries has been hypothesized to be a result of deficiencies in the exposure to microbial organisms and their products, resulting in impaired immune system development, a concept first introduced as the hygiene hypothesis (2, 3). Pioneer work has identified the first 6 months after birth as a “window of opportunity” (4–7) during which contact with specific microbe-associated molecular patterns (MAMPs) triggers a cascade of reactions crucial for infant gut maturation (8–10). Disrupting early gut community succession may lead to dysbiosis, a state of ecological imbalance ensuing when the community loses key taxa, diversity, and/or metabolic capacity. This state can lead to a reduction of colonization resistance, allowing for a subsequent bloom in

opportunistic pathogens [(11); for a definition of relevant ecological concepts refer to **Table 1**]. Concomitantly, microbial dysbiosis during infancy may also lead to health-related consequences in the neonatal stage or later in life. Preterm neonates can develop necrotizing colitis (NEC), a life-threatening disease strongly associated with microbial dysbiosis (12). Infants may also experience an elevated risk of developing inflammatory diseases such as asthma and allergies (13, 14), type 1 diabetes (15, 16), celiac disease (17), inflammatory bowel disease (18, 19), and obesity (20, 21) when exposed to a microbial dysbiosis early in life. Thus, studying the patterns of microbiome assembly and how disturbances to this process reflect in the developing immune system is of utmost importance to understand the origin of human diseases responsible for enormous health and economic burden to societies.

The infant gut microbiome is a complex ecosystem involving a great number and diversity of members (e.g. bacteria, phages, fungi, viruses, protozoans) that interact in a spatially and temporally structured environment (26–28). The neonatal

gut microbiota can be considered a complex adaptive system in which both low-level local interactions and selection mechanisms combine to create high-level patterns (22). Complex adaptive systems are non-linear (output not proportional to the input, thus impeding predictability) in that they are heavily influenced by stochastic temporal events that result in a plethora of variable outcomes (22). The infant gut microbiome supports a set of emergent properties contributing to host physiology, including nervous, metabolic, and immune development (29–31), as well as tissue differentiation (32, 33). The emergent properties of a complex adaptive system are considered to be supported by combinations of taxonomic and/or functional diversity, as well as key taxonomic and/or functional groups, both of which insure community resilience (22), and increase the difficulty of attributing a cause–effect relationship to unique features or groups. Therefore, including community ecology theory to study the temporal dynamics of the infant gut microbiome has the potential to provide key information about its influence on the host immune system maturation.

Until 2 years of age, the human infant microbiome remains highly heterogeneous and lacks stability (34), being influenced by temporally structured environmental factors such as (1) maternal factors (35–37), (2) birth (38–41), (3) neonatal nutrition (27, 42, 43), and (4) other non-temporally structured factors, such as antibiotic treatments (41, 44, 45). The initial intestinal bacterial community composition of vaginally born infants involves higher levels of a multitude of bacterial groups (e.g., *Atopobium*, *Bacteroides*, *Clostridium*, *Escherichia coli*, *Streptococcus* spp. and *Prevotella*), while the community of infants born by C-section is dominated by skin-related taxa including *Staphylococcus* spp. (38). Key bacterial groups are also transferred to the infant by breastfeeding: *Bifidobacterium* and *Lactobacillus* (46–48). The multiple studies that have shown how intestinal dysbiosis can lead to detrimental immune-mediated outcomes (e.g., asthma, allergies, NEC, etc.) [see Ref. (30) for a review] suggest that the human immune system relies on an evolutionary conserved temporally structured succession of microbiome assembly. Unraveling the links between the temporal mosaics of the gut microbiome (structured succession patterns) with the emergent properties of this ecosystem (e.g., taxonomic and functional diversity, resilience, etc.) is key to improve our understanding of the importance of the infant microbiome for the development of the immune system.

The successful identification of the mechanisms linking the infant gut microbiome and immune development depends on our capacity to disentangle the relative effects of multiple factors (host genetics, environmental factors), key actors (e.g., *Bacteroidetes*, *Bifidobacterium*, etc.) and their interactions. Resilience, the ability of a system to adjust its activity to retain its basic functionality after a disturbance, is a crucial property of complex adaptive systems (49) and could be a key characteristic protecting the infant gut microbiome from reaching a dysbiotic state. Here, we review the recent findings on the links between infant gut microbiota and immune system maturation. Our review highlights the reliance of the neonate immune system development on a complex set of host-specific, environmental, temporal, and self-organizing characteristics of the infant gut

TABLE 1 | Definition of selected ecological concepts.

Concept	Definition
Complex adaptive system	A system composed of a multitude of autonomous and interdependent actors that share a variety of interactions, and behave as a unified whole in reacting and adjusting to changes in the environment (22)
Emergent property	A system's property that its components lack individually
Stochasticity	The unpredictable fluctuation of environmental conditions
Temporally structured ecosystem	An ecosystem in which emergent properties (e.g., taxonomic and functional diversity, resilience) rely on a conserved succession of events ordered in time
Richness	The number of "species" in a community
Alpha-diversity	The number of "species" and their abundance within a community or the mean in a collection of communities (i.e., Shannon index)
Beta-diversity	The absolute turnover in community composition often measured as communities' pair-wise dissimilarity in microbial ecology, also defined as the ratio between regional and local species diversity
Taxonomic diversity	The number and the relative abundance of species or taxa in a community
Functional diversity	The variety of processes or functions in a community that are important to its structure and dynamic stability
Resilience	A system's or community's capacity to promptly return to its initial state after a perturbation
Resistance	A system's or community's capacity to resist or impede changes in its state while withholding a perturbation
Selection	A key evolutive process in which genetic and environmental pressures determine which organisms succeed at survival and reproduction
Keystone species	An exceptionally important species whose presence is crucial in maintaining the organization and diversity of the ecological community (23, 24)
Succession	A pattern of changes in specific composition of a community after a radical disturbance or after the opening of a new patch in the physical environment for colonization (25)

microbiome. We propose that future studies should consider multi-level dynamics of the infant gut microbial community by disentangling the ecosystem reliance on (1) temporally structured patterns of alpha- and beta-diversity, both taxonomic and functional; (2) keystone species or microbial groups; and (3) inter-kingdom interactions. This will require a conceptual framework based on the understanding that the infant gut harbors a complex and diverse set of microbial species interacting in a temporally structured, multi-level, and non-linear network. Rightfully recognizing these structural characteristics has the potential to enable the identification of disturbance thresholds threatening the healthy development of the infant gut microbiome and its role in immune system training.

AGE-DEPENDENT IMMUNE SYSTEM DEVELOPMENT

Multiple studies and comprehensive reviews discuss how the maturation of the immune system relies on the exposure to MAMPs (50–52). Here, we discuss the recent findings demonstrating that the efficiency of microbial exposure in immune system training can be age dependent, suggesting the importance of microbial composition and infant gut microbiome temporal succession patterns.

The gastrointestinal tract is already anatomically and functionally developed at birth in full-term infants, yet important aspects of its maturation occur postnatally and depend on exogenous stimulations with microbial cells, metabolites, hormones, growth factors, and antigens (53, 54). Recent studies in murine models have revealed that several aspects of immune development are more permissive to microbial-mediated changes during early life, and that certain microbial taxa are crucial in these interactions. For instance, oral administration of *Bifidobacterium breve* was effective in inducing proliferation of FoxP3-positive regulatory T cells (FoxP3⁺ Tregs) only if administered during the pre-weaning stage in mice (55). This age-dependent promotion of an important tolerogenic immune cell was also shown to be species specific, thereby suggesting that the tolerogenic gut immune response may have adapted to respond to specific—and important—bacterial taxa. *Bifidobacterium* species and subspecies are dominant members of the infant gut microbiome (56) and are strong modulators of the immune response (57). Their role as keystone taxa of the infant gut is proposed later in this review. Another microbial species that cause an age-dependent immune effect is the *Helicobacter pylori*, which ameliorated airway hyperresponsiveness more effectively when administered before weaning in two relevant mouse models (58), although it remains unclear if and when this bacterium colonizes the infant gastrointestinal tract.

While age-dependent modulation of the host's immune response can be attributed to specific microbial taxa, most studies point to global changes in the microbial community (diversity shifts, and metabolites of poly-microbial origin) as drivers of immune development. Cahenzli et al. (59) showed that regulation of IgE responses and amelioration of antigen-induced oral anaphylaxis is dependent upon increased microbial diversity during early life. Their work thus suggests that there may be a

diversity threshold necessary for proper maturation of these Th2 immune mechanisms. Furthermore, several other studies have demonstrated the immune consequences of the disruption of the early-life gut microbial community using antibiotics. Antibiotics induce drastic compositional and diversity shifts that lead to changes in crucial immune functions, including Treg proliferation (60, 61), IgE response (60, 62), Th-17 response (61, 63), and basophil-mediated Th2-cell responses (62). Given the influence exerted by these immune functions on widespread tissues and systems, it is not surprising that antibiotic-induced immune alterations during early life in animal models aggravate autoimmune diabetes (61, 64), allergic lung inflammation (60, 62, 63), inflammatory chronic colitis (65), and obesity (20, 21).

Early-life immune development is also reliant on the actions of a group of bacterial metabolites known as short-chain fatty acids (SCFAs). These compounds are direct by-products of bacterial colonic fermentation and are produced at very high rates (66). Acetate, propionate, and butyrate are the SCFAs produced in highest concentrations in the human gut, and are rapidly taken up by the gut epithelium through passive and active transport mechanisms (67). SCFAs are essential energy sources for colonocytes cells in the mammalian gut, and are precursors for gluconeogenesis, liponeogenesis, and protein and cholesterol synthesis (68). Among many of their immune functions [reviewed in Ref. (66)], SCFAs have been shown to induce extrathymic proliferation of FoxP3⁺ T cells (68–70), which orchestrate peripheral tolerance in mucosal tissues. This critical immune function of SCFA has been shown to be relevant for the offspring even if exposure occurred before birth. Oral administration of acetate during pregnancy was sufficient for the priming of FoxP3⁺ Treg cells and preventing allergic airway inflammation in the adult offspring (36), suggesting that *in utero* exposure to maternal gut microbial metabolites contributes to the development of immune functions in the airways of the offspring.

In addition to interactions with the developing immune system, a recent study by Kim et al. (71) suggests that the early gut microbiome confers colonization resistance through the production of bacterial metabolites resulting from age-dependent colonization with key bacterial taxa. Clostridial species from *Clostridium* clusters IV and XIVa, which increase in abundance with age, induced colonization resistance to intestinal mouse pathogens *Salmonella enterica* subsp. *typhimurium* and *Citrobacter rodentium*. Interestingly, the conferred mechanism of resistance is unrelated to immune adaptors MyD88 and TRIF, and independent of B and T cell function. The settlement of *Clostridia* in the gut of GF mice was also greatly reduced by the absence of neonatal bacteria, which may help explain the increased susceptibility of newborns and young infants to these GI infections.

Collectively, these studies constitute compelling evidence that key taxa, microbial community diversity, and bacterial metabolites constitute modulatory triggers of host immune function maturation. Although considerable research effort has been made, a great deal of the age-dependent processes through which microbial exposure drives immune system development remains to be identified. The infant gut microbiome temporal succession patterns, driven by birth, weaning, and introduction of solid foods, match marked changes in host immune function (72, 73).

Therefore, future studies designed during these events, such as human longitudinal cohorts, hold great potential to improve our understanding of the dynamics at play.

TEMPORALLY STRUCTURED ENVIRONMENTAL FACTORS

Succession in ecology is defined as the pattern of changes in a community after a disturbance or after the opening of a new patch to colonization (74). Correspondingly, succession in the infant gut microbiome starts with the arrival of pioneer species that transform the gut habitat and enable the settlement of first succession species. The identity of the infant gut pioneer and first succession species is influenced by factors such as maternal factors (e.g., body weight and stress) (35–37), delivery mode (38–41), and type of milk consumption [(27, 42); **Figure 1**]. The temporal structure of these environmental factors contributes to the identity and dynamics of the infant gut microbiome and plays a role in the immune system training.

Prenatal Life

Even before birth, fetal immune development relies on microbial products present in the placenta. In an experimental system in

which germ-free mice were transiently colonized with genetically engineered *E. coli* HA107, maternal gut colonization influenced the offspring's immune system by increasing the intestinal group 3 innate lymphoid cells and F4/80⁺CD11c⁺ mononuclear cells (iMNCs), and strongly altering the offspring's intestinal transcriptional profiles (37). These early shifts in the offspring immune system improved the capacity of the pups to avoid inflammatory responses to MAMPs and intestinal microbes' penetration, thus suggesting that microbial training of the immune system starts *in utero* (37). Despite some reports suggesting that fetal colonization may begin *in utero* (75, 76), lack of appropriate contamination controls and failure to show bacterial viability in these studies yields this work inconclusive and inadequate to disprove the currently accepted view of the placenta as a sterile environment (77). More importantly, several studies have shown that early colonization of the infant gut is strongly driven by mode of birth (39–41, 78), thus suggesting that direct colonization of the infant gut most likely begins after membrane rupture, during labor and birth. For example, Backhed et al. (39, 40) showed that the gut microbiome of vaginally born infants exhibited an enrichment in *Bifidobacterium*, *Bacteroides*, *Escherichia*, and *Parabacteroides*. In comparison, the gut microbiome of infants born through cesarean sections (C-sections) was enriched in

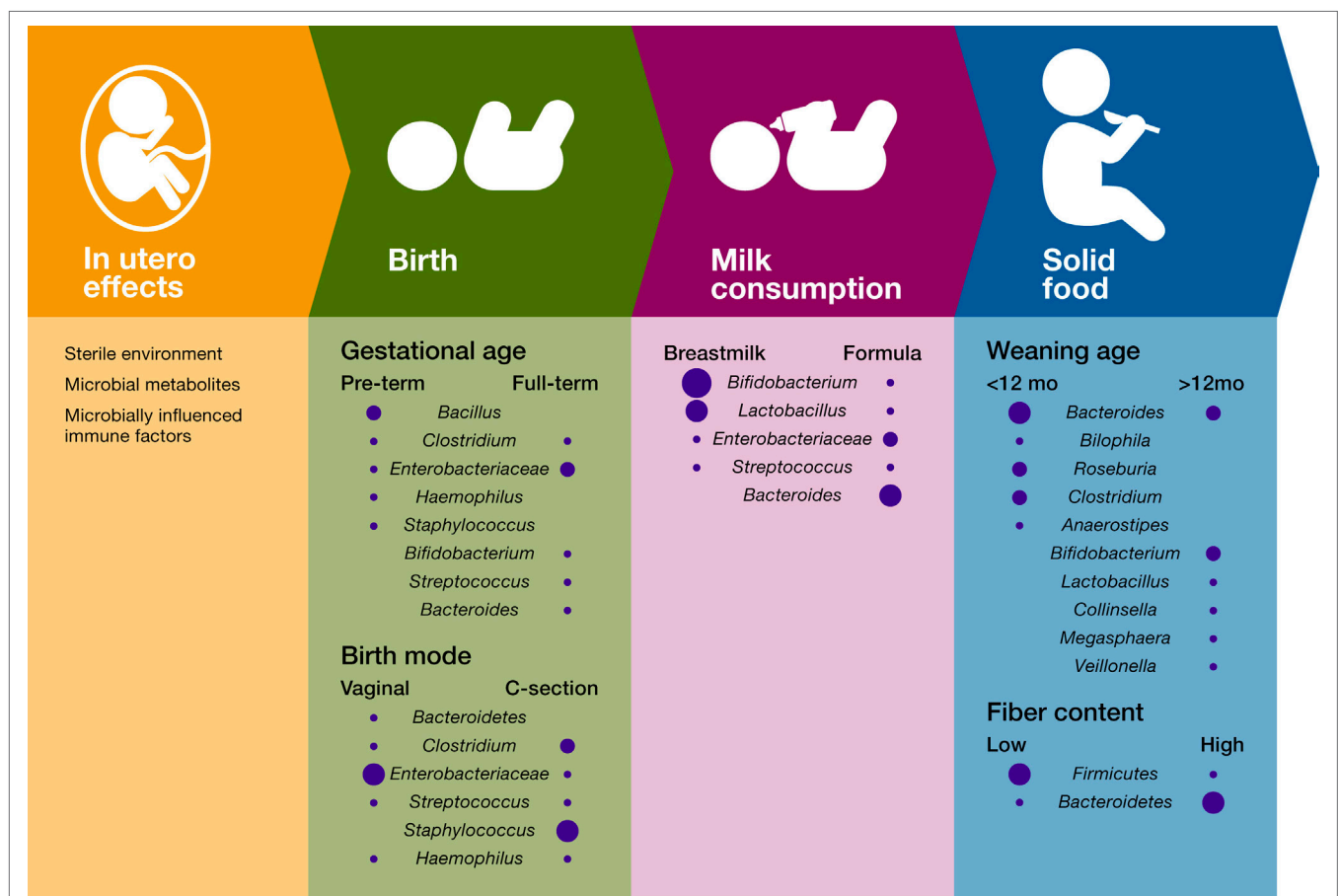


FIGURE 1 | Influence of temporal succession events and environmental factors on the infant gut bacterial microbiome. Only the most important differences in bacterial composition are included for each variable, and the size of the circle is proportional to the relative abundance of the bacterial taxa.

microbes associated with the skin, the mouth, and the surrounding environment.

Birth

The infant's gut habitat changes rapidly after birth with facultative anaerobes species (e.g., *E. coli*, *Staphylococcus*, and *Streptococcus*) colonizing first and consuming the available oxygen (79). A longitudinal study following 39 infants from birth demonstrated that mode of delivery impacts *Bacteroides* populations in the infant's gut microbiota between 6 and 18 months of age (41). Yassour et al. (41) showed that, in comparison with vaginal birth, most infants born by C-section lacked the presence of the *Bacteroides* genus until about 6–18 months of age. Their work also showed that a higher abundance of *Bifidobacterium* species, both in C-section and vaginally delivered infants, was detected concomitantly with a lower abundance of *Bacteroides*, suggesting that infant gut microbial communities are also influenced by microbe–microbe interactions. Delayed colonization with *Bacteroides* species was also associated with cesarean sections in a study of 24 infants (80), a finding that was linked to lower levels of Th-1-associated chemokines CXCL10 and CXCL11 in blood. *Bacteroides* species are important and extremely common members of the human gut microbiome, capable of fermenting a variety of fibers in the colon (81) and modulating the immune system (their potential as a keystone taxa and role in immunomodulation is discussed later in this review). Hence, vertical transmission during vaginal birth is likely a structured environmental factor that promotes colonization by members of this influential bacterial group.

Gut microbiome differences driven by mode of birth have been reported in almost all microbiome infant studies that recorded this variable (28, 38, 82–87). Although the cumulated evidence points to the mode of birth being a major influence in the gut pioneer microbiome, one recent study performed on 115 infants showed no differences in the meconium microbial communities between both mode of birth (C-section and vaginal delivery) (88). Unfortunately, key bio-statistical parameters of their analyses are missing from the paper, crucial information to assess the robustness of their results. What still remains unclear is how long these differences last, with only a few reports showing differences beyond early childhood (82). Nonetheless, changes in important microbial groups, community diversity, or functions during this critical and permissive window of immune development are likely to induce immune alterations that may remain beyond the age at which these taxonomic differences are no longer detectable.

Intriguingly, the taxonomic identity of pioneer colonizers not only depends on birth but also on gestational age. A 1-month longitudinal study of 58 preterm infants in a neonatal intensive care unit showed that time post-conception can also impact the type of early colonizers in the premature gut yet not the pattern of bacterial succession (89). Members of the *Bacilli* class appear as the initial colonizers in premature infants, which contrasts with the initial colonization with *Enterobacteriaceae* members in most term babies (Figure 1). In addition, this study showed that gut microbiome follows a progression strongly determined by host biology factors, suggesting that, during the first month

after delivery, the genetic and physiologic characteristics of the preterm infant gut drive a conserved pattern of succession in gut microbiome.

Milk Consumption

In addition to mode of birth and post-conceptual age, diet during early infancy strongly impacts community structure and diversity. Comparisons between breastfed and formula-fed infants have shown that *Bifidobacterium* spp. and *Lactobacillus* spp. predominate in breastfed infants whereas formula-fed infants exhibit higher proportions of *Bacteroides* spp., *Clostridium* spp., *Streptococcus* spp., *Enterobacter* spp., *Citrobacter* spp., and *Veillonella* spp. (39, 40, 90–93). Breast milk can modulate the infant gut microbiome through different mechanisms. First, human breast milk contains a significant number of bacteria that is passed to the infant constantly during the first months of life (46, 47, 94–98). Besides being a direct source of microbes, human milk contains a group of unconjugated glycan resistant to human enzymatic digestion known as human milk oligosaccharides (HMOs). These compounds act as prebiotics for key infant gut taxonomic bacterial groups including *Bifidobacterium* (99–103) and *Bacteroides* species (104). Importantly, fermentation of HMOs results in the production of SCFAs (102, 103), increases secretory immunoglobulin A (sIgA) production, and improves gut microbiome resistance to pathogens (105, 106).

Breastfeeding also influences the training of the infant immune system through the presence of antimicrobial compounds in the human milk (lactoferrin, lysozyme) and immune effectors [sIgA, immune cells, and cytokines; (107)]. Bridgman et al. (108) demonstrated that sIgA abundance is associated with breastfeeding status in a cohort of 47 4-month-old infants. sIgA is critical for the infant gut mucosal immune defense [see Ref. (109) for a review] mainly through a process known as immune exclusion, where sIgA adheres to bacterial cells and antigens and prevents their access to the gut epithelium (110). Although this antibody is initially acquired through breastfeeding, the infant gut microbiota will ultimately stimulate its local production through the maturation of B cells (111). Notably, the risk of developing atopy is increased if B cells maturation is delayed (112–115), stressing the importance of breastfeeding in infant gut microbiome and immune development.

Solid Food Introduction and Weaning

The introduction of solid foods constitutes the last step in early-life microbiome succession events, which leads to the consolidation of a gut microbial community that remains largely stable for the remainder of childhood and adult life. Due to the availability of new fiber sources and other substrates, transition to solid foods results in an increase of diversity and the enrichment of *Bacteroides* spp., *Clostridium* spp., *Ruminococcus* spp., *Faecalibacterium* spp., *Roseburia* spp., and *Anaerostipes* spp., as well as the reduction in *Bifidobacterium* spp. and *Enterobacteriaceae* (39, 40, 116, 117). Functionally, solid food introduction increases SCFA production, vitamin biosynthesis, and xenobiotic degradation (34, 39, 40). Notably, these changes coincide with important aspects of digestive development (e.g., pancreatic function and intestinal nutrient absorption) and shifts in immune development, some of

which are driven by microbes. For instance, the expression of the epithelial antimicrobial granule protein, Angiogenin-4 (Ang4), and of epithelial fucosylated glycans is markedly increased during weaning in conventional but not in germ-free mice. Remarkably, colonization with *Bacteroides thetaiotaomicron*, a bacterial commensal that increases in abundance post-weaning, was able to induce both Ang4 expression and fucosylated glycan reprogramming [Ang4; (118, 119)], strongly suggesting that specific host functions have adapted to rely on microbial signals that arrive in a temporally structured manner.

Furthermore, it has been suggested that cessation of breast-feeding, rather than solid food introduction, drives the main compositional shifts that result in an “adult-like” gut microbiome. In a longitudinal study of 98 infants, an early weaning age (under 12 months) was associated with an increase in *Bacteroides* spp., *Bilophila* spp., *Roseburia* spp., *Clostridium* spp., and *Anaerostipes* spp. In comparison, breastmilk supplementation beyond this age favored a more “immature” community composition, characterized by *Bifidobacterium* spp., *Lactobacillus* spp., *Collinsella* spp., *Megasphaera* spp., and *Veillonella* spp. (39, 40).

INFANT GUT COMMUNITY DIVERSITY: AN INDICATOR OF HEALTH?

The impact of early-life dysbiosis on the risk of developing several human diseases has led to the hypothesis that there is a critical window during which changes in the gut microbiome are most influential in immune development. During this “window of opportunity,” the infant gut harbors a highly variable and increasingly diverse microbial community of low resilience, which renders it easily disrupted by disturbances such as antibiotic treatments (41). During this period of time, a loss of diversity or change in community composition has the potential to disrupt the development of certain aspects of neonate immune system and to promote a bloom of pathogens, thus increasing the risk of developing immune-mediated and infectious diseases. However, it remains unclear if community diversity *per se* represents a robust indicator of infant gut microbiome disruption, especially since (1) there could be a threshold to be crossed for the gut ecosystem to suffer a significant loss of function; and (2) diversity as a diagnostic tool provides no information on the gut microbial community composition or functional properties.

Many studies have argued that a loss of community diversity could indicate a disruption of the natural infant gut microbiome community. After birth, both the taxonomic and functional diversity of the infant bacterial microbiome have been shown to increase (88). Life-threatening diseases such as NEC have been suggested to occur as an effect of disruption of the natural succession in the infant gut microbiome after antibiotic treatment (120, 121), lowering community diversity and creating an opportunity for other bacterial groups (e.g., Gammaproteobacteria) to dominate the normal bacterial community (122, 123). At that stage, a loss of community diversity can also hinder the training of the immune system by reducing its ability to recognize commensal bacteria [see Ref. (52) for a review]. Recent studies have confirmed that a significant loss in gut microbial diversity

is indicative of an increased risk of developing autoimmune diseases (80, 124). In addition, a loss of diversity can promote a long-term increase in IgE levels, which has been suggested to trigger immune-mediated disorders in mice (59).

However, it remains to be determined if the link between the development of immune diseases and the loss of microbial diversity is caused by a reduction of microbial species alone or, more precisely, by a loss of key taxonomic or functional microbial groups essential to the development of the infant immune system. The work of Arrieta et al. (13) on 319 infants in a longitudinal cohort, showed no significant relationship between fecal microbial alpha-diversity and the risk of developing asthma. Yet, four bacterial taxa (*Faecalibacterium*, *Lachnospira*, *Rothia*, and *Veillonella*), fecal acetate and deconjugated bile acids were significantly altered in babies at risk of asthma. By contrast, Kostic et al. (15) identified that a significant reduction in infant gut community alpha-diversity is a characteristic condition of the T1D state in a cohort of 33 infants predisposed to type 1 diabetes. This loss in alpha-diversity was combined with an alteration of the metabolic pathways and microbial community phylogenetic structure (15). These studies suggest that both subtle and global changes in community composition may lead to immune impairment and disease development, and that functional dysbiosis can occur independently of significant changes in community alpha-diversity.

Community alpha-diversity may also not be a reliable indicator across all human populations given its geographic variability (27). In a study comparing European to Burkina Faso children, De Filippo et al. (125) showed that the latter group had a greater gut microbial diversity and shift in community composition, potentially associated with their high fiber diet. However, other lifestyle factors and environmental exposures may also explain these differences. In addition, bacterial alpha-diversity fluctuates significantly during the first year of life, making it an unreliable ecosystem measurement unless studies are strictly age- and population matched. Further, an opposite relationship between alpha-diversity and health status occurs during the first weeks of life, where lower alpha-diversity and a predominance of a few subspecies of *Bifidobacterium longum* is associated with better growth (126).

Another factor that is rarely taken into account when assessing microbiome alpha-diversity is the impact of other non-bacterial microbes. In a unique study targeting both infant gut bacterial and fungal communities, Fujimura et al. (14) showed that infant gut bacterial alpha-diversity increased with time while the fungal alpha-diversity decreased in reciprocal correlation. This finding suggests that microbial diversity *per se* might naturally fluctuate depending on the targeted organism and that currently unexplored inter-kingdom gut microbial associations may influence these dynamics. Most interestingly, their work demonstrated that the fungal beta-diversity better predicted atopy risk than bacterial beta-diversity. Therefore, fluctuations in infant gut fungal community composition could play a role in influencing infant's susceptibility to childhood allergies and asthma.

The increase in both taxonomic and functional diversity of the infant bacterial gut microbiome in the few months after birth

appears to be associated with multiple aspects of the immune system development, providing further evidence that the immune system relies on a temporally structured succession of the gut microbiome. However, the infant gut microbial diversity *per se* might not be an indicator conveying enough information to be considered as a diagnostic tool. Notwithstanding, studies to date do suggest that the training of the immune system relies on a particular pattern of microbial diversity increasing from birth until 3 years old, and that disrupting this pattern can increase the risk of developing immune-mediated disorders. Future research disentangling the relative impact of species richness, community taxonomic, and functional composition on the retention of infant gut ecosystem emergent properties (e.g., infant immune system development) will provide key information for the development of diagnostic tools.

KEYSTONE GROUPS

In community ecology, the concept of a keystone species or group of species is described as an actor of a community that is so important to its organization and diversity that losing it provokes a massive cascade of extinctions and loss of ecosystem function (23, 24, 127). In other words, a keystone species has a remarkable impact in relation to its abundance (128). In an ecosystem, keystone species can belong to any trophic levels, from low-level species providing the resources on which a plethora of other species depends, to high-level species applying top-down regulation on the community. Keystone taxa of the infant gut microbiome contribute significantly to the ecosystem by (1) contributing to the establishment of other species; (2) by producing important metabolites including SCFAs (e.g., butyrate) that trigger local trophic cascades; (3) by improving ecosystem resistance against invading pathogenic species; and (4) by aiding in sustaining a balanced symbiosis with the host, which will in turn favor the stability of the microbial ecosystem. Because of the high inter-individual [i.e., Ref. (7)] and temporal (27, 34, 39, 40) variability of the infant gut ecosystem, identifying keystone taxa is a great challenge. Here, we discuss the potential for *Bifidobacterium* and *Bacteroides* to be keystone taxa and their role on infant immune system training.

Bifidobacterium

Bifidobacteria are dominant members of the infant gut microbiome, have a large repertoire of genes for the digestion of HMOs (104, 129), and have been isolated from maternal feces, human milk, and infant feces (130, 131), demonstrating how well adapted they are to the transmission routes and growth conditions in the infant gut. *B. longum* is the predominant species in the human gut, but several *B. longum* subspecies have different levels of adaptability and functionality in the infant gut. *B. longum* subsp. *infantis* (*Bifidobacterium infantis*), *B. longum* subsp. *longum* (*B. longum*), and *B. longum* subsp. *breve* (*B. breve*) are commonly isolated from healthy breastfed infant feces, while formula-fed infants are also colonized with *Bifidobacterium adolescentis* (132–134). Of these subspecies, *B. infantis* has the largest gene repertoire to digest all HMO structures in human milk (129). In addition, when administered as a probiotic

to preterm neonates, *B. infantis* colonizes better than other subspecies (135), which may explain why clinical trials using *B. lactis* or *B. breve* as a probiotic strain in the prevention of NEC have been unsuccessful (136, 137), while 5 out of 6 trials using *B. infantis* have shown to be effective in decreasing NEC incidence in neonates (138–143).

Bifidobacterium species decrease the intestinal luminal pH through the production of lactate and acetate, which is considered a crucial strategy in increasing intestinal nutrient absorption (144). Acetate accounts for more than 80% of the SCFA production in the infant gut (13) [compared to over 50% in the adult gut (145)] and is a key metabolite in the early establishment of colonization resistance, by preventing infections with enteropathogens (146, 147).

Through a process known as metabolic cross-feeding, where the metabolic products of a species or group of species provide growth substrates for other populations, *Bifidobacterium*'s production of lactate and acetate sustains the growth of other species, such as *Roseburia*, *Eubacterium*, *Faecalibacterium*, and *Anaerostipes* (148–151). In addition to this strong influence of microbe–microbe interactions, the sustained growth of other microbial species also enables the subsequent production of butyrate (152, 153). Notably, the lower abundance of colonization with *Bifidobacterium* in formula-fed babies is associated with a lower concentration of lactate and a higher gut luminal pH compared to breastfed babies (93, 154), and likely accounts for one of the root causes of the striking microbiome discrepancies observed between breastfed and formula-fed infants.

Bifidobacteria also play an exceptionally important role through its direct interactions with the developing immune system. Besides preventing enteropathogenic infections, *Bifidobacterium* species also protect the infant gut by modulating mucosal barrier function and promoting immunological and inflammatory responses (155, 156). The dominance of the infant gut microbiome by *Bifidobacterium* spp. was associated with an improved T-cell-mediated response to oral and parenteral vaccines and with lower neutrophilia at 15 weeks of age (126). *B. breve* has also evolved a mechanism to be protected from the immune system response by synthesizing a specific exopolysaccharide that increases its competitive power for space and colonization in the mouse gut (157, 158).

Collectively, *Bifidobacterium* species possess important strategies that insure their colonization at high abundance in the infant gut, prevent the growth of competing species that disfavor host fitness, and promote immune development. Due to the very high microbial inter-individual variation, and the number of subspecies found in the infant gut, it remains unclear if *Bifidobacterium* is a biomarker of infant gut health, yet the sub-species *B. infantis* may be a likely candidate.

Bacteroides

Together with *Bifidobacterium*, *Bacteroides* are the only groups known to use HMOs as a primary nutrient source (102, 103, 159). In addition, *Bacteroides* species are considered *generalists*—organisms with a great capacity to switch dietary nutrient sources or host-derived substrates (151). In an elegant study that followed the transcriptional profile of the human and murine

symbiont, *B. thetaiotaomicron*, and the structure of murine cecal glycans, it was demonstrated that this bacterium has the gene encoding capacity to switch from digesting food sugars to foraging host mucus glycans (160). The metabolic plasticity of this species likely improves their adaptability to the fluctuating luminal conditions of the developing infant gut, especially after weaning and introduction of solid foods. Importantly, colonization with *Bacteroides* species is heavily reliant on natural events that drive succession patterns, such as vaginal birth and breastfeeding (41, 80), suggesting that *Bacteroides* spp. transmission is advantageous for both the host and members of this taxa, and that it is highly coevolved.

Certain symbionts are thought to have evolved mechanisms through which they influence the host immune system maturation in a way that is beneficial for them. An example of these mechanisms is the development of specific metabolic capacity by *B. thetaiotaomicron* (119), a microbial species previously linked with angiogenesis in the postnatal intestine development (161). This species influences the gut microbial community by regulating the epithelial glycan synthesis (162), therefore creating a specific niche for itself and for other microorganisms with similar nutrient biochemical capacity.

Another species involved in immune system development is *Bacteroides fragilis*. Its production of polysaccharide A has been shown to suppress inflammation by downregulating interleukin (IL)-17 (163). Monocolonization of germ-free mice by *B. fragilis* has been shown to balance Th1 and Th2 responses (164). In addition, these monocolonized mice showed an increase in the conversion of CD4⁺ T cells into IL-10-producing Foxp3⁺ Treg cells, which induced a strong anti-inflammatory effect during gut inflammation (165). *B. fragilis* was also demonstrated to be negatively associated with the expression of toll-like receptor-4 and with lipopolysaccharide (LPS)-induced production of multiple inflammatory cytokines and chemokines (166).

Intriguingly, recent findings on the links between *Bacteroides* and immune system training suggest that, although they are important members of the early gut microbiome, an overabundance of *Bacteroides* spp. and a corresponding increase of exposure to their LPS, result in improper stimulation of the innate immune system and in inhibition of LPS tolerance in non-obese diabetic mice. This mechanism was proposed to explain the disparity in type 1 diabetes incidence in Northern Europe, where Russian children have reduced *Bacteroides* spp. abundance and lower disease rates, compared to Finnish and Estonian children (16). This study highlights the importance of attaining a balanced stimulation of the immune response early in life and how specific gut microbes have evolved to do so in a temporally structured manner. It also underlines the complexity of disentangling the effects of particular bacterial species and higher phylogenetic groups on the emergent properties of the infant gut ecosystem and host fitness.

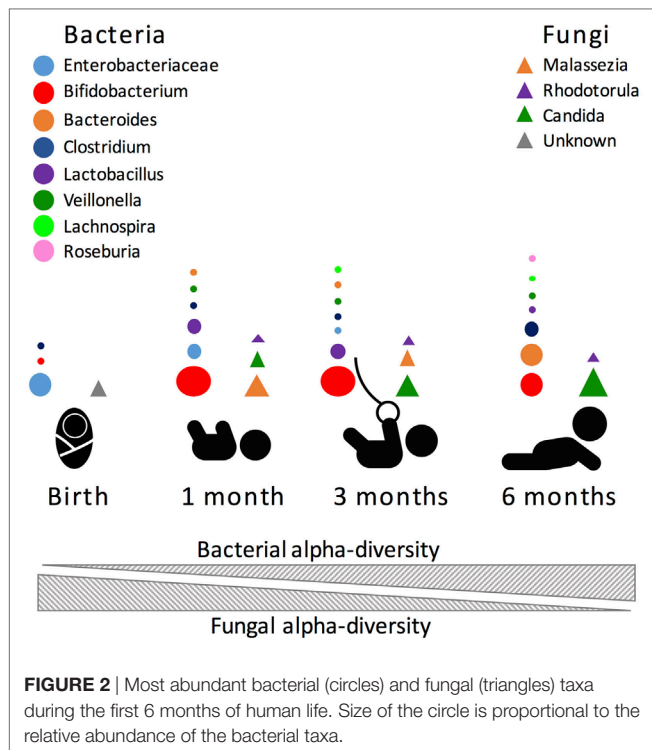
FUTURE RESEARCH

At its beginning, complexity theory suggested that ecosystems exhibiting a higher complexity were more stable when sustaining disturbances such as species loss (167, 168). However, mathematical

model simulations of food webs led to the proposal that instead of focusing on the stability of individual populations within an ecosystem, a better comprehension of complex systems could be gained from studying emergent properties such as productivity, resilience, and biomass (169, 170). From this point, studies have employed multiple properties to characterize ecosystems including species richness, taxonomic composition, functional profile, the level of interactions between species of the ecosystem, and the strength of these interactions. This transition in community ecology theory mirrors the improvement of our comprehension of complex ecosystems shifting from a singular to a multi-level perspective.

In this review, we advocate that the infant gut microbiome should be considered as a complex adaptive system crucial to the maintenance of various emergent properties (e.g., infant immune system training). These ecosystem properties are hardly attributable to a single group, instead they seem to rely on a temporally structured pattern of bacterial diversity increase after birth and the succession of particular keystone groups. The properties of complex adaptive systems highlight the great challenges faced by studies of the infant gut microbiome: a system far-from equilibrium dynamics, characterized by permanent novelty and incessant adaptation, dispersed multi-level interactions, and the absence of a global controller (171). The emergent properties of this ecosystem highlight the necessity of prospective, longitudinal infant gut microbiome studies, both taxonomic and functional, which will eventually allow us to identify the critical points at which this system loses its emergent properties and reaches a state of dysbiosis, impeding adequate immune system development. In addition, there is a need to disentangle the influence of loss of taxonomic and functional diversity, as well as of shifts in keystone taxa on immune system training and subsequent disease development. From past studies, we now understand that the maturation of the immune system relies on a temporally structured dynamic, starting *in utero* with maternal effects, influenced by environmental factors (delivery mode, type of milk consumption, and solid foods) and host biology, and depending heavily on auto-correlated local interactions between microbial groups. Further understanding of this complex adaptive system will also require (1) sampling a variety of geographically distinct human populations, (2) carrying out longitudinal cohorts that sample numerous times during the first 12 months, and (3) combining amplicon-based surveys with functional assays, such as metagenomics and metabolomics.

Another important influence in gut microbiome composition that remains vastly unexplored is the role of non-bacterial microorganisms. The role of the virome, the collection of viruses colonizing the host, has been previously explored in adult animals. Similarly to the bacteriome, the virome strongly interacts with the host immune system, with both positive and negative consequences for host health [see Ref. (172–174) for reviews]. However, it remains unknown what role the virome has during early-life immune development. Further, fungi, protozoans, and helminths, which are traditionally excluded from culture- and non-culture-based studies, are important and immunomodulatory members of the gut microbiome, albeit in



smaller proportions than bacteria. Nonetheless, it was recently shown that fungi species are present at much higher diversity in the first months of life, compared to later months, and that this change in diversity inversely correlates with bacterial diversity [(14); **Figure 2**]. Future studies directed at exploring inter-kingdom gut microbial associations during early life and how

these associations influence the host will provide a more global understanding of the microbial triggers influencing immune system development.

Eventually, the identification of the critical events and factors that influence microbiome resilience and function will enable the development of effective interventions aimed at maintaining and/or improving immune system development and disease prevention. Although an astounding amount of work has been carried out to understand the reliance of the immune system on the infant gut microbiome, much remains to be elucidated on the particular mechanisms responsible for this training. Improvements in our understanding will arise from continuing multidisciplinary joint efforts between immunologists, microbiologists, clinicians, bioinformaticians, and ecologists.

AUTHOR CONTRIBUTIONS

M-CA formulated the concept for this review, and IL-L wrote the first draft. Both authors co-wrote and revised the entire review article.

ACKNOWLEDGMENTS

The authors thank Hypothesis Media for producing the art of **Figure 1**, as well as Dr. Jens Walter for thoroughly reviewing this work.

FUNDING

M-CA is funded by grants from the Canadian Institutes of Health Research (CIHR) and by the University of Calgary.

REFERENCES

- Arrieta MC, Stiemsma LT, Amenogbe N, Brown EM, Finlay B. The intestinal microbiome in early life: health and disease. *Front Immunol* (2014) 5:427. doi:10.3389/fimmu.2014.00427
- Strachan DP. Hay fever, hygiene, and household size. *BMJ* (1989) 299(6710):1259. doi:10.1136/bmj.299.6710.1259
- Strachan DP. Family size, infection and atopy: the first decade of the 'hygiene hypothesis'. *Thorax* (2000) 55(Suppl 1):S2. doi:10.1136/thorax.55.suppl_1.S2
- Penders J, Stobberingh EE, van den Brandt PA, Thijs C. The role of the intestinal microbiota in the development of atopic disorders. *Allergy* (2007) 62(11):1223–36. doi:10.1111/j.1398-9995.2007.01462.x
- Penders J, Thijs C, van den Brandt PA, Kummeling I, Snijders B, Stelma F, et al. Gut microbiota composition and development of atopic manifestations in infancy: the KOALA birth cohort study. *Gut* (2007) 56(5):661–7. doi:10.1136/gut.2006.100164
- Russell SL, Finlay BB. The impact of gut microbes in allergic diseases. *Curr Opin Gastroenterol* (2012) 28(6):563–9. doi:10.1097/MOG.0b013e3283573017
- Hill CJ, Lynch DB, Murphy K, Ulaszewska M, Jeffery IB, O'Shea CA, et al. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET cohort. *Microbiome* (2017) 5(1):4. doi:10.1186/s40168-016-0213-y
- Prescott SL, Macaubas C, Smallacombe T, Holt BJ, Sly PD, Holt PG. Development of allergen-specific T-cell memory in atopic and normal children. *Lancet* (1999) 353(9148):196–200. doi:10.1016/S0140-6736(98)05104-6
- Van Der Velden V, Laan M, Baert M, De Waal Malefyt R, Neijens H, Savelkoul H. Selective development of a strong Th2 cytokine profile in high-risk children who develop atopy: risk factors and regulatory role of IFN- γ , IL-4 and IL-10. *Clin Exp Allergy* (2001) 31(7):997–1006. doi:10.1046/j.1365-2222.2001.01176.x
- Rautava S, Ruuskanen O, Ouwehand A, Salminen S, Isolauri E. The hygiene hypothesis of atopic disease—an extended version. *J Pediatr Gastroenterol Nutr* (2004) 38(4):378–88. doi:10.1097/00005176-200404000-00004
- Costello EK, Stagaman K, Dethlefsen L, Bohannan BJ, Relman DA. The application of ecological theory toward an understanding of the human microbiome. *Science* (2012) 336(6086):1255–62. doi:10.1126/science.1224203
- Cassir N, Benamar S, Khalil JB, Croce O, Saint-Faust M, Jacquot A, et al. *Clostridium butyricum* strains and dysbiosis linked to necrotizing enterocolitis in preterm neonates. *Clin Infect Dis* (2015) 61(7):1107–15. doi:10.1093/cid/civ468
- Arrieta MC, Stiemsma LT, Dimitriu PA, Thorson L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* (2015) 7(307):307ra152. doi:10.1126/scitranslmed.aab2271
- Fujimura KE, Sitarik AR, Havstad S, Lin DL, Levan S, Fadrosch D, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med* (2016) 22(10):1187–91. doi:10.1038/nm.4176
- Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hamalainen AM, et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host Microbe* (2015) 17(2):260–73. doi:10.1016/j.chom.2015.01.001
- Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in microbiome LPS immunogenicity contributes to autoimmunity in humans. *Cell* (2016) 165(6):1551. doi:10.1016/j.cell.2016.05.056

17. Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. *Microb Ecol Health Dis* (2015) 26(1):26191. doi:10.3402/mehd.v26.26191
18. Gevers D, Kugathasan S, Denson LA, ázquez-Baeza YV, Van Treuren W, Ren B, et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* (2014) 15(3):382–92. doi:10.1016/j.chom.2014.02.005
19. Imhann F, Bonder MJ, Vila AV, Fu J, Mujagic Z, Vork L, et al. Proton pump inhibitors affect the gut microbiome. *Gut* (2016) 65(5):740–8. doi:10.1136/gutjnl-2015-310376
20. Cho I, Yamanishi S, Cox L, Methé BA, Zavadil J, Li K, et al. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* (2012) 488(7413):621–6. doi:10.1038/nature11400
21. Cox LM, Yamanishi S, Sohn J, Alekseyenko AV, Leung JM, Cho I, et al. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* (2014) 158(4):705–21. doi:10.1016/j.cell.2014.05.052
22. Levin SA. Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* (1998) 1(5):431–6. doi:10.1007/s100219900037
23. Paine RT. Food web complexity and species diversity. *Am Nat* (1966) 100(910):65–75. doi:10.1086/282400
24. Paine RT. A note on trophic complexity and community stability. *Am Nat* (1969) 103(929):91–3. doi:10.1086/282586
25. Horn HS. The ecology of secondary succession. *Annu Rev Ecol Syst* (1974) 5(1):25–37. doi:10.1146/annurev.es.05.110174.000325
26. Spor A, Koren O, Ley R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* (2011) 9(4):279–90. doi:10.1038/nrmicro2540
27. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* (2012) 486(7402):222–7. doi:10.1038/nature11053
28. Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, ázquez-Baeza YV, et al. Meta-analyses of studies of the human microbiota. *Genome Res* (2013) 23(10):1704–14. doi:10.1101/gr.151803.112
29. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* (2012) 336(6086):1268–73. doi:10.1126/science.1223490
30. Gensollen T, Iyer SS, Kasper DL, Blumberg RS. How colonization by microbiota in early life shapes the immune system. *Science* (2016) 352(6285):539–44. doi:10.1126/science.aad9378
31. Gray LE, O'Hely M, Ranganathan S, Sly PD, Vuillermin P. The maternal diet, gut bacteria, and bacterial metabolites during pregnancy influence offspring asthma. *Front Immunol* (2017) 8:365. doi:10.3389/fimmu.2017.00365
32. Cho I, Blaser MJ. The human microbiome: at the interface of health and disease. *Nat Rev Genet* (2012) 13(4):260–70. doi:10.1038/nrg3182
33. Yu Y, Lu L, Sun J, Petrof EO, Claud EC. Preterm infant gut microbiota affects intestinal epithelial development in a humanized microbiome gnotobiotic mouse model. *Am J Physiol Gastrointest Liver Physiol* (2016) 311(3):G521–32. doi:10.1152/ajpgi.00022.2016
34. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* (2011) 108(Suppl 1):4578–85. doi:10.1073/pnas.1000081107
35. Cabrera-Rubio R, Collado MC, Laitinen K, Salminen S, Isolauri E, Mira A. The human milk microbiome changes over lactation and is shaped by maternal weight and mode of delivery. *Am J Clin Nutr* (2012) 96(3):544–51. doi:10.3945/ajcn.112.037382
36. Thorburn AN, McKenzie CI, Shen S, Stanley D, Macia L, Mason LJ, et al. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat Commun* (2015) 6:7320. doi:10.1038/ncomms8320
37. Gomez de Agüero M, Ganai-Vonarburg SC, Fuhrer T, Rupp S, Uchimura Y, Li H, et al. The maternal microbiota drives early postnatal innate immune development. *Science* (2016) 351(6279):1296–302. doi:10.1126/science.aad2571
38. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* (2010) 107(26):11971–5. doi:10.1073/pnas.1002601107
39. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* (2015) 17(5):690–703. doi:10.1016/j.chom.2015.04.004
40. Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* (2015) 17(6):852. doi:10.1016/j.chom.2015.05.012
41. Yassour M, Vatanen T, Siljander H, Hamalainen AM, Harkonen T, Ryhanen SJ, et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci Transl Med* (2016) 8(343):343ra81. doi:10.1126/scitranslmed.aad0917
42. Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the human infant intestinal microbiota. *PLoS Biol* (2007) 5(7):e177. doi:10.1371/journal.pbio.0050177
43. Round JL, Lee SM, Li J, Tran G, Jabri B, Chatila TA, et al. The toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* (2011) 332(6032):974–7. doi:10.1126/science.1206095
44. Tanaka S, Kobayashi T, Songjinda P, Tateyama A, Tsubouchi M, Kiyohara C, et al. Influence of antibiotic exposure in the early postnatal period on the development of intestinal microbiota. *FEMS Immunol Med Microbiol* (2009) 56(1):80–7. doi:10.1111/j.1574-695X.2009.00553.x
45. Zeissig S, Blumberg RS. Life at the beginning: perturbation of the microbiota by antibiotics in early life and its role in health and disease. *Nat Immunol* (2014) 15(4):307–10. doi:10.1038/ni.2847
46. Martín R, Heilig G, Zoetendal E, Smidt H, Rodríguez J. Diversity of the *Lactobacillus* group in breast milk and vagina of healthy women and potential role in the colonization of the infant gut. *J Appl Microbiol* (2007) 103(6):2638–44. doi:10.1111/j.1365-2672.2007.03497.x
47. Martín R, Heilig HG, Zoetendal EG, Jiménez E, Fernández L, Smidt H, et al. Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women. *Res Microbiol* (2007) 158(1):31–7. doi:10.1016/j.resmic.2006.11.004
48. Solís G, de Los Reyes-Gavilan C, Fernández N, Margolles A, Gueimonde M. Establishment and development of lactic acid bacteria and bifidobacteria microbiota in breast-milk and the infant gut. *Anaerobe* (2010) 16(3):307–10. doi:10.1016/j.anaerobe.2010.02.004
49. Gao J, Barzel B, Barabási A-L. Universal resilience patterns in complex networks. *Nature* (2016) 530(7590):307–12. doi:10.1038/nature16948
50. Horner AA. Toll-like receptor ligands and atopy: a coin with at least two sides. *J Allergy Clin Immunol* (2006) 117(5):1133–40. doi:10.1016/j.jaci.2006.02.035
51. Romagnani S. Coming back to a missing immune deviation as the main explanatory mechanism for the hygiene hypothesis. *J Allergy Clin Immunol* (2007) 119(6):1511. doi:10.1016/j.jaci.2007.04.005
52. Round JL, Mazmanian SK. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* (2009) 9(5):313–23. doi:10.1038/nri2515
53. Cummins AG, Thompson FM. Effect of breast milk and weaning on epithelial growth of the small intestine in humans. *Gut* (2002) 51(5):748–54. doi:10.1136/gut.51.5.748
54. Forchielli ML, Walker WA. The effect of protective nutrients on mucosal defense in the immature intestine. *Acta Paediatr Suppl* (2005) 94(449):74–83. doi:10.1080/08035320510043592
55. Lyons A, O'mahony D, O'brien F, MacSharry J, Sheil B, Ceadia M, et al. Bacterial strain-specific induction of Foxp3+ T regulatory cells is protective in murine allergy models. *Clin Exp Allergy* (2010) 40(5):811–9. doi:10.1111/j.1365-2222.2009.03437.x
56. Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, et al. Diversity of bifidobacteria within the infant gut microbiota. *PLoS One* (2012) 7(5):e36957. doi:10.1371/journal.pone.0036957
57. Isolauri E, Sutas Y, Kankaanpää P, Arvilommi H, Salminen S. Probiotics: effects on immunity. *Am J Clin Nutr* (2001) 73(2 Suppl):444S–50S.
58. Arnold IC, Dehzad N, Reuter S, Martin H, Becher B, Taube C, et al. *Helicobacter pylori* infection prevents allergic asthma in mouse models through the induction of regulatory T cells. *J Clin Invest* (2011) 121(8):3088–93. doi:10.1172/JCI45041

59. Cahenzli J, Koller Y, Wyss M, Geuking MB, McCoy KD. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe* (2013) 14(5):559–70. doi:10.1016/j.chom.2013.10.004
60. Russell SL, Gold MJ, Hartmann M, Willing BP, Thorson L, Wlodarska M, et al. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep* (2012) 13(5):440–7. doi:10.1038/embor.2012.32
61. Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, McRitchie S, et al. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. *Nat Microbiol* (2016) 1(11):16140. doi:10.1038/nmicrobiol.2016.140
62. Hill DA, Siracusa MC, Abt MC, Kim BS, Kobuley D, Kubo M, et al. Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. *Nat Med* (2012) 18(4):538–46. doi:10.1038/nm.2657
63. Russell SL, Gold MJ, Reynolds LA, Willing BP, Dimitriu P, Thorson L, et al. Perinatal antibiotic-induced shifts in gut microbiota have differential effects on inflammatory lung diseases. *J Allergy Clin Immunol* (2015) 135(1):100–9. doi:10.1016/j.jaci.2014.06.027
64. Li J, Yang K, Ju T, Ho T, McKay CA, Gao Y, et al. Early life antibiotic exposure affects pancreatic islet development and metabolic regulation. *Sci Rep* (2017) 7:41778. doi:10.1038/srep41778
65. Scheer S, Medina TS, Murison A, Taves MD, Antignano F, Chenery A, et al. Early-life antibiotic treatment enhances the pathogenicity of CD4+ T cells during intestinal inflammation. *J Leukoc Biol* (2017) 101(4):893–900. doi:10.1189/jlb.3MA0716-334RR
66. Corrêa-Oliveira R, Fachi JL, Vieira A, Sato FT, Vinolo MAR. Regulation of immune cell function by short-chain fatty acids. *Clin Transl Immunol* (2016) 5(4):e73. doi:10.1038/cti.2016.17
67. Nicholson JK, Holmes E, Kinross J, Burcelin R, Gibson G, Jia W, et al. Host-gut microbiota metabolic interactions. *Science* (2012) 336(6086):1262–7. doi:10.1126/science.1223813
68. Smith PM, Howitt MR, Panikov N, Michaud M, Gallini CA, Bohlooly-y M, et al. The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. *Science* (2013) 341(6145):569–73. doi:10.1126/science.1241165
69. Arpaia N, Campbell C, Fan X, Dikiy S, van der Veeken J, deRoos P, et al. Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* (2013) 504(7480):451–5. doi:10.1038/nature12726
70. Furusawa Y, Obata Y, Fukuda S, Endo TA, Nakato G, Takahashi D, et al. Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* (2013) 504(7480):446–50. doi:10.1038/nature12721
71. Kim YG, Sakamoto K, Seo SU, Pickard JM, Gilliland MG III, Pudlo NA, et al. Neonatal acquisition of *Clostridia* species protects against colonization by bacterial pathogens. *Science* (2017) 356(6335):315–9. doi:10.1126/science.aag2029
72. Bomba L, Minuti A, Moisés SJ, Trevisi E, Eufemi E, Lizier M, et al. Gut response induced by weaning in piglet features marked changes in immune and inflammatory response. *Funct Integr Genomics* (2014) 14(4):657–71. doi:10.1007/s10142-014-0396-x
73. Parigi SM, Eldh M, Larssen P, Gabriellsson S, Villablanca EJ. Breast milk and solid food shaping intestinal immunity. *Front Immunol* (2015) 6:415. doi:10.3389/fimmu.2015.00415
74. Clements FE. *Plant Succession: An Analysis of the Development of Vegetation*. Washington: Carnegie Institution of Washington (1916).
75. Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J. The placenta harbors a unique microbiome. *Sci Transl Med* (2014) 6(237):237ra65. doi:10.1126/scitranslmed.3008599
76. Collado MC, Rautava S, Aakko J, Isolauri E, Salminen S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci Rep* (2016) 6:23129. doi:10.1038/srep23129
77. Perez-Muñoz ME, Arrieta M-C, Ramer-Tait AE, Walter J. A critical assessment of the sterile womb and in utero colonization hypotheses: implications for research on the pioneer infant microbiome. *Microbiome* (2017) 5(1):48. doi:10.1186/s40168-017-0268-4
78. Dominguez-Bello MG, Cienfuentes C, Romero R, Garcia P, Gomez I, Mago V, et al. PCR detection of *Helicobacter pylori* in string-absorbed gastric juice. *FEMS Microbiol Lett* (2001) 198(1):15–6. doi:10.1111/j.1574-6968.2001.tb10612.x
79. Bezirtzoglou E. The intestinal microflora during the first weeks of life. *Anaerobe* (1997) 3(2–3):173–7. doi:10.1006/anae.1997.0102
80. Jakobsson HE, Abrahamsson TR, Jenmalm MC, Harris K, Quince C, Jernberg C, et al. Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* (2014) 63(4):559–66. doi:10.1136/gutjnl-2012-303249
81. Chen T, Long W, Zhang C, Liu S, Zhao L, Hamaker B. Fiber utilizing capacity varies with *Prevotella* versus *Bacteroides* enterotypes. *FASEB J* (2016) 30 (1 Suppl):683.2.
82. Salminen S, Gibson G, McCartney A, Isolauri E. Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut* (2004) 53(9):1388–9. doi:10.1136/gut.2004.041640
83. Adlerberth I, Lindberg E, Åberg N, Hesselmar B, Saalman R, Strannegård I-L, et al. Reduced enterobacterial and increased staphylococcal colonization of the infantile bowel: an effect of hygienic lifestyle? *Pediatr Res* (2006) 59(1):96–101. doi:10.1203/01.pdr.0000191137.12774.b2
84. Penders J, Thijs C, Vink C, Stelma FF, Snijders B, Kummeling I, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* (2006) 118(2):511–21. doi:10.1542/peds.2005-2824
85. Biasucci G, Benenati B, Morelli L, Bessi E, Boehm G. Cesarean delivery may affect the early biodiversity of intestinal bacteria. *J Nutr* (2008) 138 (9):1796S–800S.
86. Matamoros S, Gras-Leguen C, Le Vacon F, Potel G, de La Cochetiere M-F. Development of intestinal microbiota in infants and its impact on health. *Trends Microbiol* (2013) 21(4):167–73. doi:10.1016/j.tim.2012.12.001
87. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* (2016) 8(343):343ra82. doi:10.1126/scitranslmed.aad7121
88. Chu DM, Ma J, Prince AL, Antony KM, Seferovic MD, Aagaard KM. Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat Med* (2017) 23(3):314–26. doi:10.1038/nm.4272
89. La Rosa PS, Warner BB, Zhou Y, Weinstock GM, Sodergren E, Hall-Moore CM, et al. Patterned progression of bacterial populations in the premature infant gut. *Proc Natl Acad Sci U S A* (2014) 111(34):12522–7. doi:10.1073/pnas.1409497111
90. Penders J, Vink C, Driessen C, London N, Thijs C, Stobberingh EE. Quantification of *Bifidobacterium* spp., *Escherichia coli* and *Clostridium difficile* in faecal samples of breast-fed and formula-fed infants by real-time PCR. *FEMS Microbiol Lett* (2005) 243(1):141–7. doi:10.1016/j.femsle.2004.11.052
91. Adlerberth I, Wold AE. Establishment of the gut microbiota in Western infants. *Acta Paediatr* (2009) 98(2):229–38. doi:10.1111/j.1651-2227.2008.01060.x
92. Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, et al. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J Pediatr Gastroenterol Nutr* (2010) 51(1):77–84. doi:10.1097/MPG.0b013e3181d1b11e
93. Bezirtzoglou E, Tsiotsias A, Welling GW. Microbiota profile in feces of breast- and formula-fed newborns by using fluorescence in situ hybridization (FISH). *Anaerobe* (2011) 17(6):478–82. doi:10.1016/j.anaerobe.2011.03.009
94. Heikkilä M, Saris P. Inhibition of *Staphylococcus aureus* by the commensal bacteria of human milk. *J Appl Microbiol* (2003) 95(3):471–8. doi:10.1046/j.1365-2672.2003.02002.x
95. Gueimonde M, Laitinen K, Salminen S, Isolauri E. Breast milk: a source of bifidobacteria for infant gut development and maturation? *Neonatology* (2007) 92(1):64–6. doi:10.1159/000100088
96. Collado M, Delgado S, Maldonado A, Rodríguez J. Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett Appl Microbiol* (2009) 48(5):523–8. doi:10.1111/j.1472-765X.2009.02567.x
97. Martín R, Jiménez E, Heilig H, Fernández L, Marín ML, Zoetendal EG, et al. Isolation of bifidobacteria from breast milk and assessment of the bifidobacterial population by PCR-denaturing gradient gel electrophoresis

- and quantitative real-time PCR. *Appl Environ Microbiol* (2009) 75(4):965–9. doi:10.1128/AEM.02063-08
98. Hunt KM, Foster JA, Forney LJ, Schütte UM, Beck DL, Abdo Z, et al. Characterization of the diversity and temporal stability of bacterial communities in human milk. *PLoS One* (2011) 6(6):e21313. doi:10.1371/journal.pone.0021313
 99. Ward RE, Niaeonuevo M, Mills DA, Lebrilla CB, German JB. Research article in vitro fermentability of human milk oligosaccharides by several strains of bifidobacteria. *Mol Nutr Food Res* (2007) 51:1398–405. doi:10.1002/mnfr.200700150
 100. Sela DA, Mills DA. Nursing our microbiota: molecular linkages between bifidobacteria and milk oligosaccharides. *Trends Microbiol* (2010) 18(7):298–307. doi:10.1016/j.tim.2010.03.008
 101. Barile D, Rastall RA. Human milk and related oligosaccharides as prebiotics. *Curr Opin Biotechnol* (2013) 24(2):214–9. doi:10.1016/j.copbio.2013.01.008
 102. Yu Z-T, Chen C, Kling DE, Liu B, McCoy JM, Merighi M, et al. The principal fucosylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. *Glycobiology* (2013) 23(2):169–77. doi:10.1093/glycob/cws138
 103. Yu ZT, Chen C, Newburg DS. Utilization of major fucosylated and sialylated human milk oligosaccharides by isolated human gut microbes. *Glycobiology* (2013) 23(11):1281–92. doi:10.1093/glycob/cwt065
 104. Marcobal A, Sonnenburg JL. Human milk oligosaccharide consumption by intestinal microbiota. *Clin Microbiol Infect* (2012) 18(Suppl 4):12–5. doi:10.1111/j.1469-0691.2012.03863.x
 105. Newburg DS, Walker WA. Protection of the neonate by the innate immune system of developing gut and of human milk. *Pediatr Res* (2007) 61(1):2–8. doi:10.1203/01.pdr.0000250274.68571.18
 106. Newburg DS. Neonatal protection by an innate immune system of human milk consisting of oligosaccharides and glycans. *J Anim Sci* (2009) 87(13 Suppl):26–34. doi:10.2527/jas.2008-1347
 107. Moles L, Gomez M, Jimenez E, Fernandez L, Bustos G, Chaves F, et al. Preterm infant gut colonization in the neonatal ICU and complete restoration 2 years later. *Clin Microbiol Infect* (2015) 21(10):936.e1–10. doi:10.1016/j.cmi.2015.06.003
 108. Bridgman SL, Konya T, Azad MB, Sears MR, Becker AB, Turvey SE, et al. Infant gut immunity: a preliminary study of IgA associations with breastfeeding. *J Dev Orig Health Dis* (2016) 7(1):68–72. doi:10.1017/S2040174415007862
 109. Brandtzaeg P. Secretory IgA: designed for anti-microbial defense. *Front Immunol* (2013) 4:222. doi:10.3389/fimmu.2013.00222
 110. Brandtzaeg P. Mucosal immunity: induction, dissemination, and effector functions. *Scand J Immunol* (2009) 70(6):505–15. doi:10.1111/j.1365-3083.2009.02319.x
 111. Macpherson AJ, Uhr T. Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. *Science* (2004) 303(5664):1662–5. doi:10.1126/science.1091334
 112. Kukkonen K, Kuitunen M, Haahtela T, Korpela R, Poussa T, Savilahti E. High intestinal IgA associates with reduced risk of IgE-associated allergic diseases. *Pediatr Allergy Immunol* (2010) 21(1 Pt 1):67–73. doi:10.1111/j.1399-3038.2009.00907.x
 113. Sandin A, Björkstén B, Böttcher MF, Englund E, Jenmalm MC, Bråbäck L. High salivary secretory IgA antibody levels are associated with less late-onset wheezing in IgE-sensitized infants. *Pediatr Allergy Immunol* (2011) 22(5):477–81. doi:10.1111/j.1399-3038.2010.01106.x
 114. Orivuori L, Loss G, Roduit C, Dalphin JC, Depner M, Genuneit J, et al. Soluble immunoglobulin A in breast milk is inversely associated with atopic dermatitis at early age: the PASTURE cohort study. *Clin Exp Allergy* (2014) 44(1):102–12. doi:10.1111/cea.12199
 115. Orivuori L, Mustonen K, de Goffau MC, Hakala S, Paasela M, Roduit C, et al. High level of fecal calprotectin at age 2 months as a marker of intestinal inflammation predicts atopic dermatitis and asthma by age 6. *Clin Exp Allergy* (2015) 45(5):928–39. doi:10.1111/cea.12522
 116. Fallani M, Amarri S, Uusijarvi A, Adam R, Khanna S, Aguilera M, et al. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology* (2011) 157(5):1385–92. doi:10.1099/mic.0.042143-0
 117. Vallès Y, Artacho A, Pascual-García A, Ferrús ML, Gosálbes MJ, Abellán JJ, et al. Microbial succession in the gut: directional trends of taxonomic and functional change in a birth cohort of Spanish infants. *PLoS Genet* (2014) 10(6):e1004406. doi:10.1371/journal.pgen.1004406
 118. Bry L, Falk PG, Midtvedt T, Gordon JI. A model of host-microbial interactions in an open mammalian ecosystem. *Science* (1996) 273(5280):1380. doi:10.1126/science.273.5280.1380
 119. Hooper LV, Stappenbeck TS, Hong CV, Gordon JI. Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol* (2003) 4(3):269–73. doi:10.1038/ni888
 120. Cotten CM, Taylor S, Stoll B, Goldberg RN, Hansen NI, Sánchez PJ, et al. Prolonged duration of initial empirical antibiotic treatment is associated with increased rates of necrotizing enterocolitis and death for extremely low birth weight infants. *Pediatrics* (2009) 123(1):58–66. doi:10.1542/peds.2007-3423
 121. Alexander VN, Northrup V, Bizzarro MJ. Antibiotic exposure in the newborn intensive care unit and the risk of necrotizing enterocolitis. *J Pediatr* (2011) 159(3):392–7. doi:10.1016/j.jpeds.2011.02.035
 122. Wang Y, Hoenig JD, Malin KJ, Qamar S, Petrof EO, Sun J, et al. 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *ISME J* (2009) 3(8):944–54. doi:10.1038/ismej.2009.37
 123. Mai V, Young CM, Ukhanova M, Wang X, Sun Y, Casella G, et al. Fecal microbiota in premature infants prior to necrotizing enterocolitis. *PLoS One* (2011) 6(6):e20647. doi:10.1371/journal.pone.0020647
 124. Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin Exp Allergy* (2014) 44(6):842–50. doi:10.1111/cea.12253
 125. De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A* (2010) 107(33):14691–6. doi:10.1073/pnas.1005963107
 126. Huda MN, Lewis Z, Kalanetra KM, Rashid M, Ahmad SM, Raqib R, et al. Stool microbiota and vaccine responses of infants. *Pediatrics* (2014) 134(2):e362–72. doi:10.1542/peds.2013-3937
 127. Mills LS, Soule ME, Doak DF. The keystone-species concept in ecology and conservation. *Bioscience* (1993) 43(4):219–24. doi:10.2307/1312122
 128. Power ME, Tilman D, Estes JA, Menge BA, Bond WJ, Mills LS, et al. Challenges in the quest for keystones. *Bioscience* (1996) 46(8):609–20. doi:10.2307/1312990
 129. Underwood MA, German JB, Lebrilla CB, Mills DA. *Bifidobacterium longum* subspecies infantis: champion colonizer of the infant gut. *Pediatr Res* (2015) 77(1–2):229–35. doi:10.1038/pr.2014.156
 130. Jost T, Lacroix C, Braegger C, Chassard C. Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. *Br J Nutr* (2013) 110(7):1253–62. doi:10.1017/S0007114513000597
 131. Jost T, Lacroix C, Braegger C, Chassard C. Stability of the maternal gut microbiota during late pregnancy and early lactation. *Curr Microbiol* (2014) 68(4):419–27. doi:10.1007/s00284-013-0491-6
 132. Sakata S, Tonooka T, Ishizeki S, Takada M, Sakamoto M, Fukuyama M, et al. Culture-independent analysis of fecal microbiota in infants, with special reference to *Bifidobacterium* species. *FEMS Microbiol Lett* (2005) 243(2):417–23. doi:10.1016/j.femsle.2005.01.002
 133. Turroni F, Foroni E, Pizzetti P, Giubellini V, Ribbera A, Merusi P, et al. Exploring the diversity of the bifidobacterial population in the human intestinal tract. *Appl Environ Microbiol* (2009) 75(6):1534–45. doi:10.1128/AEM.02216-08
 134. Roger LC, Costabile A, Holland DT, Hoyle L, McCartney AL. Examination of faecal *Bifidobacterium* populations in breast- and formula-fed infants during the first 18 months of life. *Microbiology* (2010) 156(Pt 11):3329–41. doi:10.1099/mic.0.043224-0
 135. Underwood MA, Kalanetra KM, Bokulich NA, Lewis ZT, Mirmiran M, Tancredi DJ, et al. A comparison of two probiotic strains of bifidobacteria in premature infants. *J Pediatr* (2013) 163(6):1585–1591.e9. doi:10.1016/j.jpeds.2013.07.017
 136. Szajewska H, Guandalini S, Morelli L, Van Goudoever JB, Walker A. Effect of *Bifidobacterium animalis* subsp *lactis* supplementation in preterm infants: a systematic review of randomized controlled trials. *J Pediatr Gastroenterol Nutr* (2010) 51(2):203–9. doi:10.1097/MPG.0b013e3181dc0d93
 137. Costeloe K, Hardy P, Juszczak E, Wilks M, Millar MR; Probiotics in Preterm Infants Study Collaborative Group. *Bifidobacterium breve* BBG-001 in

- very preterm infants: a randomised controlled phase 3 trial. *Lancet* (2016) 387(10019):649–60. doi:10.1016/S0140-6736(15)01027-2
138. Bin-Nun A, Bromiker R, Wilschanski M, Kaplan M, Rudensky B, Caplan M, et al. Oral probiotics prevent necrotizing enterocolitis in very low birthweight neonates. *J Pediatr* (2005) 147(2):192–6. doi:10.1016/j.jpeds.2005.03.054
 139. Lin HC, Su BH, Chen AC, Lin TW, Tsai CH, Yeh TF, et al. Oral probiotics reduce the incidence and severity of necrotizing enterocolitis in very low birth weight infants. *Pediatrics* (2005) 115(1):1–4. doi:10.1542/peds.2004-1463
 140. Samanta M, Sarkar M, Ghosh P, Ghosh J, Sinha M, Chatterjee S. Prophylactic probiotics for prevention of necrotizing enterocolitis in very low birth weight newborns. *J Trop Pediatr* (2009) 55(2):128–31. doi:10.1093/tropej/fmn091
 141. Underwood MA, Salzman NH, Bennett SH, Barman M, Mills DA, Marcobal A, et al. A randomized placebo-controlled comparison of 2 prebiotic/probiotic combinations in preterm infants: impact on weight gain, intestinal microbiota, and fecal short-chain fatty acids. *J Pediatr Gastroenterol Nutr* (2009) 48(2):216–25. doi:10.1097/MPG.0b013e31818de195
 142. Fernandez-Carrocera LA, Solis-Herrera A, Cabanillas-Ayon M, Gallardo-Sarmiento RB, Garcia-Perez CS, Montano-Rodriguez R, et al. Double-blind, randomised clinical assay to evaluate the efficacy of probiotics in preterm newborns weighing less than 1500 g in the prevention of necrotising enterocolitis. *Arch Dis Child Fetal Neonatal Ed* (2013) 98(1):F5–9. doi:10.1136/archdischild-2011-300435
 143. Jacobs SE, Tobin JM, Opie GF, Donath S, Tabrizi SN, Pirota M, et al. Probiotic effects on late-onset sepsis in very preterm infants: a randomized controlled trial. *Pediatrics* (2013) 132(6):1055–62. doi:10.1542/peds.2013-1339
 144. Macfarlane GT, Macfarlane S. Bacteria, colonic fermentation, and gastrointestinal health. *JAOAC Int* (2012) 95(1):50–60. doi:10.5740/jaoacint.SGE_Macfarlane
 145. Louis P, Scott KP, Duncan SH, Flint HJ. Understanding the effects of diet on bacterial metabolism in the large intestine. *J Appl Microbiol* (2007) 102(5):1197–208. doi:10.1111/j.1365-2672.2007.03322.x
 146. Fukuda S, Toh H, Hase K, Oshima K, Nakanishi Y, Yoshimura K, et al. Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* (2011) 469(7331):543–7. doi:10.1038/nature09646
 147. Fukuda S, Toh H, Taylor TD, Ohno H, Hattori M. Acetate-producing bifidobacteria protect the host from enteropathogenic infection via carbohydrate transporters. *Gut Microbes* (2012) 3(5):449–54. doi:10.4161/gmic.21214
 148. Duncan SH, Louis P, Flint HJ. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* (2004) 70(10):5810–7. doi:10.1128/AEM.70.10.5810-5817.2004
 149. Belenguer A, Duncan SH, Calder AG, Holtrop G, Louis P, Lobley GE, et al. Two routes of metabolic cross-feeding between *Bifidobacterium adolescentis* and butyrate-producing anaerobes from the human gut. *Appl Environ Microbiol* (2006) 72(5):3593–9. doi:10.1128/AEM.72.5.3593-3599.2006
 150. Falony G, Vlachou A, Verbrugghe K, De Vuyst L. Cross-feeding between *Bifidobacterium longum* BB536 and acetate-converting, butyrate-producing colon bacteria during growth on oligofructose. *Appl Environ Microbiol* (2006) 72(12):7835–41. doi:10.1128/AEM.01296-06
 151. Flint HJ, Duncan SH, Scott KP, Louis P. Links between diet, gut microbiota composition and gut metabolism. *Proc Nutr Soc* (2015) 74(1):13–22. doi:10.1017/S0029665114001463
 152. Duncan SH, Barcenilla A, Stewart CS, Pryde SE, Flint HJ. Acetate utilization and butyryl coenzyme A (CoA):acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microbiol* (2002) 68(10):5186–90. doi:10.1128/AEM.68.10.5186-5190.2002
 153. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol* (2014) 12(10):661–72. doi:10.1038/nrmicro3344
 154. Ogawa K, Ben RA, Pons S, de Paolo MI, Bustos Fernandez L. Volatile fatty acids, lactic acid, and pH in the stools of breast-fed and bottle-fed infants. *J Pediatr Gastroenterol Nutr* (1992) 15(3):248–52. doi:10.1097/00005176-199210000-00004
 155. Butel M-J, Suau A, Campeotto F, Magne F, Aires J, Ferraris L, et al. Conditions of bifidobacterial colonization in preterm infants: a prospective analysis. *J Pediatr Gastroenterol Nutr* (2007) 44(5):577–82. doi:10.1097/MPG.0b013e3180406b20
 156. Abdulkadir B, Nelson A, Skeath T, Marrs EC, Perry JD, Cummings SP, et al. Routine use of probiotics in preterm infants: longitudinal impact on the microbiome and metabolome. *Neonatology* (2016) 109(4):239–47. doi:10.1159/000442936
 157. Fanning S, Hall LJ, Cronin M, Zomer A, MacSharry J, Goulding D, et al. Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. *Proc Natl Acad Sci U S A* (2012) 109(6):2108–13. doi:10.1073/pnas.1115621109
 158. Fanning S, Hall LJ, van Sinderen D. *Bifidobacterium breve* UCC2003 surface exopolysaccharide production is a beneficial trait mediating commensal-host interaction through immune modulation and pathogen protection. *Gut Microbes* (2012) 3(5):420–5. doi:10.4161/gmic.20630
 159. Marcobal A, Barboza M, Sonnenburg ED, Pudlo N, Martens EC, Desai P, et al. *Bacteroides* in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* (2011) 10(5):507–14. doi:10.1016/j.chom.2011.10.007
 160. Sonnenburg JL, Xu J, Leip DD, Chen CH, Westover BP, Weatherford J, et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* (2005) 307(5717):1955–9. doi:10.1126/science.1109051
 161. Stappenbeck TS, Hooper LV, Gordon JI. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A* (2002) 99(24):15451–5. doi:10.1073/pnas.202604299
 162. Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. *Science* (2001) 292(5519):1115–8. doi:10.1126/science.1058709
 163. Round JL, Mazmanian SK. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A* (2010) 107(27):12204–9. doi:10.1073/pnas.0909122107
 164. Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* (2005) 122(1):107–18. doi:10.1016/j.cell.2005.05.007
 165. Telesford KM, Yan W, Ochoa-Reparaz J, Pant A, Kircher C, Christy MA, et al. A commensal symbiotic factor derived from *Bacteroides fragilis* promotes human CD39+ Foxp3+ T cells and Treg function. *Gut Microbes* (2015) 6(4):234–42. doi:10.1080/19490976.2015.1056973
 166. Sjogren YM, Tomicic S, Lundberg A, Botcher MF, Bjorksten B, verremark-Ekstrom ES, et al. Influence of early gut microbiota on the maturation of childhood mucosal and systemic immune responses. *Clin Exp Allergy* (2009) 39(12):1842–51. doi:10.1111/j.1365-2222.2009.03326.x
 167. MacArthur R. Fluctuations of animal populations and a measure of community stability. *Ecology* (1955) 36(3):533–6.
 168. Elton CC. The reasons for conservation. *The Ecology of Invasions by Animals and Plants*. Netherlands, London, UK: Springer; Methuen and Co. (1958). p. 143–53. doi:10.1007/978-94-009-5851-7
 169. May RM. *Stability and Complexity in Model Ecosystems*. Princeton: Princeton University Press (1973).
 170. Tilman D. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* (1999) 80(5):1455–74. doi:10.2307/176540
 171. Arthur WB, Durlauf SN, Lane DA. *The Economy As an Evolving Complex System II*. Reading, MA: Addison-Wesley (1997).
 172. Kernbauer E, Cadwell K. Autophagy, viruses, and intestinal immunity. *Curr Opin Gastroenterol* (2014) 30(6):539–46. doi:10.1097/MOG.0000000000000121
 173. Virgin HW. The virome in mammalian physiology and disease. *Cell* (2014) 157(1):142–50. doi:10.1016/j.cell.2014.02.032
 174. Cadwell K. The virome in host health and disease. *Immunity* (2015) 42(5):805–13. doi:10.1016/j.immuni.2015.05.003

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

Copyright © 2017 Laforest-Lapointe and Arrieta. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The *SpTransformer* Gene Family (Formerly *Sp185/333*) in the Purple Sea Urchin and the Functional Diversity of the Anti-Pathogen rSpTransformer-E1 Protein

L. Courtney Smith* and Cheng Man Lun†

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

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Coenraad Adema,
University of New Mexico,
United States
Gerardo R. Vasta,
University of Maryland,
Baltimore, United States

*Correspondence:

L. Courtney Smith
csmith@gwu.edu

†Present address:

Cheng Man Lun,
HIV Dynamics and Replication
Program, Virus-Cell Interaction
Section, Center for Cancer Research,
National Cancer Institute, Frederick,
MD, United States

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 10 April 2017

Accepted: 08 June 2017

Published: 30 June 2017

Citation:

Smith LC and Lun CM (2017) The
SpTransformer Gene Family (Formerly
Sp185/333) in the Purple Sea Urchin
and the Functional Diversity of the
Anti-Pathogen rSpTransformer-E1
Protein.
Front. Immunol. 8:725.
doi: 10.3389/fimmu.2017.00725

The complex innate immune system of sea urchins is underpinned by several multigene families including the *SpTransformer* family (*SpTrf*; formerly *Sp185/333*) with estimates of ~50 members, although the family size is likely variable among individuals of *Strongylocentrotus purpuratus*. The genes are small with similar structure, are tightly clustered, and have several types of repeats in the second of two exons and that surround each gene. The density of repeats suggests that the genes are positioned within regions of genomic instability, which may be required to drive sequence diversification. The second exon encodes the mature protein and is composed of blocks of sequence called elements that are present in mosaics of defined element patterns and are the major source of sequence diversity. The *SpTrf* genes respond swiftly to immune challenge, but only a single gene is expressed per phagocyte. Many of the mRNAs appear to be edited and encode proteins with altered and/or missense sequence that are often truncated, of which some may be functional. The standard *SpTrf* protein structure is an N-terminal glycine-rich region, a central RGD motif, a histidine-rich region, and a C-terminal region. Function is predicted from a recombinant protein, rSpTransformer-E1 (rSpTrf-E1), which binds to *Vibrio* and *Saccharomyces*, but not to *Bacillus*, and binds tightly to lipopolysaccharide, β -1,3-glucan, and flagellin, but not to peptidoglycan. rSpTrf-E1 is intrinsically disordered but transforms to α helical structure in the presence of binding targets including lipopolysaccharide, which may underpin the characteristics of binding to multiple targets. *SpTrf* proteins associate with coelomocyte membranes, and rSpTrf-E1 binds specifically to phosphatidic acid (PA). When rSpTrf-E1 is bound to PA in liposome membranes, it induces morphological changes in liposomes that correlate with PA clustering and leakage of luminal contents, and it extracts or removes PA from the bilayer. The multitasking activities of rSpTrf-E1 infer multiple and perhaps overlapping activities for the hundreds of native *SpTrf* proteins that are produced by individual sea urchins. This likely generates a flexible and highly protective immune system for the sea urchin in its marine habitat that it shares with broad arrays of microbes that may be pathogens and opportunists.

Keywords: *Sp185/333*, multitasking, anti-pathogen, purple sea urchin, *Strongylocentrotus*, echinoderm, invertebrate, intrinsically disordered proteins

INTRODUCTION

Immune activities in animals that survive the arrays of pathogens with which they share their habitats, display a wide range of innate functions irrespective of whether they also deploy adaptive immunity. The underlying attributes of many genes that act in pathogen detection or anti-pathogen responses typically show significant sequence diversity in the encoded proteins that can be derived from gene diversification mechanisms, mRNA processing that may include posttranscriptional changes, and posttranslational modifications to the proteins. Single copy genes that function in immunity can also display significant sequence diversity through large numbers of alleles in a population. Some examples are genes linked in the fusion/histocompatibility locus in the compound tunicate *Botryllus schlosseri*, and genes in the major histocompatibility locus in mammals and other vertebrates [reviewed in Ref. (1)]. However, many of the genes that encode innate immune functions are expanded into families such as Toll-like receptors and NOD-like receptors in most animals, fibrinogen-related proteins in mollusks, and killer immunoglobulin-like receptors in mammals. Common attributes of immune gene family members include clustering, shared sequences, repeats, plus elevated levels of duplications, deletions, and recombination (2). These attributes typically generate pseudogenes, but also generate new genes that can be expressed and are then subject to selection based on increased host fitness in responses to and protection from pathogens. A gene family with these attributes in the purple sea urchin, *Strongylocentrotus purpuratus*, is the *Sp185/333* gene family, which will be the focus of this review. A recombinant (r)Sp185/333 protein shows multitasking activities with characteristics for binding to different types of microbes and multiple pathogen-associated molecular patterns (PAMPs) (3), and transforms from intrinsic disorder to α helical structure upon binding a target (4, 5). These attributes underlie the new name for this particular recombinant protein from rSp0032, which was based on the cDNA nomenclature as reported by Terwilliger et al. (6), to rSpTransformer-E1 (rSpTrf-E1) that is based on a combination of its structural conformational changes and its E1 element pattern (4). In keeping with maintaining continuity between names for genes and their encoded proteins, the gene family has also been renamed from *Sp185/333* to *SpTransformer* (*SpTrf*) and the general name for the proteins have been changed from *Sp185/333* proteins to *SpTrf*

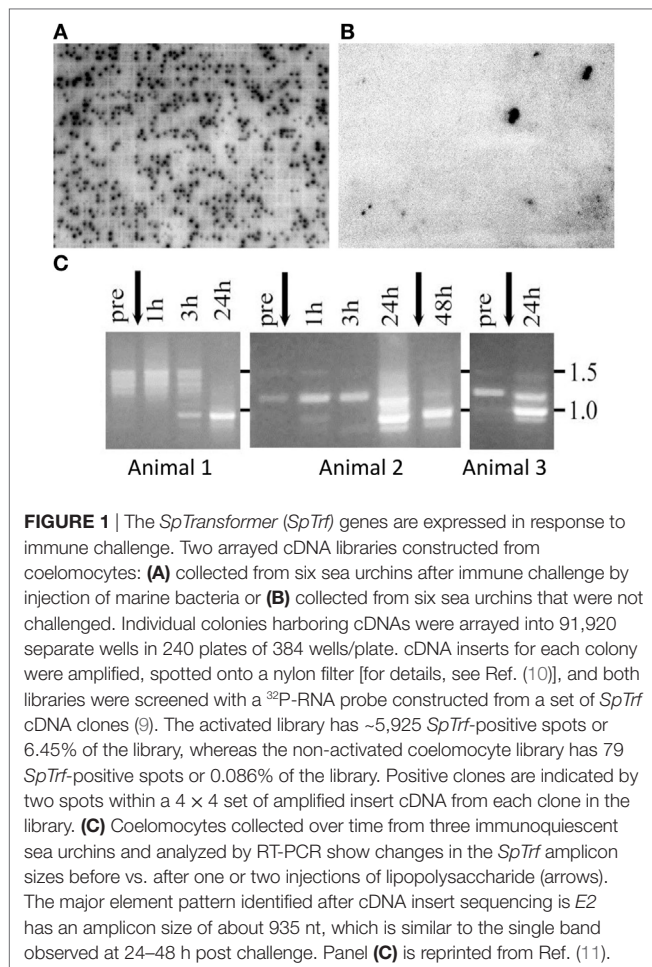
proteins. These updated names will be used in this review and in all future reports on this system.

DISCOVERY; *SpTrf* GENE EXPRESSION AND SEQUENCE DIVERSITY OF THE mRNAs

The first reports of *SpTrf* sequences included an expressed sequence tag (EST; equivalent of an RNA-Seq read) from a cDNA library constructed from coelomocytes after challenge with lipopolysaccharide (LPS) (7) and a full-length coelomocyte cDNA sequence identified after challenge with marine bacteria and injury based on results from analysis by differential display (8). Both were noteworthy because of significant upregulated gene expression in coelomocytes in response to immune challenge. When an arrayed cDNA library constructed from immune activated coelomocytes was screened with a subtracted probe specific for mRNAs in LPS-activated coelomocytes, clones identified in the library indicated a striking upregulation in gene expression of these same sequences, and which constituted ~60% of the sequenced clones (9). The names of the original EST and differential display clones, 333 and 185, were used in the original name of the gene family and collection of cDNAs because the deduced protein sequences did not match to any proteins in any other organism and offered no prediction for function. Upon re-screening the arrayed cDNA libraries for clones with *SpTrf* sequences, positive clones constituted 6.45% of the library constructed from bacteria challenged coelomocytes and 0.086% of the non-activated library (Figures 1A,B). This 75-fold increase in gene expression in response to challenge correlates with results from the original Northern blots (8). Comparisons among the cDNA sequences show significant and intriguing sequence diversity that, in addition to the gene expression characteristics, was the basis for additional investigations.

Sea urchins in their normal marine habitat are in constant contact with microbes in the water, on the substrate, and associated with their diet, and healthy animals maintain a constant level of immune activity. However, this immune activity complicates experimental evaluation of immune responsiveness of sea urchins to a particular PAMP or microbe. This problem was resolved by the discovery that when sea urchins are kept in closed, recirculating marine aquaria for more than 6- to 8-months and away from the input of "wild" sea water, they turn down their immune responsiveness and, therefore, have been called immunoquiescent (IQ) (12). Examples of downregulated gene expression in IQ animals include the complement homolog, *SpC3* (12–14), and the *Sp056* gene that encodes the small C-type lectin, *SpEchinoidin* (11). Consequently, when IQ sea urchins are immune challenged to determine activators of the *SpTrf* genes, expression is induced with one or two injections of LPS (Figure 1C), β -1,3-glucan (a fungal PAMP), double stranded (ds)RNA (polyGC to represent a viral challenge), or injury that includes injection of buffer (11). Prior to challenge or injury in IQ sea urchins, *SpTrf* amplicons are either absent or show a spread of weak bands of about 1.2–1.5 kB (Figure 1C). After challenge, an increase in the intensity of the amplicons is noted and the amplicon sizes change differently

Abbreviations: 2D, two dimensional; BAC, bacterial artificial chromosome; CD, circular dichroism; CF, coelomic fluid; aCF, artificial CF; dextran-488, dextran labeled with Alexa Fluor® 488; dsRNA, double stranded RNA; EST, expressed sequence tag; gDNA, genomic DNA; HeTrf, the transformer family from *Heliocidaris erythrogramma*; IDP, intrinsically disordered protein; IQ, immunoquiescent; PA, phosphatidic acid; PC, phosphatidylcholine; r, recombinant; rC-Gly, recombinant C-terminal end of the glycine-rich region; RGD, arginine, glycine, aspartic acid motif; rGly-rich, recombinant glycine-rich fragment; rHis-rich, recombinant histidine-rich fragment; rSpTrf-E1, recombinant *SpTrf* protein with an E1 element pattern; SDS, sodium dodecyl sulfate; *SpRAG1L*, recombinase activating gene 1-like homolog from *Strongylocentrotus purpuratus*; *SpRAG2L*, recombinase activating gene 2-like homolog from *Strongylocentrotus purpuratus*; *SpTrf*, the transformer family from *Strongylocentrotus purpuratus*; STRs, short tandem repeats; TFE, 2,2,2-trifluoroethanol.



among individual animals but tend to focus on a single major size of ~0.9 kb. This indicates a change from diverse or no expression in non-challenged IQ sea urchins to a focus on a major band that likely corresponds with cDNAs of similar size that are the most common version of the *SpTrf* cDNA sequences (see below).

Automated alignments of the *SpTrf* cDNA sequences fail when using standard alignment programs with default parameters, which forced alignments to be done manually. Challenges for generating alignments are due to the unusual characteristic of the *SpTrf* sequences in which insertions of large artificial gaps are required for optimal alignments. These gaps identify and define recognizable blocks of sequence called *elements* (Figure 2A) (6, 11). The initial alignments were based on the cDNA sequences and identified a maximum of 25 elements, of which, subsets of elements are present as mosaics in individual sequences; no sequences have the full complement of possible elements. Different mosaics of elements are repeatedly identified and are termed *element patterns* and correlate with the sequence variants of element 15. This highly diverse element is present in a range of sizes and is employed as the basis for naming the element patterns of A through G (Figure 2A). Some sequences do not include element 15 and are termed *0* patterns. Other attributes of the cDNA sequences include repeats identified as

tandem type 1 repeats, interspersed repeats of types 2–5, and one to three possible stop codons in element 25 defined as element 25a, b, or c. The swift upregulation of the *SpTrf* genes in response to immune challenge and the striking sequence diversity of the cDNAs strongly suggest that this family has important activities in the sea urchin immune response.

Ongoing and repeated searches of sequence repositories have only identified *Trf* sequences in other euechinoids. In phylogenetic analyses of the euechinoid order within the echinoid class of echinoderms, it clusters separately from the cidaroid order, which is more ancient [for details on echinoderm phylogeny, see Ref. (17, 18)]. Searches of the genome sequences from the euechinoid sea urchins, *Mesocentrotus franciscanus*, *Strongylocentrotus fragilis* [see (19) for genus revisions in the strongylocentrotid sea urchins], and *Lytechinus variagatus* identify matches to *Trf* genes. A single cDNA sequence has been reported for *Strongylocentrotus intermedius* (20), and 39 *HeTrf* (formerly *He185/333*) gene sequences have been characterized from *Heliocidaris erythrogramma*, another sea urchin species (21). However, searches of the genome sequence of the pencil sea urchin, *Eucidaris tribuloides*, in addition to other cidaroid species and other classes of echinoderms show no matches to *Trf* genes. Given the outcomes of these searches, the *Trf* gene family appears to be a derived character of innate immunity that is present only within the regular euechinoid sea urchins.

THE *SpTrf* GENES ARE SMALL, ARRANGED IN TIGHT CLUSTERS, AND HAVE SHARED BUT DIVERSE SEQUENCES

Alignments that demonstrate the interesting element-based *SpTrf* cDNA sequence structure is superficially consistent with and suggestive of extensive alternative splicing similar to that documented for *DSCAM* (22). However, when genomic DNA (gDNA) from three sea urchins is digested with restriction enzymes, used in Southern blots, and analyzed with probes from the 5' and 3' ends of cDNA templates, both probes hybridize to bands of 1.5–2 kb, which are similar in size to the mRNA sequences (Figure 3A) (6). This prediction of a small gene size does not fit with the *DSCAM* gene structure of ~100 exons and correlates with results from a search of the initial assembly of the sea urchin genome sequence (9/2003) that shows *SpTrf* genes of less than 2 kb with two exons (Figure 3B). Alternative splicing to generate the cDNA sequence diversity is impossible for two exons, and no cryptic splice sites are present in the genes that might generate unexpected splicing patterns (23). Because the *SpTrf* genes are small, they could be amplified by PCR from gDNA and sequenced, and all show the same basic structure of two exons (15). Comparisons among 121 genes of unique sequence (of 171 sequenced gene amplicons) show significant sequence diversity. Although the first exon encodes a relatively conserved hydrophobic leader, the second exon is highly diverse with regard to both size and sequence and encodes the mature protein with mosaic element patterns corresponding to those characterized in the cDNAs (Figure 2A) (15). When the coding regions of the genes

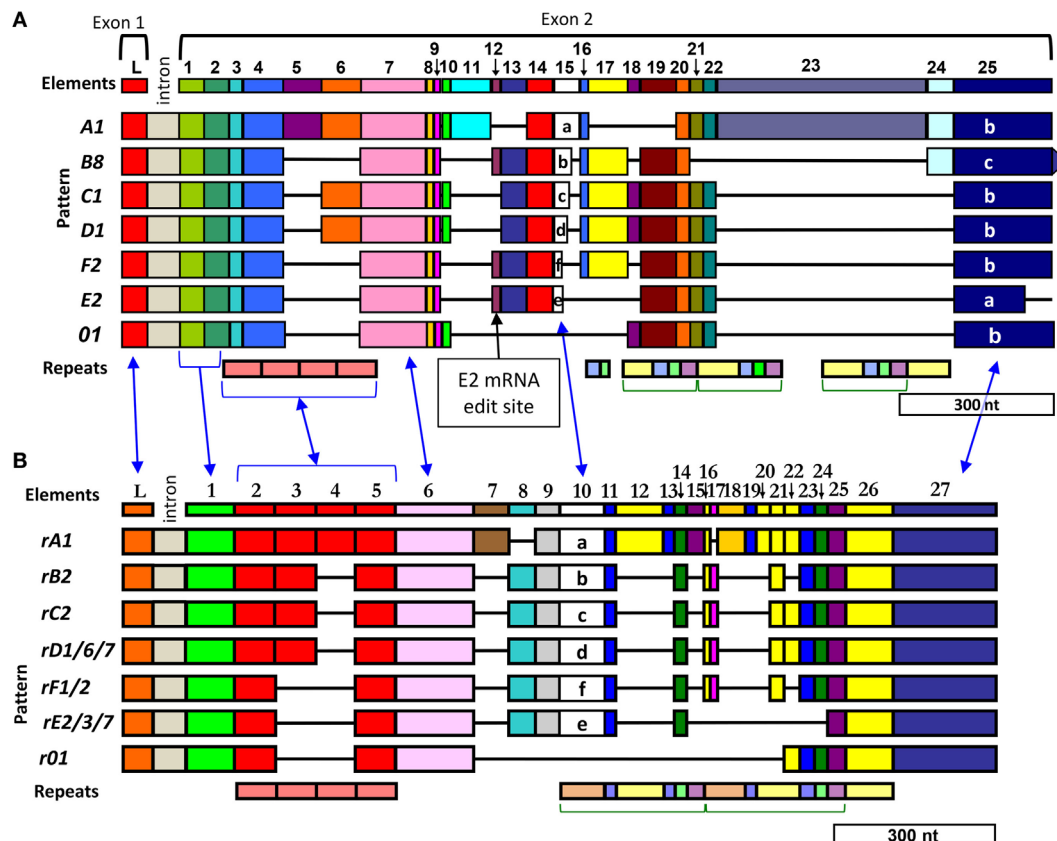


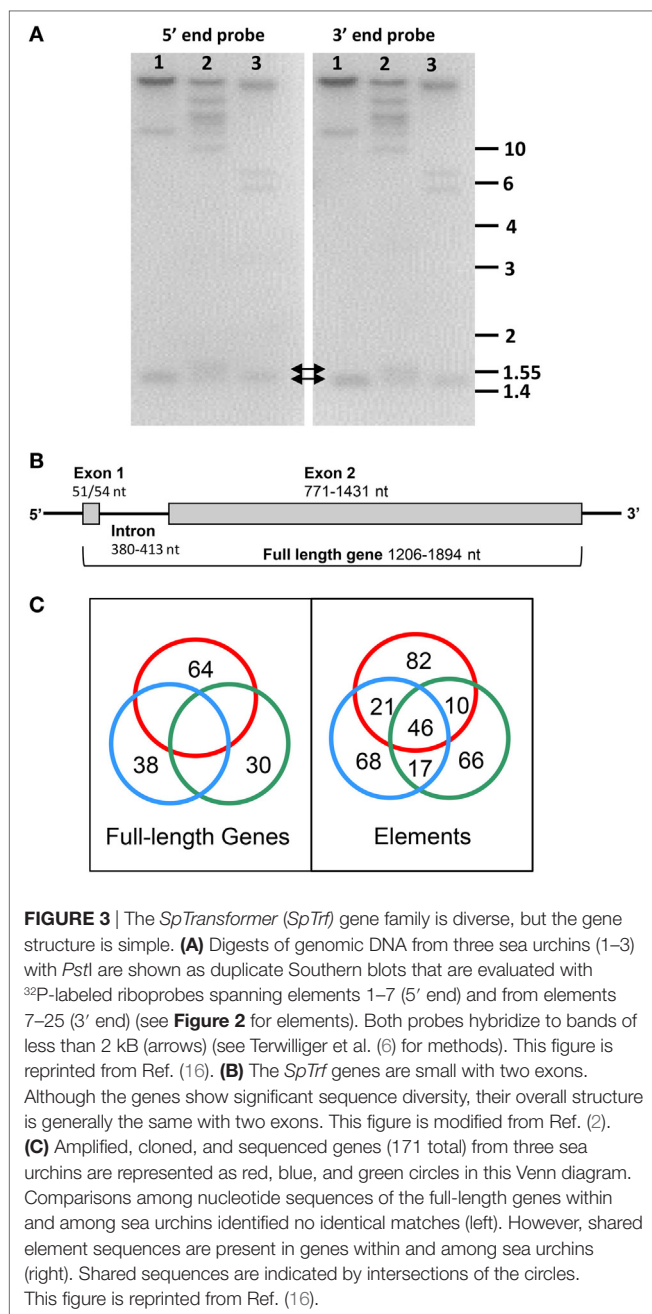
FIGURE 2 | Diversity of elements and repeats in the *SpTransformer* (*SpTrf*) sequences enable two equally optimal alignments. Optimal alignments of the *SpTrf* sequences require the insertion of artificial gaps (black horizontal lines) that delineate individual elements shown as colored blocks [numbered across the top (L, leader)]. The intron is not to scale. Different element patterns are based on the variable presence or absence of elements and are listed to the left of each alignment. The combination of elements defines the element pattern that is named according to the distinctive and highly diverse sequence of element 15 (6). **(A)** The initial alignment employed fragments of cDNA sequences, expressed sequence tags, and full-length cDNA sequences according to Terwilliger et al. (6, 11) and is called a “cDNA-based” alignment. Element 25 in panel **(A)** is listed as three types, a, b, and c, which are defined by the location of the first of three possible stop codons in the element. These three stops are also present in element 27 in panel **(B)** but are not indicated. A very common mRNA editing site in element 12 is indicated for the *E2* element pattern (indicated) that changes a codon to a stop to encode a truncated *E2.1* sequence that omits the histidine-rich region of the protein (see also **Figure 4B**). **(B)** A subsequent alignment optimizes the edges of the elements with the edges of the repeats according to Buckley and Smith (15) and is called the “repeat-based” alignment. Both types of alignments are feasible with cDNA and gene sequences. The blue arrows between the two alignments indicate identical regions. Repeats are shown at the bottom of each alignment, occur as tandem repeats or interspersed tandem repeats, and are denoted by different colors (type 1, red; type 2, blue; type 3, green; type 4, yellow; type 5, purple; type 6, peach). This figure is modified from Ref. (16).

are aligned using the cDNA-based alignment parameters according to Terwilliger et al. (6), the first four elements in the second exon are not defined by the insertion of artificial gaps (**Figure 2A**). Furthermore, the edges of the elements and the edges of the repeats do not correspond. Consequently, an alternative alignment that matches the edges of the repeats with the edges of the elements, where possible, resulted in the “repeat based” alignment for both genes and cDNAs (**Figure 2B**). The repeat-based alignment collapses some repeats, identifies the type 6 repeat, and increases the number of possible elements to 27 although it shortens the overall length of the alignment. As expected, the intron sequences are more diverse than the exons, although comparisons among the introns suggest five types that are usually, but not always, associated with a specific element pattern in the second exon (15). Alignments of the genes reveal several surprising results besides the presence of elements and repeats. Comparisons among gene sequences from

different sea urchins show that no full-length gene sequence is shared among animals, but that sequences of individual elements, which have different sequence variants, can be shared among genes from individual animals and among different animals (**Figure 3C**). The *SpTrf* genes are unique and highly unusual based on their significant sequence diversity that is derived from the element-based structure of the second exon in addition to sequence variations in many of the elements.

HeTrf GENES ARE ALSO STRUCTURED WITH ELEMENTS

The sea urchin *H. erythrogramma* is local to Australia and the southern hemisphere and is morphologically similar to *S. purpuratus*. They are about the same size, are generally purple, and have similar types of coelomocytes in the coelomic fluid (CF) (24).



Although, their life histories are quite different—*S. purpuratus* is an indirect developer with larvae that feed in the zooplankton prior to undergoing metamorphosis to a juvenile sea urchin, whereas *H. erythrogramma* skips the larval stage and develops directly from an embryo to a juvenile—both species have *Trf* gene families (21). The *HeTrf* cDNA sequences are 68–74% identical to the *SpTrf* cDNA sequences, tend to be shorter, and have 31 elements arranged into 29 different element patterns that are different from those in the *SpTrf* cDNAs and genes. The *HeTrf* genes also have two exons, although the intron has large variations in length. There are four types of imperfect tandem and interspersed repeats that are similar to four of the six repeats

in the *SpTrf* sequences, although the copy numbers and positions of the repeats within the genes are different. Codons under positive selection for diversification [for methods, see references in Ref. (21)] are positioned throughout the sequences for the *HeTrf* genes but tend to be located within the first 200 codons in the *SpTrf* genes. These two *Trf* gene families are clearly homologous but the two families separate into different clades in phylogenetic analyses suggesting diverging evolutionary histories likely based on different sets of pathogens that the two species face not only as adults but also during the larval phase of *S. purpuratus*, which is absent in *H. erythrogramma*.

EVOLUTIONARY HISTORY OF THE *SpTrf* GENES ESTIMATED FROM THE TYPE 1 REPEAT DIVERSITY

The varieties of repeats in the *SpTrf* genes are a notable and unusual attribute of the second exon. The five types of interspersed repeats positioned toward the 3' end of the second exon are present in complex patterns that are repeated two or three times depending on the alignment (**Figure 2**) (25). The tandem type 1 repeats that are present in two to four copies are positioned toward the 5' end of the second exon and show imperfect sequence matches in addition to mosaic patterns that vary among genes (**Figure 2**). A computational evaluation of the type 1 repeats and their phylogenetic clustering into four clades demonstrated that clade membership correlates with their position in the second exon and defines the correct position of the repeats when two or three are present rather than four (25). When two type 1 repeats are present in a gene, they are always the first and fourth repeat, and when three repeats are present, they are always the first, second, and fourth repeat (**Figure 2B**). Sequence variations among the type 1 repeats may be the outcome of duplication, deletion, and recombination of two theoretical ancestral type 1 repeat sequences that are based on a computational prediction from extant sequences. This led to questions of whether recombination hot spots could be identified within the genes, which was underpinned by observations that sequences of adjacent regions did not match among different genes. For example, these included (i) the sequence of the 5' UTR relative to the adjoining first exon, (ii) the sequence of the 5' end vs. the 3' end of the genes, and (iii) the 5' vs. 3' ends of some elements irrespective of whether they correspond to repeats (25). Predictions strongly suggest significant recombination between the two ends of the second exon, between adjacent elements, and within larger elements, with no clear hot spots of recombination (**Table 1**). Furthermore, the frequency of predicted recombination within the second exon is similar to results for the well-known somatic recombination that occurs among the variable and joining segments of the T cell receptor and is very different from the lack of recombination between the two ends of the sea urchin histone *H3* gene. Molecular clock analysis of the *SpTrf* genes indicates that the genes are young (26) and about the same age as the species (27), which is in agreement with the generally accepted concept that immune genes encoding proteins that interact with the environment and/or pathogens are under pressure to diversify and show swift evolution [reviewed

TABLE 1 | Recombination is predicted throughout the *SpTrf* gene sequences.^a

	L	IntA	IntB	1	6	6a ^b	6b	26	27	27a	27b	H3.1 ^c	H3.2	TcRV ^d	TcRJ
L		***	***	**	**	□	□	***	***	□	□	□	□	□	□
IntA	**		***	**	***	□	□	***	***	□	□	□	□	□	□
IntB	**	**		ns	***	□	□	***	***	□	□	□	□	□	□
1	**	**	*		***	□	□	***	***	□	□	□	□	□	□
6	*	**	**	**		□	□	***	***	□	□	□	□	□	□
6a	□	□	□	□	□		*	□	□	□	□	□	□	□	□
6b	□	□	□	□	□	*		□	□	□	□	□	□	□	□
26	**	**	**	**	***	□	□		***	□	□	□	□	□	□
27	**	**	**	**	***	□	□	*		□	□	□	□	□	□
27a	□	□	□	□	□	□	□	□	□		*	□	□	□	□
27b	□	□	□	□	□	□	□	□	□	**		□	□	□	□
H3.1	□	□	□	□	□	□	□	□	□	□	□	ns		□	□
H3.2	□	□	□	□	□	□	□	□	□	□	□	ns		□	□
TcRV	□	□	□	□	□	□	□	□	□	□	□	□	□	**	***
TcRJ	□	□	□	□	□	□	□	□	□	□	□	□	□		

^aRecombination within and among elements is based on results from the incongruence length difference test shown above the diagonal and the incongruence permutation test shown below. This table is modified from Ref. (25).

^bFor some individual elements, comparisons are carried out between the 5' end (a) and the 3' end (b). Element numbers are based on the repeat-based alignment (see Figure 2B).

^cEvaluation of the sea urchin histone H3 gene sequence employed two regions of the gene sequence, H3.1 and H3.2, which were of similar size to the average element size in the *SpTrf* (*SpTrf*) sequences.

^dEvaluation of the TcR employed the variable region (TcRV) and the joining region (TcRJ) (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; ns, not significant; L, leader; Int, intron; □, not done).

in Ref. (1, 2)]. The occurrence of recombination throughout the *SpTrf* gene family is likely to be much greater than suggested by shared and unshared element sequences and may perhaps be driven by the clustered nature of the genes (2, 28, 29) (see below).

THE *SpTrf* mRNAs ARE LIKELY EDITED

A surprising result from the *SpTrf* cDNA sequences reported by Terwilliger et al. (11) is that only about half (306 of 608) encode full-length proteins, whereas the rest have frameshifts leading to missense sequence and early stop codons or have a single nucleotide change that inserts an early stop codon at a particular position in element 13 (Figure 2A). Similarly, point mutations, indels, missense sequence, and early stop codons are also present in about 10% (11 of 112) of the *HeTrf* cDNA sequences from *H. erythrogramma* (21). In striking comparison, all but one of the 198 *SpTrf* gene sequences [171 amplified from gDNA from three sea urchins, 12 amplified from clones in the small insert bacterial artificial chromosome (BAC) library (15), and 15 assembled from BAC inserts (29)] have perfect open reading frames. The unusual difference of perfect vs. altered reading frames in the genes vs. the cDNAs, respectively, is an outcome of comparisons between genes and cDNAs from individual animals (30). Very few of the genes match identically to the cDNA sequences from individual sea urchins, but more noteworthy are the differences between the sequences of the genes and cDNAs of the same element pattern. The comparison shows that 30% of the nucleotide differences are a cytidine in the gene and a uracil at the same position in the cDNA, which is consistent with cytidine deaminase activity (30). Other changes in the cDNAs, such as the indels, may be the outcome of low fidelity RNA polymerases, such as polymerase μ . Genes encoding several cytidine deaminases plus polymerase μ are present in the sea urchin genome sequence (31). These results suggest editing of the *SpTrf* mRNAs,

which, although quite unexpected, could have the disadvantage of yielding transcripts that encode non-functional proteins, but also the advantage of expanding the diversity of the proteins produced in response to immune challenge irrespective of whether the editing may be random, directed, or both.

The identification of RNA editing of both *SpTrf* and *HeTrf* transcripts resulting in indels and frameshifts led to an initial assumption that these mRNAs would be recycled and not transcribed. However, predicted missense sequences from edited cDNA sequences with frame shifts are present in the *SpTrf* proteins isolated from the CF, indicating that the edited mRNAs are translated (32). This is noteworthy because the frequency of edited vs. non-edited mRNAs changes relative to immune challenge. Edited *SpTrf* mRNAs encoding truncated proteins including some with missense sequence tend to be present more often in coelomocytes from IQ sea urchins prior to immune challenge, whereas mRNAs that are not edited and encode full-length proteins tend to increase in coelomocytes responding to immune challenge (Figure 4A) (11, 33). This change is detected in many sequence versions of the cDNAs but is most easily identified in those that encode the E2 element pattern (494 of 608 cDNAs) of which 57% have a nucleotide change in element 13 that changes a glycine codon to an early stop (Figure 2A). This single edit results in truncated proteins that are missing the histidine-rich region and are defined as the E2.1 element pattern (11). Edits to the E2 mRNA can also insert indels that induce frameshifts, such as the E2.4 sequence that has missense sequence and an early stop (Figure 4B). An alignment of deduced protein sequences with the E2 element pattern illustrates the position of the common RNA-editing event that produces the E2.1 truncated protein (Figure 4B). RNA editing that deletes the histidine-rich region of proteins is consistent with difficulties in isolating many *SpTrf* proteins by nickel affinity prior to challenge (33). Speculation on the underlying basis for the change in editing relative to an

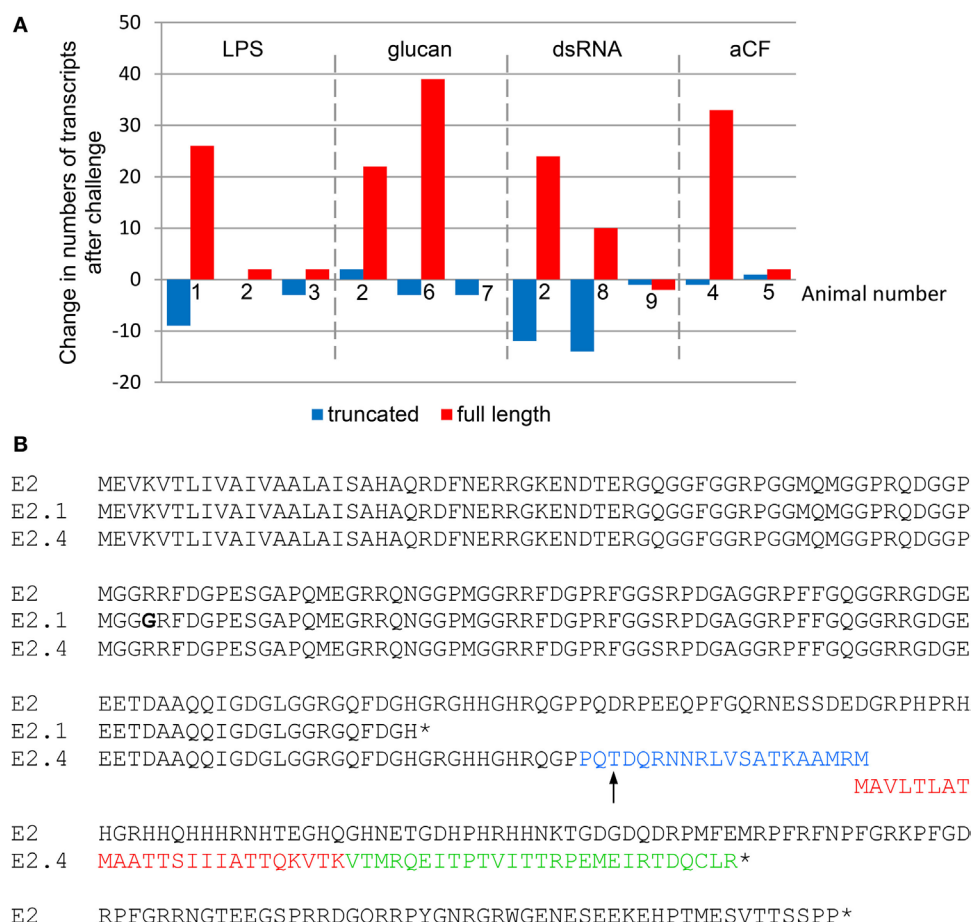


FIGURE 4 | Immune challenge decreases edited *SpTransformer* (*SpTrf*) mRNAs. **(A)** The change in transcripts encoding truncated (edited) vs. full-length sequences with regard to immune challenge with lipopolysaccharide (LPS), β -1,3-glucan (glucan), double stranded RNA (dsRNA), or sham injection (aCF; artificial coelomic fluid) relative to pre-challenge transcript numbers is shown for nine animals based on cDNA sequences reported previously (11). Animal 2 received separate challenges from all pathogen-associated molecular patterns. Amplicons from RT-PCR for animals that received LPS that were used for this analysis are shown in **Figure 1C**. Bars below 0 indicate fewer transcripts after challenge and bars above 0 indicate more. Missing bars indicate no change. This figure is modified from Ref. (33). **(B)** An alignment of deduced amino acid sequences from a full-length E2 protein and two truncated E2 proteins shows mismatches, frameshifts, and early stops. The *SpTrf* protein with an E2 element pattern is a full-length protein encoded by cDNA clone Sp0016 [GenBank accession number DQ183104.1 (6)]. In some cDNA sequences denoted E2.1, the sequence is edited at a specific glycine codon to a stop that is not encoded by the gene. The E2.1 truncated sequence is encoded by cDNA clone 1-1539 [GenBank accession number EF066308.1 (11)] and prior to the early stop is not identical to the E2 sequence used in the alignment (bold glycine is indicated). The E2.4 element pattern is an edited mRNA and encodes a truncated protein with missense sequence (cDNA clone 8-2415; GenBank accession number EF065834.1 (11)). The point of the frameshift is indicated with an arrow, which is followed by missense sequences that have been identified by proteomic methods (blue and red text) (32). Additional missense sequence in E2.4 is shown in green followed by an early stop codon. The alignment was done with BioEdit (34) and modified by hand. Stop codons are indicated by the (*).

immune response suggests that at least the E2.1-truncated proteins may have broad immuno-surveillance functions, whereas the full-length proteins may be more targeted to particular pathogens (3) (see below).

SpTrf GENE FAMILY SIZE AND STRUCTURE

The extraordinary diversity of 121 (~71%) unique sequences of 171 amplified *SpTrf* genes from three *S. purpuratus* sea urchins predicts that the gene family is likely large. Detailed analysis plus three different approaches for estimating the gene family size predicted 50 ± 10 *SpTrf* genes per genome [reviewed in

Ref. (16)]. In stark contrast to this estimate, only six genes are assembled in the sea urchin genome sequence. This lack of correlation may be the outcome of significant artifacts in genome assembly for genes with shared sequences that are tightly linked and associated with repeats (35). The apparent underestimation of the *SpTrf* gene family in the assembled genome sequence may be the result of assembling similar genes into hybrid sequences that do not actually exist in the real genome (2, 28, 29). Finding the correct structure and sequence of the *SpTrf* gene family led to a screen of the sea urchin gDNA BAC library followed by insert sequencing, assembly, and annotation that identified three clusters for a total of 15 *SpTrf* genes (**Figure 5**) (28, 29). Although 15 genes are many fewer than predicted, it is consistent with 18

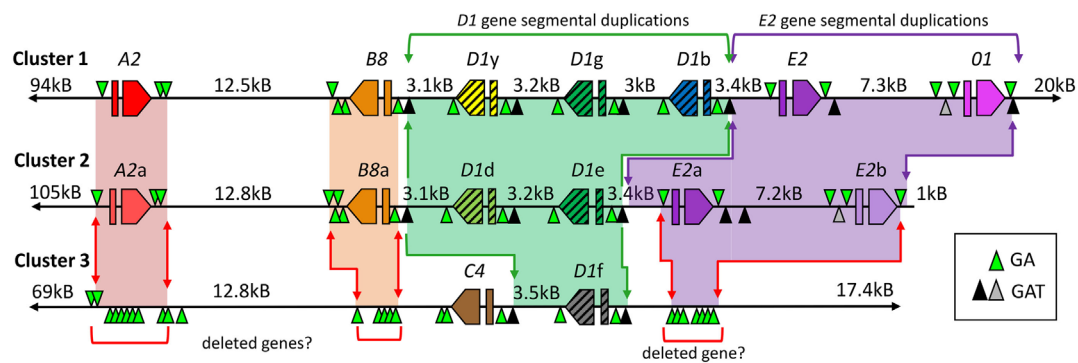


FIGURE 5 | Three clusters of *SpTransformer* (*SpTrf*) genes are present in the *Strongylocentrotus purpuratus* genome. The three clusters of genes are likely located at two loci within the genome. See **Figure 3B** for an illustration of the standard gene structure. Clusters 1 and 2 are likely allelic based on matches in the flanking regions outside of the gene clusters, even though the numbers of genes within the loci do not match. Genes are labeled by element pattern; however, those with the same element pattern are not necessarily of identical sequence. All genes are flanked by GA short tandem repeats (STRs) and may be the basis of deleted regions (red arrows), including genes, in Cluster 3 that are indicated by regions of GA STRs that are as long as 3 kB. Segmental duplications including *D1* genes (green shading and green arrows) and *E2* genes (purple shading and purple arrows) are flanked by GAT STRs (black triangles indicate >35 repeats, gray triangles indicate 4–17 repeats). Red and orange shading indicate likely alleles in Clusters 1 and 2. Regions of missing or deleted genes in Cluster 3 are indicated by red brackets. This figure is modified from Ref. (29).

genes predicted from the genome sequence traces available prior to assembly. Although it is possible that the *SpTrf* gene clusters may be unstable in BAC clones (see below), it is also feasible that the animal that provided gDNA for genome sequencing may have had a particularly small *SpTrf* gene family.

The clusters of *SpTrf* genes in the sea urchin genome sequence are positioned on both the positive and negative DNA strands in mixtures of genes with different element patterns that show significant sequence diversity within the clusters (**Figure 5**) (29). The genes are linked as tightly as 3 kB, although the flanking genes in Clusters 1 and 2 are positioned much farther from their nearest neighbor. All genes are flanked by short tandem repeats (STRs) of GA sequences. Moreover, all six of the *D1* genes and two of three of the *E2* genes are positioned within segmental duplications that are flanked by GAT STRs. The segments harboring the six *D1* genes are highly similar as are those with the three *E2* genes in addition to the *O1* gene in Cluster 1 (**Figure 5**) (28, 29). The long flanking regions on either side of Clusters 1 and 2 are very similar indicating that these two clusters are likely allelic even though the numbers of genes and their element patterns do not match. Clusters 1 and 2 are most similar to the *SpTrf* gene cluster on scaffold 125 of the sea urchin genome sequence; however, the genes on the scaffold appear to be hybrid sequences of both allelic clusters (and consequently are artificial sequences) and do not include the *O1* gene in Cluster 1. Hybrid gene sequences are predicted based on assembly approaches that use sequence reads from both alleles at a locus, compounded by efforts to avoid assembling both alleles in what would appear as tandem gene duplicates. Cluster 3 is quite different from Clusters 1 and 2 and is positioned at a different locus because the flanking regions do not match those of Clusters 1 and 2 (**Figure 5**). The two genes in Cluster 3 are positioned in the same orientation and are both surrounded by GA STRs, but only the *D1f* gene is positioned within a segmental duplication flanked by GAT STRs that shows sequence similarity to the *D1* duplications in the other two clusters. Outside of the

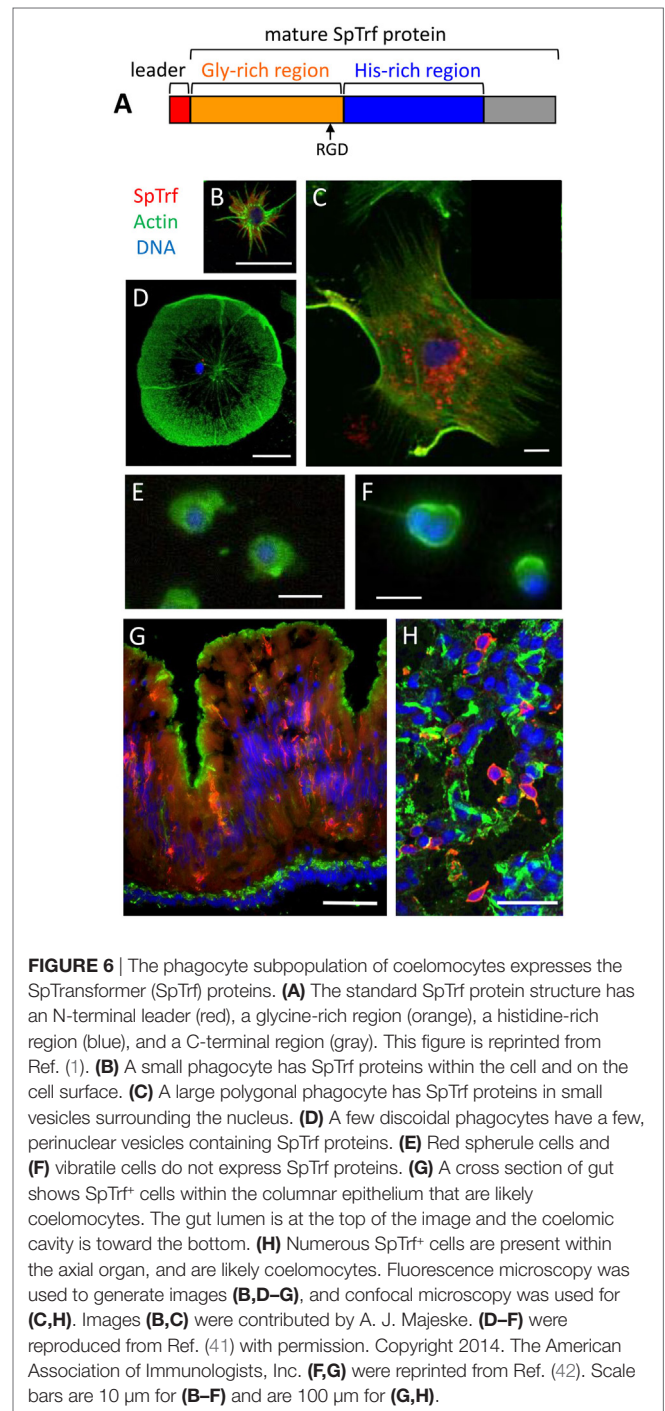
two *SpTrf* genes in Cluster 3 are flanking sequences with GA STRs of about 3 kB that are positioned at locations of ~3 kB and ~12 kB from the two genes, which match the locations of genes in the other two clusters. Speculations on the positions and functions of the STRs in the *SpTrf* gene clusters suggest that the GAT STRs may drive segmental duplications of regions that include the *D1* and *E2* genes (29). Sequence similarities among regions between the GA STRs that include the genes suggest that they may drive gene duplications (28). However, the size and locations of GA STRs flanking the genes in Cluster 3 are also consistent with gene deletions (2, 29). The non-matching allelic loci in Clusters 1 and 2 that include both different numbers of genes and variations in the element patterns in the second exon among the genes is consistent with the concept of genomic instability that may be based on shared sequences, shared repeats, and the association with many STRs within the clusters of this gene family [(2) and see below].

Although the concept of genomic instability intuitively seems lethal in that it could compromise both coding and regulatory regions, there can be advantages to genomic instability in localized and restricted regions. The advantage of small, tightly linked genes with shared patches of sequence, nearly identical segmental duplications, and tightly associated STRs, is that these attributes are likely essential for the sequence diversification of the *SpTrf* gene family (2). Rapid diversification is common for many innate immune genes that are under pathogen pressure and must keep pace in the arms race for host survival (1). This is consistent with swift changes in the members of the *SpTrf* gene family with the advantage of driving broad diversity of the expressed proteins (33, 36) that may be essential for interactions with the populations of microbial and other pathogens in the ocean that are simultaneously under selection for virulence to improve invasion, proliferation, and survival. A characteristic of many clustered genes that encode proteins with activities for interacting in some way with the environment such as pathogen recognition receptors or odorant or taste receptors (among others) is that although

the genes tend to change rapidly, the diversification process generates pseudogenes. For example, 25% of the 253 clustered *SpTLR* genes in the sea urchin genome sequence are pseudogenes (37), and 54% of the clustered human odorant receptor gene superfamily are pseudogenes (38). Mechanisms for correcting the reading frames in *SpTrf* pseudogenes have been speculated upon and may be an aspect of gene sequence diversification mechanisms, which are related to tight gene clustering (2, 28). Crossing over and gene conversion are enhanced in regions of the *Arabidopsis* genome that contain shared sequences, such as the disease resistance gene family (39). This process may also function for the *SpTrf* gene clusters based on the abundant shared sequences within and among the clusters. However, there must be some level of balance for gene conversion that would correct reading frames but with controls to block sequence homogenization among multiple linked family members. Homogenization of gene sequences within clusters would be disadvantageous in the arms race against pathogens. Hence, the conversion process that runs through a gene may be initiated by sequences shared among genes, but that progression to tightly linked genes may be limited by the presence of the GA STRs that surround all genes (28). This is consistent with increased sequence diversity in intergenic regions (excluding intergenic regions that are part of segmental duplications). However, a single *SpTrf* pseudogene that has been identified from 198 sequenced *SpTrf* genes has a deletion that alters the reading frame and is unusual because it is intronless and may be a retroposon. Possibilities as to why a retroposon may show a frameshift could be that it may not be expressed if it is not associated with a regulatory region and, therefore, may not be under pathogen pressure to maintain the ORF. Furthermore, if it was retro transposed into the genome in isolation away from clustered *SpTrf* genes, the theoretical mechanisms for diversification and reading frame corrections may not extend to isolated genes. The overall genomic instability predicted for the *SpTrf* gene family that is based on multiple types of repeats within and surrounding the clustered genes is consistent with the observation of differences in the repertoire of genes in the *SpTrf* family among individual sea urchins (29). Ongoing diversification of the *SpTrf* genes and the advantages of this process for host protection against pathogens require the input of new genes to the family as others are modified and/or deleted, and fits a description of swift evolution and the birth–death or duplication–deletion concept for duplicated genes (40).

DIVERSITY OF THE *SpTrf* PROTEINS

The rapid onset and increase in *SpTrf* gene expression in sea urchins upon immune challenge from microbes or PAMPs (8, 9, 11), the sequence diversity of the genes, and messages (6, 11, 15, 28, 29) in addition to putative mRNA editing (30) suggest that the encoded proteins are highly diverse and likely have immunological functions. The deduced structure of the *SpTrf* proteins indicates a hydrophobic leader and a mature protein of variable sizes that includes a glycine-rich region near the N-terminus with an arginine–glycine–aspartic acid (RGD) motif near the middle of most proteins suggestive of integrin binding, followed by a histidine-rich region, and a C-terminal region (Figure 6A).



The deduced sizes and sequences of the glycine-rich and the histidine-rich regions are highly variable based on the presence and absence of elements and the sequence variability within elements in the genes and messages (see Figure 2). The *HeTrf* proteins from *H. erythrogramma* have a similar structure including a C-terminal histidine-rich region with poly-histidine patches that vary from 6 to 13 histidines (21), which is more histidines than that have been identified in most of the *SpTrf* proteins. Only a few of the *HeTrf* proteins have an RGD motif whereas it is present in most

of the SpTrf proteins. The HeTrf proteins are composed of subsets of 26 possible elements and have four types of imperfect repeats that are positioned in both tandem and interspersed patterns. The sequences of both the elements and the repeats in HeTrf proteins are somewhat similar to those in SpTrf proteins, although the organization is different. These two homologous gene families encode proteins predicted to have similar anti-pathogen functions; however, their characteristics are not identical (21).

SpTrf PROTEINS ARE EXPRESSED IN A SUBSET OF PHAGOCYTES

There are four major morphotypes of coelomocytes in *S. purpuratus* that include phagocytes, red and colorless spherule cells, and vibratile cells (24), and only some of the phagocyte class of coelomocytes express the SpTrf proteins (41, 43). Surprisingly, the cells with the highest SpTrf expression are the small phagocytes in which the proteins are localized to cytoplasmic vesicles and the cell surface (Figure 6B). Some of the large phagocytes have SpTrf proteins localized to vesicles surrounding the nucleus but the proteins are never found on the cell surface (Figures 6C,D). The red spherule cells and the vibratile cells are consistently negative for SpTrf expression (Figures 6E,F). The expression patterns for HeTrf proteins in *H. erythrogramma* are similar to patterns of the SpTrf proteins, are localized to perinuclear vesicles, and are on the surface of some phagocytes (21). Analysis of the SpTrf protein expression patterns has benefited from the use of IQ sea urchins that tend to have decreased numbers of coelomocytes in the CF (43). When IQ sea urchins are challenged with LPS, there is a twofold increase in the total number of coelomocytes in the

CF after 24 h and a 10-fold increase in the SpTrf⁺ cells in the CF after 48–96 h (36, 43). Of those increased numbers of cells in the CF, the small phagocytes show a significant increase including more cells that express SpTrf proteins. In parallel, the percentage of polygonal phagocytes in the CF does not change in response to LPS; however, these cells tend to increase expression of the SpTrf proteins. These results may be interpreted as the production and secretion of SpTrf proteins from the polygonal phagocytes and the secretion plus acquisition of SpTrf proteins onto the surface of small phagocytes.

The swift pattern of *SpTrf* gene expression in phagocytes responding to immune challenge or injury can be imagined conceptually as the expression of as many of the *SpTrf* genes as quickly as possible and production of as many of the SpTrf proteins as appropriate to control or eliminate the detected pathogen. This would be advantageous in responding to infections and to protect the host from being overwhelmed by and succumbing to a pathogen. Surprisingly, when single phagocytes are evaluated for *SpTrf* transcripts, not only do most of the individual cells yield *SpTrf* amplicons of the same size (Figure 7) but the amplicon sequences from single cells are the same (41). This implies that one gene from the *SpTrf* family is expressed per individual phagocyte. Because sea urchins show a significant increase in messages (11) and SpTrf protein arrays (33, 36) in response to immune challenge, this swift response was considered feasible only if multiple *SpTrf* genes were expressed per phagocyte. Consequently, expression of a single *SpTrf* gene per phagocyte was an unexpected outcome. The mechanism for how this is regulated including expression of one gene and suppression of all the others, perhaps in response to the particular pathogen, is not known.

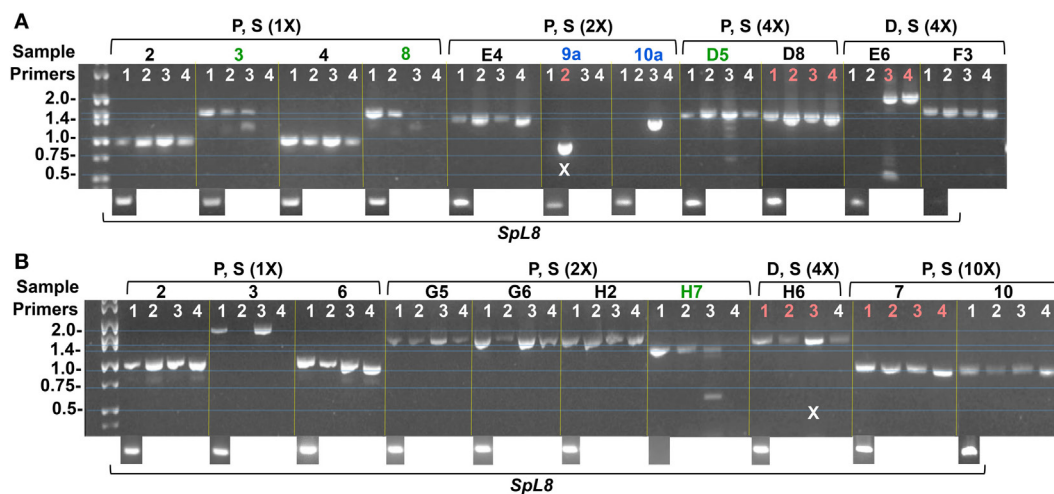


FIGURE 7 | Amplicons from single phagocyte indicate expression of single *SpTransformer* (*SpTrf*) genes in single cells. Coelomocytes were collected from two sea urchins (A,B), fractionated by Percoll density gradient into fractions of polygonal plus small phagocytes (P, S), and discoidal plus small phagocytes (D, S). Fractions were diluted to an estimate of 1 cell/μl followed by further dilutions of 2x, 4x, and 10x to ensure 1 cell/sample. Samples were first tested by nested RT-PCR using primers for *SpL8* (shown at the bottom of the images) that encodes the sea urchin homolog of protein 8 from the human large ribosomal subunit, and indicates samples that contain a cell. Samples with cells were evaluated for *SpTrf* transcripts by nested RT-PCR using four pairs of primers (1–4) on each sample that would amplify different sequence versions of *SpTrf* cDNAs. Green sample numbers indicate multiple bands amplified by the four primer pairs. Blue sample numbers indicate a single amplicon for a single pair of primers. Samples indicated in red were chosen for sequencing. X indicates failed or ambiguous sequence results. This figure is reproduced from Ref. (41) with permission. Copyright 2014. The American Association of Immunologists, Inc.

SpTrf EXPRESSION IN ADULT AND LARVAL SEA URCHIN TISSUES

In addition to expression in the phagocyte class of coelomocytes in adult sea urchins, SpTrf protein expression is also associated with non-immune tissues. Some of the cells within the columnar epithelium of the gut express SpTrf proteins (**Figure 6G**) (42), and similarly, the HeTrf proteins are localized to membranes of transport vesicles and the plasma membrane in gut associated amoebocytes (or phagocytes) (44). In addition to the gut epithelium, SpTrf proteins are also expressed in the pharynx, esophagus, and gonads (42). It is noteworthy that expression of the SpTrf proteins also occurs in the axial organ (**Figure 6H**), which shows increased expression after immune challenge. Although SpTrf proteins in sea urchin larvae have not been reported, *SpTrf* gene expression is restricted to a subset of blastocoelar cells that are localized in the blastocoel, extend filopodia across the blastocoel, form syncytia (45), and function as the primary larval phagocytes and act in host protection (46). The larval blastocoelar cells appear to be the functional equivalent of the large phagocytes in adult sea urchins based on cellular morphology, localization in the body cavity, phagocytic activity, and syncytia formation (47). Given that the blastocoelar cells are the only cell type in larvae to express the *SpTrf* genes, it is likely that the SpTrf protein expression in adult tissues is similarly restricted to phagocytes.

DIVERSE ARRAYS OF SpTrf PROTEINS ARE EXPRESSED IN RESPONSE TO IMMUNE CHALLENGE

The predicted sizes of the SpTrf and HeTrf proteins from cDNA sequences range from ~4 kDa for the smallest truncated protein to 54 kDa for the largest full-length protein, and overall, the most common size range is 35–40 kDa (6, 11, 21). However, the actual average size of SpTrf and HeTrf proteins on Western blots is 65–80 kDa with much larger sizes of over 200 kDa, which is likely the result of multimerization (21, 36, 43). The patterns and sizes of bands on standard one-dimensional Western blots for SpTrf and HeTrf proteins are different among sea urchins and change differently in response to challenge, illustrating the level of diversity of these proteins within and among animals (**Figure 8A**) (21, 36). When the Trf proteins are isolated from the CF and evaluated by 2D Western blots after isoelectric focusing, the extent of protein diversity is displayed as arrays of spots of which many appear as horizontal trains of spots mostly in the acidic range suggesting variations in pI for proteins of the same molecular weight (**Figure 8B**) (21, 36). Full-length SpTrf proteins with sufficient numbers of histidines can be isolated by nickel affinity and they also appear on 2D Western blots as horizontal trains but are found in the basic region of the blot in accordance with the positive charge on the histidines (**Figure 8C**) (33). When nickel-isolated SpTrf protein arrays are compared among sea urchins, the arrays differ among animals and show differences in the numbers and intensities of the SpTrf spots. Furthermore, the SpTrf arrays among individual sea urchins change differently in response to a series of challenges from different species of

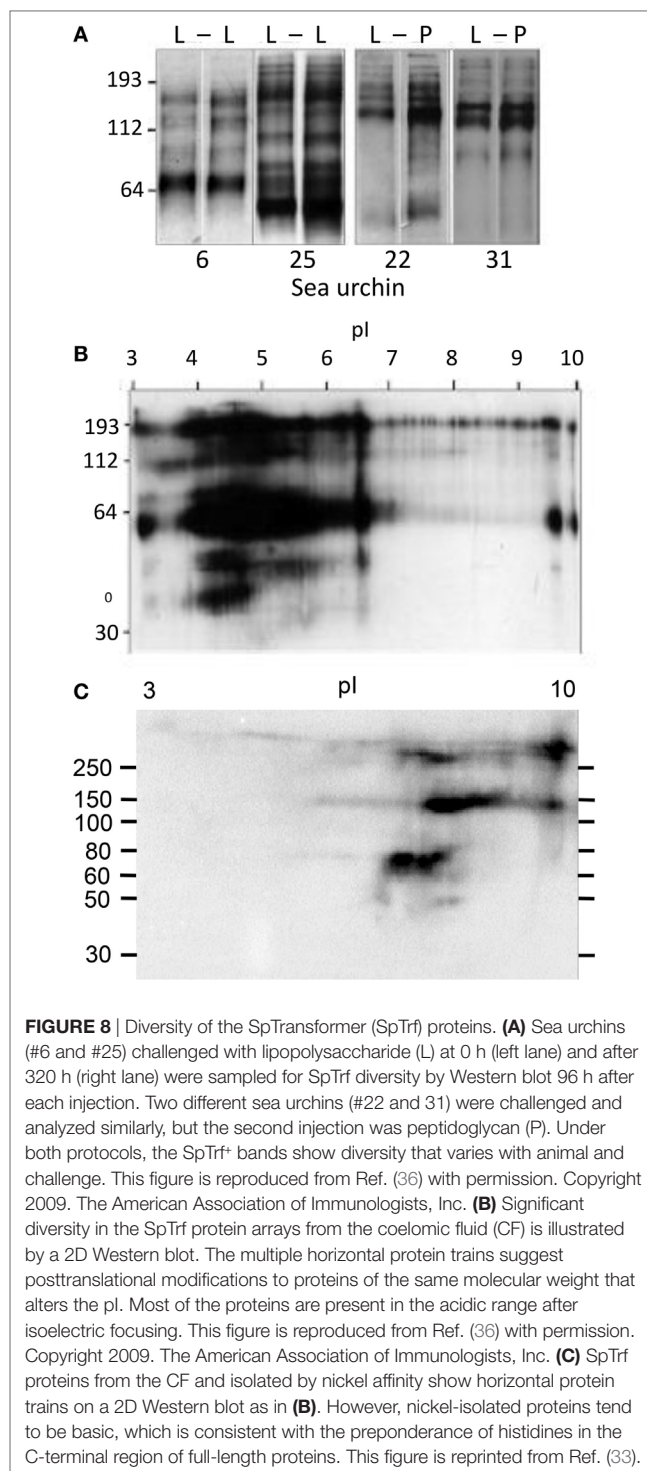


FIGURE 8 | Diversity of the SpTrf proteins. **(A)** Sea urchins (#6 and #25) challenged with lipopolysaccharide (L) at 0 h (left lane) and after 320 h (right lane) were sampled for SpTrf diversity by Western blot 96 h after each injection. Two different sea urchins (#22 and 31) were challenged and analyzed similarly, but the second injection was peptidoglycan (P). Under both protocols, the SpTrf⁺ bands show diversity that varies with animal and challenge. This figure is reproduced from Ref. (36) with permission. Copyright 2009. The American Association of Immunologists, Inc. **(B)** Significant diversity in the SpTrf protein arrays from the coelomic fluid (CF) is illustrated by a 2D Western blot. The multiple horizontal protein trains suggest posttranslational modifications to proteins of the same molecular weight that alters the pI. Most of the proteins are present in the acidic range after isoelectric focusing. This figure is reproduced from Ref. (36) with permission. Copyright 2009. The American Association of Immunologists, Inc. **(C)** SpTrf proteins from the CF and isolated by nickel affinity show horizontal protein trains on a 2D Western blot as in **(B)**. However, nickel-isolated proteins tend to be basic, which is consistent with the preponderance of histidines in the C-terminal region of full-length proteins. This figure is reprinted from Ref. (33).

bacteria (33). The extensive variations in the arrays of proteins in this family may be a combination of differences in numbers and varieties of genes in the *SpTrf* gene family among sea urchins plus the notion that changes in expression patterns may be tailored to the type of pathogenic challenge that is detected. This, in turn, suggests a detection system that has the ability to differentiate to some level among pathogens (36).

NATIVE SpTrf PROTEINS BIND FOREIGN CELLS

The association between SpTrf protein expression and immune challenge or injury suggests that these proteins impart important functions in host immune protection. This notion is also based, in part, on the unexpected level of diversity among the *SpTrf* genes, messages, and deduced protein sequences. Although bioinformatic analyses do not detect conserved domains and thus do not provide insights as to possible functions of the proteins, the hypothesis of immune activity has been tested initially with native SpTrf proteins isolated by nickel affinity. SpTrf proteins bind to Gram-negative and Gram-positive bacteria but show variations in binding capabilities among sea urchins (3, 33). Because individual sea urchins can express hundreds of SpTrf protein variants (33, 36), functional characterization of separated SpTrf proteins requires isolated variants. Efforts to achieve expression of six different recombinant SpTrf proteins in a bacterial expression system was successful for only one, suggesting that most of the SpTrf variants are highly toxic and may have antimicrobial activity (3). The single recombinant, rSpTrf-E1 (formerly rSp0032), has an E1 element pattern that is rarely identified among the reported cDNA sequences (2.5% of 688 cDNA sequences) (Figure 9A) (6, 11) and is the first SpTrf protein to be evaluated for function. When rSpTrf-E1 is incubated with two Gram-positive *Bacillus* species, the marine Gram-negative *Vibrio diazotrophicus*, and Baker's yeast, *Saccharomyces cerevisiae*, saturable binding is observed for *Vibrio* and *Saccharomyces*, but no binding is detected for either of the *Bacillus* species (Figures 9B,C) (3). Competition binding between labeled and unlabeled rSpTrf-E1 indicates specific binding sites on *Vibrio* and *Saccharomyces* (Figures 9D,E), and the two binding curves observed for *Saccharomyces* are also observed for competition binding (Figures 9C,E). These results demonstrate an unexpected outcome of a single protein binding selectively to multiple foreign targets with strong affinity. Furthermore, based on the variations in sequences among the native SpTrf proteins, binding results for rSpTrf-E1 infers that other versions may have different and perhaps overlapping ranges of targets.

SpTransformer proteins share a standard structure (Figure 6A) despite the sequence diversity; however, the differences in the amino acid compositions for the glycine-rich and histidine-rich regions of individual proteins have led to the notion that these regions may have different functions. Consequently, the recombinant fragments of rSpTrf-E1, the recombinant glycine-rich fragment (rGly-rich), recombinant C-terminal end of the gly-rich region (rC-Gly), and recombinant histidine-rich (rHis-rich) fragments (Figure 9A) show different binding characteristics compared to the full-length rSpTrf-E1 when tested against microbial targets (3). The three recombinant fragments bind to all tested foreign cells including the *Bacillus* species indicating altered and broadened binding relative to rSpTrf-E1. The central region of rSpTrf-E1, rC-Gly, multimerizes either in the presence or absence of binding targets and in the absence of other sea urchin proteins. Neither the rGly-rich nor the rHis-rich fragments include the rC-Gly region, and they do not multimerize indicating that this central region of the protein is responsible for multimerization of rSpTrf-E1 and likely for the native SpTrf proteins. The rGly-rich and rHis-rich

fragments show similar binding toward *Vibrio* and *Saccharomyces* compared to full-length rSpTrf-E1; however, they both show broadened binding toward the two *Bacillus* species unlike the full-length protein. Binding competition for *Saccharomyces* between the rGly-rich and rHis-rich fragments shows that each reduces binding by the other by 40% suggesting distinct but overlapping binding sites for each fragment. Similarly, when the competitor is the full-length rSpTrf-E1, it reduces binding to *Saccharomyces* by the rGly-rich fragment by 40% and fully competes with the rHis-rich fragment (Figure 10A). These results illustrate that rSpTrf-E1 and the rHis-rich fragment bind to the same sites on yeast, likely with the same mechanisms. However, the rGly-rich fragment when expressed separately binds to additional sites that are not recognized by either rSpTrf-E1 or the rHis-rich fragment. Given mRNA editing and the presence of Gly-rich truncated proteins in the CF [(32) and see Figures 2A and 4B], the broadened binding characteristic suggests possible immune surveillance activities in sea urchins (3). It is apparent that the regions of the full-length SpTrf proteins likely interact and may function together to define binding selectivity to certain target cells.

rSpTrf-E1 IS INTRINSICALLY DISORDERED AND UNDERGOES STRUCTURAL TRANSFORMATION

The multitasking activities of rSpTrf-E1 (i.e., binding to a range of foreign cells) are unique because most other anti-pathogen proteins bind to a single category of foreign cell types and suggest that several molecular targets may be the basis for cellular binding. When rSpTrf-E1 is incubated with *Vibrio*, analysis by gel electrophoresis and mass spectrometry shows that flagellin is co-localized in an SpTrf-positive band. This raises the possibility that binding by rSpTrf-E1 to foreign cells may be mediated through PAMPs (3). In addition to flagellin from *Vibrio*, rSpTrf-E1 also shows strong and specific binding to flagellin from *Salmonella typhimurium*, LPS from *Escherichia coli*, and β -1,3-glucan from *Saccharomyces*, but does not bind to peptidoglycan from *Bacillus subtilis* (Figure 10B). Competition assays among PAMPs shows that binding by rSpTrf-E1 to LPS can be competed by LPS, flagellin, and β -1,3-glucan, but not by peptidoglycan (Figure 10C). This demonstrates that rSpTrf-E1 binds specifically, tightly, and irreversibly to very different types of PAMPs; glucose polymers in β -1,3-glucan, a complex of sugars or lipids in LPS, and amino acids in the non-glycosylated flagellin from *Salmonella*. In contemplating the broad multitasking binding characteristics of rSpTrf-E1, the bioinformatic prediction is that this protein is likely an intrinsically disordered protein (IDP), which is composed of unfolded loops without any ordered relationships and with no secondary structure. This led to the hypotheses that the lack of secondary structure and the possibility of conformational plasticity, or the ability to acquire different sets of secondary folds such as α helices or β strands without energy input, may be a basis for how rSpTrf-E1 may bind and/or interact with such different targets (3, 4). The structural analysis of rSpTrf-E1 by circular dichroism (CD) confirms intrinsic disorder and shows that the protein transforms from disorder to mostly α helical

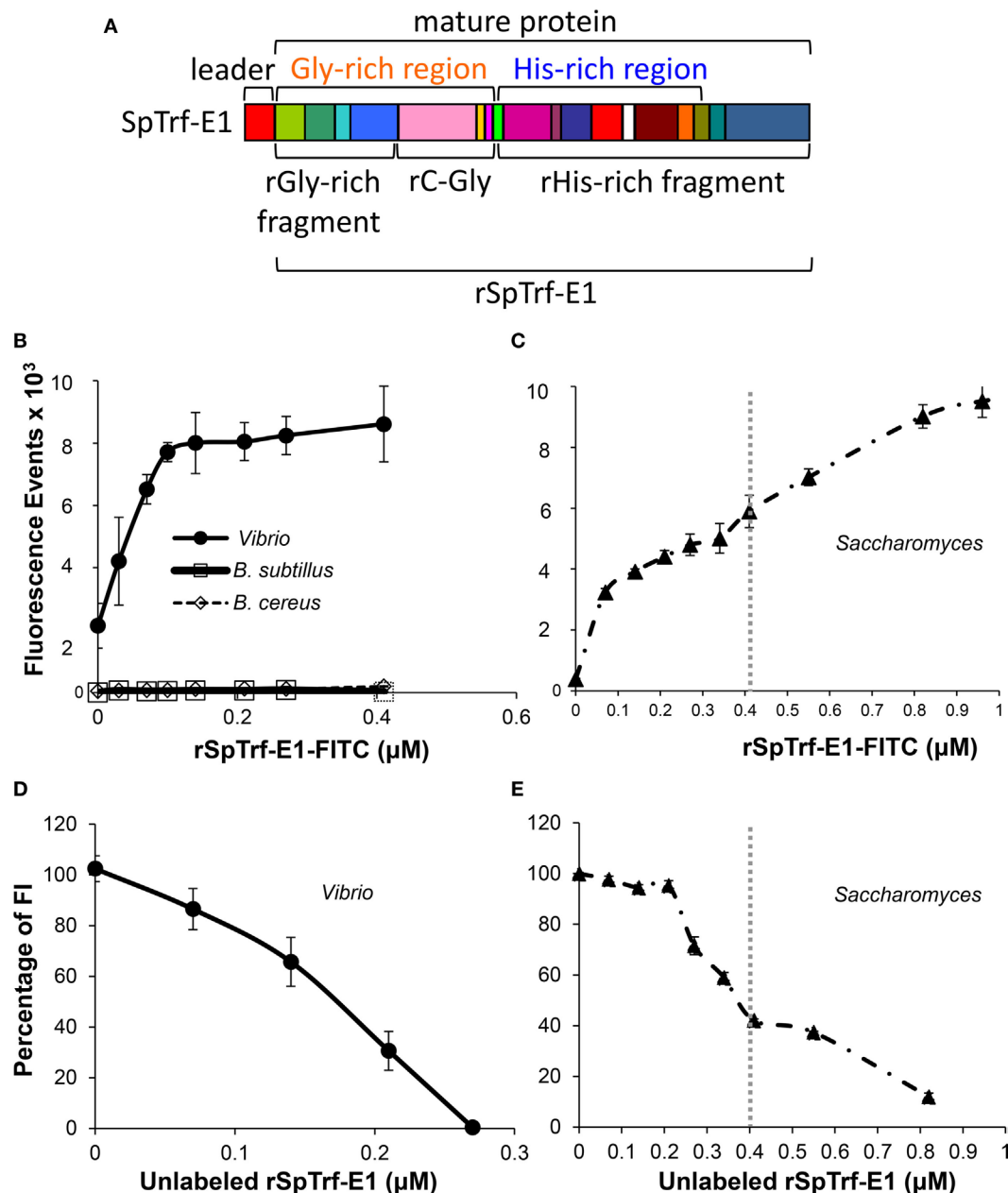


FIGURE 9 | The deduced structure and element pattern of rSpTransformer-E1 (rSpTrf-E1) and binding characteristics toward bacteria and yeast. **(A)** The deduced, full-length rSpTrf-E1 sequence predicts a leader (indicated), which is likely cleaved from the mature protein, plus a glycine-rich region (orange text) and a histidine-rich region (blue text). This structure is consistent with the standard SpTransformer (SpTrf) structure (see Figure 6A). The mature rSpTrf-E1 protein is composed of a mosaic of elements (colored blocks) that are defined by gaps based on the "cDNA-based" alignment (see Figure 2A for matching element colors) and is defined as an E1 element pattern according to Terwilliger et al. (6). The full-length rSpTrf-E1 and the recombinant fragments are indicated. This figure is modified from Ref. (48). **(B)** rSpTrf-E1 labeled with FITC (rSpTrf-E1-FITC) shows saturable binding to *Vibrio diazotrophicus* based on the increasing fluorescence events by flow cytometry with increasing protein concentration. rSpTrf-E1-FITC does not bind to *Bacillus subtilis* or *B. cereus*. **(C)** rSpTrf-E1-FITC binds to *Saccharomyces cerevisiae* and shows two independent non-linear binding curves (separated by gray dotted vertical line) based on fluorescence events from flow cytometry. Both curves indicate strong binding and the second curve (right of the dotted line) shows a saturable binding plateau. Results suggest specific saturable binding either to different sites on *S. cerevisiae*, or by different mechanisms. **(D)** rSpTrf-E1-FITC binds to specific sites on *V. diazotrophicus*. Binding competition with a fixed saturable concentration of rSpTrf-E1-FITC (as determined in (B) and set to 100% fluorescence) and mixed with increasing concentrations of unlabeled rSpTrf-E1 results in decreased fluorescence intensity (FI) of *V. diazotrophicus* by flow cytometry. This indicates that the proteins compete for the same sites. Data are shown as the mean \pm 1 SD of three independent experiments. This figure is from Ref. (3). **Figure 4E. (E)** As in (D), competition binding using a saturable level of rSpTrf-E1-FITC [as determined in (C) and set to 100% fluorescence] with increasing concentrations of unlabeled rSpTrf-E1 results in decreased FI of *S. cerevisiae* by flow cytometry. Results show two competition curves that correlate with the binding curves in (C). Data are shown as the mean \pm 1 SD of three independent experiments. Panels (B–E) are reprinted from Ref. (3) with permission from Elsevier.

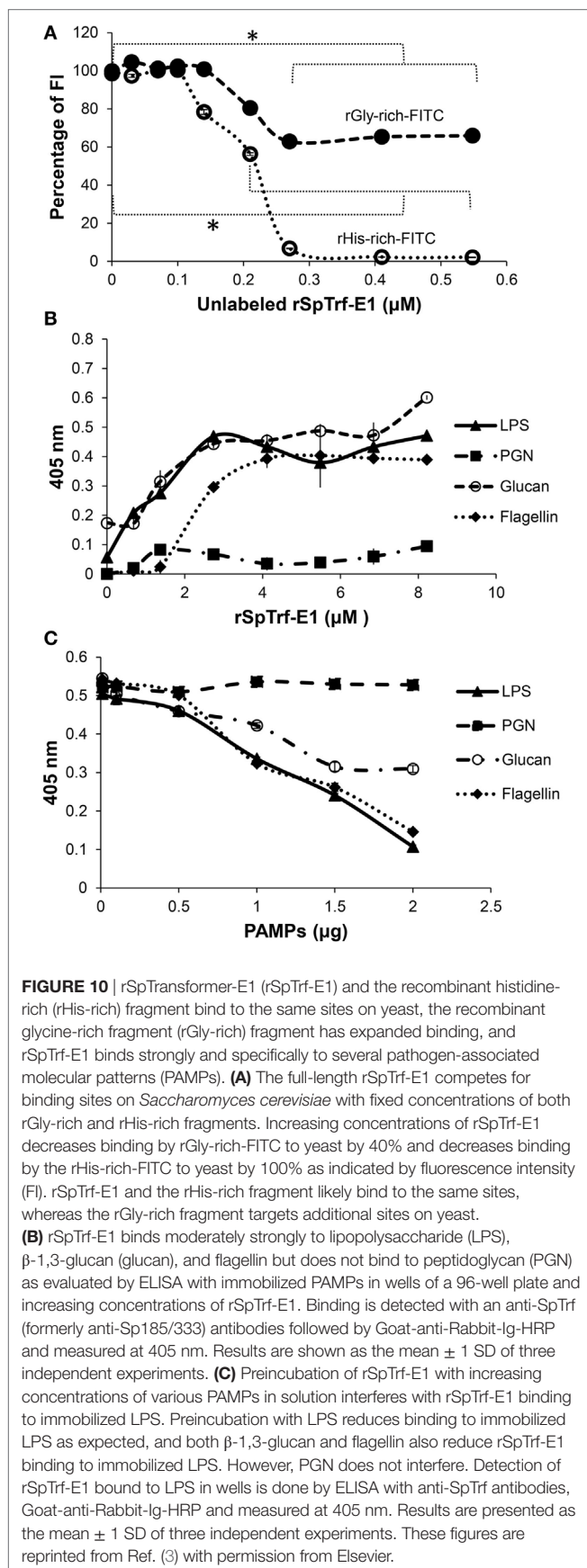


TABLE 2 | rSpTransformer-E1 (rSpTrf-E1) and the recombinant fragments show changes in secondary structure in different reagents and binding targets.^a

Protein structure	Secondary structure in different reagents				
	PO ₄	SDS	TFE	LPS	PA
rSpTrf-E1	Disordered	α	α	α	α
% α Helical	1–2 ^c	79	95.1	78.5	71.8
Helix tightness ^b		0.59	1	0.66	0.7
Recombinant glycine-rich fragment	α	α	β	β	n/d
% α Helical	15–17	75.1	N/A	N/A	
Helix tightness		0.78	N/A	N/A	
Recombinant histidine-rich fragment	α	α	β	α	n/d
% α Helical	19–30	70.7	46.2	72.8	
Helix tightness		0.76	N/A	0.78	

^aThese data are from Ref. (4, 5).

PO₄, 10 mM phosphate buffer pH = 7.4; SDS, sodium dodecyl sulfate; TFE, 2,2,2-trifluoroethanol; LPS, lipopolysaccharide from *Escherichia coli*; PA, phosphatidic acid in the form of small vesicles; n/d, not done; N/A, not applicable; the deconvolution to calculate the β strand percentage is not feasible for these samples (4).

^bHelix tightness is estimated from the R value obtained from circular dichroism (CD) spectra and is used to infer the width of an α helical twist. A standard helix has an R value of 1. A 3₁₀ helix has an R value of 0.4, which has a smaller diameter and is longer for a similar number of amino acids (4, 53).

^cThe percentage of secondary structure for either α helix or β strand is deconvoluted from the CD spectra using DichroWeb online server (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) (54, 55).

structure in the presence of sodium dodecyl sulfate (SDS), an anionic detergent that is used to simulate anionic environment (49), and 2,2,2-trifluoroethanol (TFE), which tends to promote secondary structure of α helices and β strands, and are commonly used reagents in CD studies (Table 2). Furthermore, rSpTrf-E1 readily transforms from disordered to α helical in the presence of LPS. The rGly-rich and rHis-rich fragments also show structural flexibility, but tend to be partially α helical in phosphate buffer, which is not predicted from sequence (4). In the presence of SDS, both the rHis-rich and rGly-rich fragments increase their α helical structure and in TFE both transform to β strand; however, in the presence of LPS, the rGly-rich fragment transforms to β strand and the rHis-rich fragment increases its α helical content (Table 2). These results not only led to the name change from Sp185/333 to SpTransformer to reflect the structural properties of the proteins, but also led to hypotheses for rSpTrf-E1-binding mechanisms. rSpTrf-E1 may have a transient initial binding state that can be established with multiple binding targets and is based on its unique amino acid sequence that is rich in polar and charged amino acids. This characteristic may be responsible for initiating “polyelectrostatic” interactions (50, 51) with negatively charged binding targets on pathogens, perhaps chemically similar to the sulfate group on SDS. The initial interaction may be followed quickly by a secondary step that is based on the hydrophilic nature and structural flexibility of rSpTrf-E1 as an IDP and its transformation to secondary folds for establishing tight binding with multiple targets. Although, the actual underlying chemical mechanism(s) for the binding process remain speculative, the extent of the transformation from disorder to secondary structure may be induced and/or

guided by the characteristics of the target. This provides an interesting parallel to an aspect of Linus Pauling's template theory of antibody formation and the generation of diversity in which direct interactions with an antigen induce the formation of the binding pocket from the unfolded variable domain (52). Since the time of Pauling's speculations, the mechanisms have been well characterized for generating and selecting for antigen receptors in jawed vertebrates with specific binding only to non-self. Non-rearranging anti-pathogen molecules in both vertebrates and invertebrates also target non-self, but through a wide range of mechanisms. In general, germ-line encoded molecules are evolutionarily selected for binding to PAMPs and not to self. The complexities presented by the *SpTrf* proteins, including their predicted sequence diversity (6, 11), disordered structure (4, 5), and predictions of *SpTrf* mRNA editing that can change the amino acid sequence or truncate the proteins (30), challenge the concepts of selection for non-self binding by germ-line encoded proteins. Furthermore, these attributes of the *SpTrf* system suggests that the mRNA editing may not be random (see Figure 4A).

rSpTrf-E1 BINDS PHOSPHATIDIC ACID (PA) AND DEFORMS MEMBRANES

The association of *SpTrf* proteins with coelomocyte membranes has been well documented (43, 44) but remains a mystery because there are no predicted transmembrane regions or conserved glycosphosphatidylinositol linkages from the primary amino acid sequences (11). Consequently, when tested for lipid binding, rSpTrf-E1, the rGly-rich, and the rHis-rich fragments all bind to PA, the rHis-rich fragment also binds weakly to phosphatidylinositol 4 phosphate, and rC-Gly binds weakly to phosphatidylserine (5). PA has a similar amphipathic structure as SDS except it has a phosphate head group, which is the likely binding site as none of the proteins bind to diacylglycerol. rSpTrf-E1 displays the same structural transformation from disordered to α helical in the presence of PA as it does with SDS (Table 2). When PA is incorporated into liposome membranes, rSpTrf-E1 alters liposome morphology, inducing budding or fission, fusion, and invagination (Figures 11A,B). Budding is illustrated by a liposome that buds and forms a total of three liposomes (Figure 11Aa–d; white arrows), fusion is shown between two different sized liposomes that form a single bean-shaped liposome (Figure 11Ba,b; orange arrows), and invagination is illustrated by the bean-shaped liposome that proceeds to a multi-lamellar liposome in which the internal liposome contains no luminal dextran labeled with Alexa Fluor® 488 (dextran-488) (Figure 11Bc,d). The uneven distribution of the luminal dextran-488 noted as dark regions within some liposomes suggests dextran-488 leakage (Figure 11Ac,d; white circles). To verify luminal leakage, liposomes loaded with both ANTS (fluorescent dye) and DPX (quencher) show that rSpTrf-E1 induces fluorescent dye leakage (Figure 11C). Only monomeric rSpTrf-E1 and the rHis-rich fragment induce leakage indicating that the histidine-rich region of the full-length protein is solely responsible for the leakage activity on membranes with PA. It is also noteworthy that pre-dimerized rSpTrf-E1 has no effect on liposomes, suggesting that dimerization and

multimerization of the *SpTrf* proteins deactivate or block their binding activity.

The morphological changes in the liposomes in the presence of rSpTrf-E1 are consistent with the unique structure of PA and the structural change in rSpTrf-E1 from disordered to α helical in the presence of PA (Table 2). PA is a conical phospholipid with a small phosphate head group (56) and its enrichment or clustering in a membrane is known to promote curvature (57). It is noteworthy that the dark luminal region near the convex portion of the liposome membrane in Figure 11Bc (white arrow) suggests leakage and that this is the site of invagination observed 1 min later (Figure 11Bd). These complex morphological changes occur at the same area of the liposome membrane and may be the result of PA bound to rSpTrf-E1. When liposomes composed of blue fluorescently labeled PA (NBD-PA, see legend to Figure 11) and phosphatidylcholine (PC) are incubated with rSpTrf-E1 for 20 min, NBD-PA appears as clusters of bright blue fluorescent patches in the membranes. There is usually a single NBD-PA cluster per liposome, and many are observed at intersections of two liposomes (Figure 11Da–c) and at regions of membranes showing concave curvature (Figure 11E). In one case, an NBD-PA cluster appears in a liposome with an extension from the cluster to outside of the membrane (Figure 11G; arrow). Control liposomes in the absence of rSpTrf-E1 show an even distribution of NBD-PA after 20 min (Figure 11F). When liposomes with NBD-PA are incubated with rSpTrf-E1 for 2 h, NBD-PA appears as disordered tangles outside of the liposome membranes (Figure 11H; arrow), whereas liposomes in the absence of rSpTrf-E1 continue to show an even distribution of NBD-PA in the membranes (Figure 11I). It is likely that the phosphate head group of PA is the binding target for rSpTrf-E1 based on the overall structural similarity to SDS and the amino acid composition of rSpTrf-E1 of which ~25% are positively charged and some or all may be involved with PA binding, although the exact mechanism is not known (5). The hypothesis of structural conformation and plasticity of rSpTrf-E1 is strengthened by the secondary structural changes from disorder to α helical in the presence of PA and the correlated morphological changes in liposomes containing PA. Although these results suggest how one version of the *SpTrf* proteins may associate with cell membranes, it is unknown whether PA is important for the observed association of *SpTrf* proteins on the surface of small phagocytes (see Figure 6B) (43). PA is usually present in small quantities in cells but is responsible for many physiological functions as a precursor for synthesis of other phospholipids, part of signaling pathways in response to stress, and other cellular activities (58–61). Although PA is known to be elevated on the cytoplasmic side of the cell membrane for vertebrate phagocytes (62) during phagocytosis (63), it is possible that *SpTrf* proteins bound to PA on a phagocyte surface may drive membrane curvature for phagocytosis or endocytosis during pathogen clearance (5).

CONCLUSION AND OVERVIEW OF THE *SpTrf* SYSTEM IN SEA URCHINS

The activities of rSpTrf-E1 and its recombinant fragments show unexpected multitasking activities with tight binding [e.g., $K_d = 0.2$ nM for *Vibrio*; (3)] toward certain microbes, PAMPs,

and lipids. The recombinant proteins provide new insights into how some of the *SpTrf* proteins may associate with potential pathogens and, perhaps, with membranes of both sea urchin phagocytes and bacterial surfaces. Activities of r*SpTrf*-E1 suggest that the sequence diversity of the *SpTrf* proteins may predict varying ranges of multitasking activities, with possible differing but overlapping activities toward varying groups or species of marine pathogens. We propose an overall model for *SpTrf* protein function in response to bacterial challenge that attempts

to include the results described in this review (**Figure 12**). Individual phagocytes appear to express a single *SpTrf* gene and produce a single *SpTrf* protein (41), given minor changes from mRNA editing (30). *SpTrf* proteins are stored in perinuclear vesicles of phagocytes (**Figures 7B–D**) (41, 43) and are speculated to be inactive with regard to binding and multimerization. Upon pathogen detection, different *SpTrf* protein isoforms are secreted into the CF by exocytosis from different phagocytes and may subsequently bind to the surface membrane of small phagocytes

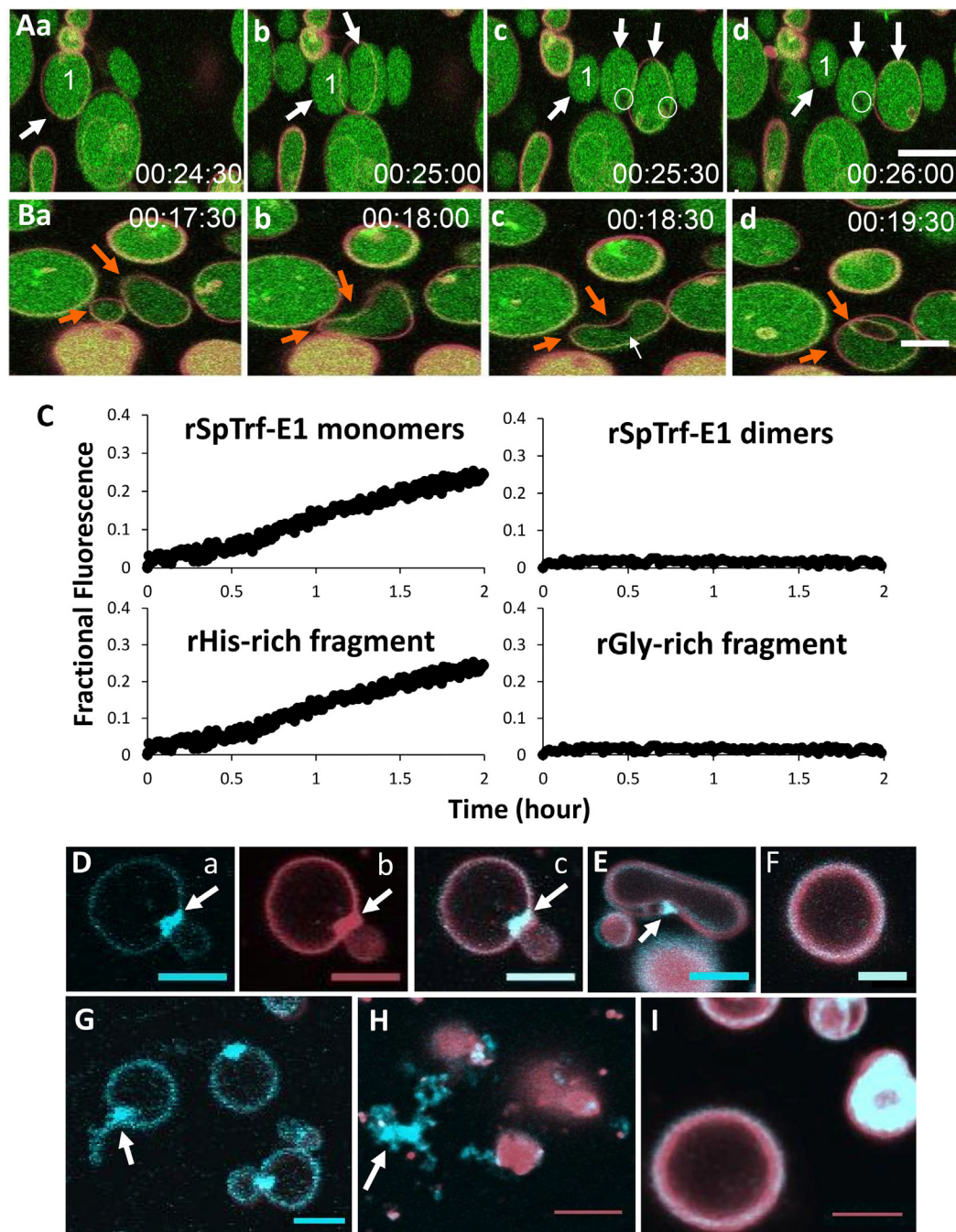


FIGURE 11 | Continued

FIGURE 11 | Continued

rSpTrf-transformer-E1 (rSpTrf-E1) causes membrane instability and induces liposomes to bud, fuse, invaginate, and leak contents. **(A)** Liposomes composed of 10% phosphatidic acid (PA) and 90% phosphatidylcholine are shown filled with dextran labeled with Alexa Fluor® 488 (green) and the membranes labeled with DiD (red). When in the presence of rSpTrf-E1, liposome labeled #1 shows budding or fission resulting in three liposomes (a–d, arrows). Images were captured by confocal microscopy every 30 s as indicated. Leakage of luminal green dextran is suggested from the black areas in the lumens of some liposomes (c,d, circles). **(B)** Two liposomes of different sizes fuse in the presence of rSpTrf-E1 (a,b, orange arrows). The fused liposome proceeds to invagination (c,d, orange arrows). Note the dark region in the lumen near the convex region of the liposome in (c), which is the site of invagination (d) that forms an internal liposome without luminal dextran labeled with Alexa Fluor® 488 (dextran-488). Images were captured by confocal microscopy every 30 s as indicated. **(C)** Only the monomeric rSpTrf-E1 and the recombinant histidine-rich (rHis-rich) fragment induce dextran-488 leakage from liposomes. Liposomes loaded with 10 mM 8-aminonaphthalene-1,3,6-trisulfonic acid disodium salt (ANTS; fluorescent dye) and 15 mM p-xylylene-Bis-pyridinium bromide (DPX; quencher) are incubated with 10 µM recombinant proteins. Luminal leakage separates ANTS from DPX by dilution into the buffer, which is excited at 360 nm and detected at 520 nm as fractional fluorescence relative to the control (lysed to measure 100% release). In the presence of monomeric rSpTrf-E1 and the rHis-rich fragment, luminal content leakage increases over time. Neither dimeric rSpTrf-E1 nor the rGly-rich fragment induce luminal content leakage from liposomes. **(D)** rSpTrf-E1 clusters PA in liposome membranes. A liposome composed of 10% fluorescent blue PA (1-oleoyl-2-{6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl}-sn-glycero-3-phosphate; NBD-PA; a, blue channel) plus the lipophilic dye DiD (b, red channel; c, merge) shows a PA cluster (arrows) at the intersection of two liposomes after 20 min of incubation with rSpTrf-E1. **(E)** NBD-PA is clustered (arrow) at the convex curve in a liposome membrane after 20 min of incubation with rSpTrf-E1. This image is a merge of the blue and red channels. **(F)** A control liposome shows no change in the distribution of NBD-PA after 20 min without rSpTrf-E1. This image is a merge of the blue and red channels. **(G)** Liposomes show clusters of NBD-PA after 20 min in the presence of rSpTrf-E1. One liposome shows extraction of NBD-PA from the membrane (arrow; blue channel only). **(H)** NBD-PA is extracted from liposome membranes after 2 h of incubation with rSpTrf-E1 and forms disordered clusters that are separated from liposomes (arrow). **(I)** Control liposomes show an even distribution of NBD-PA in the liposome membrane after 2 h in the absence of rSpTrf-E1 (merge of blue and red channels). Images in **(A,B,D–I)** were captured by confocal microscopy and all scale bars indicate 10 µm. These figures are reprinted from Ref. (5).

(Figure 12; green cell). In addition, the perinuclear vesicles may also contain membrane-bound SpTrf proteins that become associated with the cell surface upon incorporation of the vesicle membrane with the plasma membrane during exocytosis (44). The membrane association of SpTrf proteins may involve a putative membrane receptor(s) rather than or in addition to binding through PA. The SpTrf proteins that are likely secreted as IDPs, bind quickly to pathogens through strong affinity to PAMPs, followed by structural transformation to α helices (Figure 12) or other secondary folds. It is noteworthy that the concentration of SpTrf proteins in the cell-free CF is very low and that nickel-isolated native SpTrf proteins often appear as multimers (33, 36, 43), suggesting that the active proteins have a short half-life as IDPs and either bind to pathogens or multimerize and are inactivated (3) (see Figure 11C). We hypothesize that multimerization of different SpTrf variants secreted from different phagocytes occurs upon pathogen binding and opsonization that leads to pathogen clearance by triggering phagocytosis through putative receptor(s) (potentially including PA) on the polygonal phagocytes (Figure 12). In support of this hypothesis, HeTrf proteins have been observed in phagosomes in association with bacteria in the sea urchin, *H. erythrogramma* (44). Alternatively, there may be membrane-bound SpTrf proteins on phagocytes that function as putative receptors for SpTrf proteins that have opsonized bacteria. The subsequent multimerization among proteins on both the microbe and the coelomocyte surface may lead to phagocytosis. This notion is particularly interesting if PA is present on the coelomocyte plasma membrane and is clustered as a result of SpTrf binding to induce membrane curvature, which would assist with progression to phagocytosis (Figure 12; top left insert).

THE SpTrf SYSTEM HAS MULTIPLE LEVELS OF DIVERSIFICATION

The host–pathogen arms race drives diversification of pathogens to improve their abilities to infect, proliferate, disseminate, and

survive. The requirement for the host to survive the arms race also drives diversification mechanisms of the host immune system to detect and respond to constantly changing pathogens (1, 16, 64). The best example of host immune diversification is the well-understood vertebrate somatic recombination of the Ig and TcR genes that function in immune detection and response and that are diversified by the recombinase enzymes encoded by the *RAG1/2* genes (65, 66). Interest in the evolutionary origins of the RAGs has led to the identification of homologs in a few invertebrates (67–69). *SpRAG1L* and *SpRAG2L* homologs are present and linked in the sea urchin genome, are expressed in embryos and coelomocytes (67), and the *SpRAG1L* enzyme functions with mouse RAG2 to generate a low level of DNA recombination (70). Although intriguing, it is not clear whether *SpRAG1L* and *SpRAG2L* function together in sea urchin cells, and neither the DNA sequences that they may recognize nor the genes that they may impact are known. Although swift changes in the *SpTrf* gene family structure and diversity may be considered as theoretical connections to *SpRAGL* recombinase activity, it is not known whether these enzymes are involved in changes in the diversity of this gene family.

The diversity of the SpTrf system has been attributed to five levels of diversification with the beneficial outcome of generating a range of SpTrf proteins in the CF that extend beyond the diversity of the *SpTrf* gene family encoded in the genome (Figure 13). Level 1: the sequence diversity among the members of the *SpTrf* gene family, including the structure of the family in clusters of genes with shared sequences, in addition to possible gene conversion, segmental duplications, and putative gene deletions that appear to be associated with STRs, suggest localized genomic instability that may be required for gene diversification in this system (2, 28, 29). Genomic instability is consistent with differences in the members of the *SpTrf* gene family among sea urchins (29). Level 2: *SpTrf* gene expression from single phagocytes has inferred that only a single *SpTrf* gene is expressed per cell (41). This leads to the hypothesis that variations in the *cis* and/or *trans* regulatory regions associated with the *SpTrf* genes may control

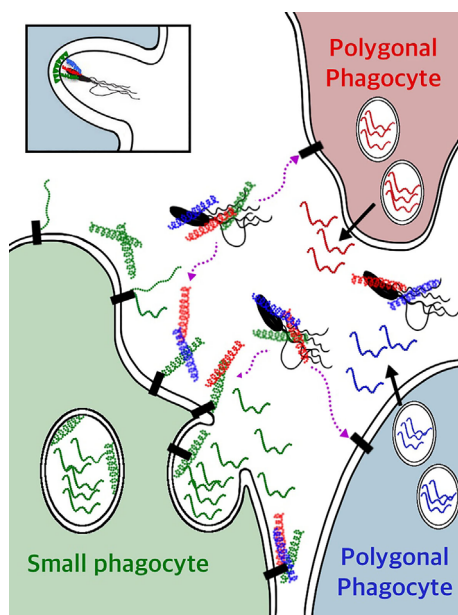


FIGURE 12 | A model for SpTransformer (SpTrf) protein functions for clearance of bacteria from the coelomic fluid (CF). Individual phagocytes secrete a single SpTrf protein variant (41), which is illustrated by individual phagocytes (red, green, blue) producing different (color coded) SpTrf protein variants. Bioinformatic predictions of many deduced SpTrf sequences and circular dichroism results for rSpTransformer-E1 (4) indicate that these proteins are likely intrinsically disordered proteins (IDPs) (squiggles). Upon interaction with or binding to targets in the CF, they transform to α helical structures (corkscrews). Whether the SpTrf proteins associate directly with phospholipids on the surface of small phagocytes (green cell) or whether SpTrf proteins associate with any phagocyte type through putative membrane receptor(s) (black rectangles) remain unknown and await investigation. When vesicle membranes fuse with the cell membrane, the membrane-bound SpTrf proteins are exposed on the surface of the small phagocyte (green cell) (44). Other SpTrf proteins that are secreted by nearby polygonal phagocytes and released into the CF likely bind quickly to pathogens through pathogen-associated molecular patterns (lipopolysaccharide, flagellin, or both) on the pathogen surface and swiftly transform from IDPs to proteins with ordered structure forming helices. Alternatively, secreted SpTrf proteins may bind to the surface of small phagocytes through multimerization with other membrane-bound SpTrf proteins, or may bind directly to phospholipids or to putative receptor(s) (black rectangles). The secreted SpTrf proteins that bind to pathogens may function as opsonins and trigger phagocytosis and pathogen clearance. The insert at the top left illustrates a theoretical clustering of phosphatidic acid (green triangles) in the outer leaflet of a phagocyte plasma membrane (represented as the double black line) by SpTrf proteins bound to the bacterium and induce the concave curvature in the membrane that may aid in the formation of the phagosome and uptake of a microbe. Other mechanisms that are known to be involved with phagosome formation are not shown.

whether specific or subsets of genes are expressed (or repressed) in phagocyte responses to particular pathogens or categories of pathogens. This putative second level of gene expression control could limit or target the diversity of the expressed proteins to optimize protection against particular pathogens and is expected to require coordination among responding and non-responding phagocytes. Level 3: the prediction of mRNA

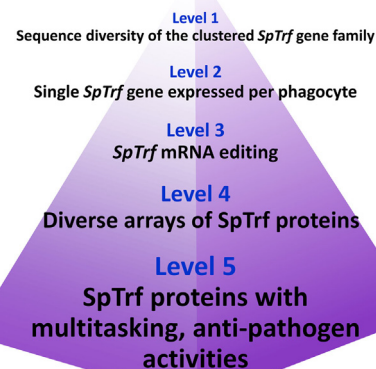


FIGURE 13 | Five levels of diversification in the SpTransformer (SpTrf) system.

editing increases the diversity of the mRNAs particularly when they are translated (edited or not) to both full-length and truncated proteins that may include missense sequence (30, 32). Editing is expected to expand the diversity of the proteins relative to the sequences encoded by the genes, including the possibility of expanded binding capabilities for truncated SpTrf proteins that are missing the histidine-rich region (3, 11). The increased presence of edited mRNAs encoding truncated and/or missense proteins prior to immune challenge suggests an active, non-random editing process with an outcome of altered functions for truncated proteins. Level 4: the diverse arrays of SpTrf proteins are the outcome of the diversification processes described in the preceding levels, which are putatively broadened further by posttranslational modifications that may alter protein function. These types of modifications have been suggested from the arrays of SpTrf proteins with the same molecular weight but with wide ranges of pI and *vice versa* (33). This may be the result of a number of types of posttranslational changes to proteins including multimerization, glycosylation for which there are a number of conserved linkage sites within and among the SpTrf isoforms (6), in addition to possibilities for phosphorylation and acetylation (33). Level 5: the new diversification level for this system is the unexpected range of rSpTrf-E1 protein functions and its unusual structural characteristics that may apply to many, if not most of the SpTrf proteins (3–5). The variety of SpTrf proteins that are expressed in response to a particular pathogen may each display differing but also overlapping ranges of multitasking activities that are based on the hydrophilic character of the proteins, the prediction that they are flexible IDPs, and the expectation that they undergo structural transformation upon binding to a range of targets. Nickel-isolated native SpTrf proteins bind to bacteria and yeast (3) and may function as opsonins to augment phagocytosis. The ability to bind selectively and tightly to multiple PAMPs is likely to confound the abilities of potential marine pathogens and opportunists to alter simultaneously multiple molecular attributes to avoid recognition, opsonization, and

possible killing by the SpTrf proteins. These multiple levels of diversification plus the flexibility and the predicted multitasking activities of SpTrf proteins are novel solutions in the immunological arms race and provide evidence for how this immune protein family may act as an extraordinarily effective component of the immune system in echinoids.

AUTHOR CONTRIBUTIONS

LCS and CML wrote, edited, and approved the manuscript.

REFERENCES

- Smith LC, Coscia MR. Tuning the host-pathogen relationship through evolution with a special focus on the echinoid *Sp185/333* system. *Invertebr Survival J* (2016) 13:355–73.
- 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 echinoid *185/333* and the plant *NBS-LRR*. In: Malagoli D, editor. *The Evolution of the Immune System: Conservation and Diversification*. London, UK: Elsevier-Academic Press (2016). p. 295–310.
- Lun CM, Schrankel CS, Chou H-Y, 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) 221:889–903. doi:10.1016/j.imbio.2016.03.006
- 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) 198:2957–66. doi:10.4049/jimmunol.1601795
- Lun CM, Samuel RL, Gillmor SD, Boyd A, Smith LC. The recombinant sea urchin immune effector protein, rSpTransformer-E1, binds to phosphatidic acid and deforms membranes. *Front Immunol* (2017) 8:481. doi:10.3389/fimmu.2017.00481
- 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) 26:134–44. doi:10.1152/physiolgenomics.00011.2006
- Smith LC, Chang L, Britten RJ, Davidson EH. Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes. *J Immunol* (1996) 156:593–602.
- Rast JP, Pancer Z, Davidson EH. New approaches towards an understanding of deuterostome immunity. *Curr Top Microbiol Immunol* (2000) 248:3–16.
- Nair SV, Del Valle H, Gross PS, Terwilliger DP, Smith LC. Microarray 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
- 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 U S A* (2000) 97:9514–8. doi:10.1073/pnas.160261897
- 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, beta-1,3-glucan, and dsRNA. *BMC Mol Biol* (2007) 8:16. doi:10.1186/1471-2199-8-16
- Gross PS, Al-Sharif WZ, Clow LA, Smith LC. Echinoderm immunity and the evolution of the complement system. *Dev Comp Immunol* (1999) 23:429–42. doi:10.1016/S0145-305X(99)00022-1
- Clow LA, Gross PS, Shih CS, Smith LC. Expression of SpC3, the sea urchin complement component, in response to lipopolysaccharide. *Immunogenetics* (2000) 51:1021–33. doi:10.1007/s002510000233
- Gross PS, Clow LA, Smith LC. SpC3, the complement homologue from the purple sea urchin, *Strongylocentrotus purpuratus*, is expressed in two subpopulations of the phagocytic coelomocytes. *Immunogenetics* (2000) 51:1034–44. doi:10.1007/s002510000234
- 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
- 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
- Janies DA, Voight JR, Daly M. Echinoderm phylogeny including *Xyloplax*, a progenetic asteroid. *Syst Biol* (2011) 60:420–38. doi:10.1093/sysbio/syr044
- Telford MJ, Lowe CJ, Cameron CB, Ortega-Martinez O, Aronowicz J, Oliveri P, et al. Phylogenomic analysis of echinoderm class relationships supports Asterozoa. *Proc Biol Sci* (2014) 281:20140479. doi:10.1098/rspb.2014.0479
- Kober KM, Bernardi G. Phylogenomics of strongylocentrotid sea urchins. *BMC Evol Biol* (2013) 13:88. doi:10.1186/1471-2148-13-88
- Wang U, Ding J, Liu Y, Liu X, Chang Y. Isolation of immune-relating *185/333-1* gene from sea urchin (*Strongylocentrotus intermedius*) and its expression analysis. *J Ocean Univ China* (2016) 15:163–70. doi:10.1007/s11802-016-2707-4
- 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:e62079. doi:10.1371/journal.pone.0062079
- Brites D, McTaggart S, Morris K, Anderson J, Thomas K, Colson I, et al. The Dscam homologue of the crustacean *Daphnia* is diversified by alternative splicing like in insects. *Mol Biol Evol* (2008) 25:1429–39. doi:10.1093/molbev/msn087
- Buckley KM, Florea LD, Smith LC. A method of identifying alternative or cryptic 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
- Smith LC, Ghosh J, Buckley KM, Clow LA, Dheilly NM, Haug T, et al. Echinoderm immunity. In: Soderhall K, editor. *Invertebrate Immunity*. Austin, TX: Landes Bioscience and Spring Science+Business Media (2010). p. 260–301.
- 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) 379:912–28. doi:10.1016/j.jmb.2008.04.037
- Ghosh J, Buckley KM, Nair SV, Raftos DA, Miller CA, 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
- Lee YH. Molecular phylogenies and divergence times of sea urchin species of Strongylocentrotidae, Echinoidea. *Mol Biol Evol* (2003) 20:1211–21. doi:10.1093/molbev/msg125
- 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) 11:575. doi:10.1186/1471-2164-11-575
- 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) 17:900. doi:10.1186/s12864-016-3241-x

ACKNOWLEDGMENTS

The authors are grateful to Audrey Majeske for providing the images in **Figures 6B,C**. The authors are indebted to Megan Barela Hudgell for improvements in **Figures 5** and **12**.

FUNDING

Support for research on sea urchin immunology and writing this review was awarded by the National Science Foundation (IOS-1146124 and IOS-1550474) to LCS.

30. Buckley KM, Terwilliger DP, Smith LC. Sequence variations in 185/333 messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. *J Immunol* (2008) 181:8585–94. doi:10.4049/jimmunol.181.12.8585
31. Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, et al. The genome of the sea urchin, *Strongylocentrotus purpuratus*. *Science* (2006) 314:941–52. doi:10.1126/science.1133609
32. Dheilly NM, Raftos DA, Haynes PA, Smith LC, Nair SV. Shotgun proteomics of coelomocytes from the purple sea urchin, *Strongylocentrotus purpuratus*. *Dev Comp Immunol* (2013) 40:35–50. doi:10.1016/j.dci.2013.01.007
33. Sherman LS, Schrankel CS, Brown KJ, Smith LC. Extraordinary diversity of immune response proteins among sea urchins: nickel-isolated Sp185/333 proteins show broad variations in size and charge. *PLoS One* (2015) 10:e0138892. doi:10.1371/journal.pone.0138892
34. 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.
35. Tang H. Genome assembly, rearrangement, and repeats. *Chem Rev* (2007) 107:3391–406. doi:10.1021/cr0683008
36. Dheilly NM, Nair SV, Smith LC, Raftos DA. Highly variable immune-response proteins (185/333) from the sea urchin *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. *J Immunol* (2009) 182:2203–12. doi:10.4049/jimmunol.07012766
37. 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
38. Olender T, Lancet D, Nebert DW. Update on the olfactory receptor (OR) gene superfamily. *Hum Genomics* (2008) 3:87–97. doi:10.1186/1479-7364-3-1-87
39. Mondragon-Palomino M, Meyers BC, Micheltore RW, Gaut BS. Patterns of positive selection in the complete NBS-LRR gene family of *Arabidopsis thaliana*. *Genome Res* (2002) 12:1305–15. doi:10.1101/gr.159402
40. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* (2000) 290:1151–5. doi:10.1126/science.290.5494.1151
41. Majeske AJ, Oren M, Sacchi S, Smith LC. Single sea urchin phagocytes express messages of a single sequence from the diverse Sp185/333 gene family in response to bacterial challenge. *J Immunol* (2014) 193:5678–88. doi:10.4049/jimmunol.1401681
42. Majeske AJ, Oleksyk TK, Smith LC. The Sp185/333 immune response genes and proteins are expressed in cells dispersed within all major organs of the adult purple sea urchin. *Innate Immun* (2013) 19:569–87. doi:10.1177/1753425912473850
43. Brockton V, Henson JH, Raftos DA, Majeske AJ, Kim YO, 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) 121:339–48. doi:10.1242/jcs.012096
44. Dheilly N, Birch D, Nair SV, Raftos DA. Ultrastructural localization of the highly variable immune response protein (He185/333) within coelomocytes and the gut tissues of the sea urchin, *Heliocidaris erythrogramma*. *Immunol Cell Biol* (2011) 89:861–9. doi:10.1038/icb.2011.3
45. Tamboline CR, Burke RD. Secondary mesenchyme of the sea urchin embryo: ontogeny of blastocoelar cells. *J Exp Zool* (1992) 262:51–60. doi:10.1002/jez.1402620108
46. Ho ECH, Buckley KM, Schrankel CS, Schuh NW, Hibino T, Sulek CM, et al. Perturbation of gut bacteria induces a coordinated cellular immune response in the purple sea urchin larva. *Immunol Cell Biol* (2016) 94:861–74. doi:10.1038/icb.2016.51
47. Majeske A, Bayne CJ, Smith LC. Aggregation of sea urchin phagocytes is augmented *in vitro* by lipopolysaccharide. *PLoS One* (2013) 8:e61419. doi:10.1371/journal.pone.0061419
48. Smith LC, Lun CM. Multitasking rSp0032 has anti-pathogen binding activities predicting flexible and effective immune responses in sea urchins mediated by the Sp185/333 system. *Pathog Infect Dis* (2016) 2:e1394. doi:10.14800/pid.1394
49. De Latour FA, Amer LS, Papanastasiou EA, Bishop BM, Van Hoek ML. Antimicrobial activity of the *Naja atra* cathelicidin and related small peptides. *Biochem Biophys Res Commun* (2010) 396:825–30. doi:10.1016/j.bbrc.2010.04.158
50. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Chan HS. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc Natl Acad Sci U S A* (2007) 104:9650–5. doi:10.1073/pnas.0702580104
51. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit* (2010) 23:105–16. doi:10.1002/jmr.961
52. Pauling L. A theory of the structure and process of formation of antibodies. *J Am Chem Soc* (1940) 62:2643–57. doi:10.1021/ja01867a018
53. Vieira-Pires RS, Morais-Cabral JH. 3₁₀ helices in channels and other membrane proteins. *J General Physiol* (2010) 136:585–92. doi:10.1085/jgp.201010508
54. Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* (2004) 32:W668–73. doi:10.1093/nar/gkh371
55. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* (2008) 89:392–400. doi:10.1002/bip.20853
56. Zimmerberg J, Kozlov MM. How proteins produce cellular membrane curvature. *Nat Rev Mol Cell Biol* (2006) 7:9–19. doi:10.1038/nrm1784
57. Kooijman EE, Chupin V, Fuller NL, Kozlov MM, De Kruijff B, Burger KNJ, et al. Spontaneous curvature of phosphatidic acid and lysophosphatidic acid. *Biochemistry* (2005) 44:2097–102. doi:10.1021/bi0478502
58. Testerink C, Munnik T. Phosphatidic acid: a multifunctional stress signaling lipid in plants. *Trends Plant Sci* (2005) 10:368–75. doi:10.1016/j.tplants.2005.06.002
59. Zeniou-Meyer M, Zabari N, Ashery U, Chasserot-Golaz S, Haeblerle AM, Demais V, et al. Phospholipase D1 production of phosphatidic acid at the plasma membrane promotes exocytosis of large dense-core granules at a late stage. *J Biol Chem* (2007) 282:21746–57. doi:10.1074/jbc.M702968200
60. Liu Y, Su Y, Wang X. Phosphatidic acid-mediated signaling. *Adv Exp Med Biol* (2013) 991:159–76. doi:10.1007/978-94-007-6331-9_9
61. Yao J, Rock CO. Phosphatidic acid synthesis in bacteria. *Biochim Biophys Acta* (2013) 1831:495–502. doi:10.1016/j.bbalip.2012.08.018
62. Bohdanowicz M, Schlam D, Hermansson M, Rizzuti D, Fairn GD, Ueyama T, et al. Phosphatidic acid is required for the constitutive ruffling and macropinocytosis of phagocytes. *Mol Biol Cell* (2013) 24(700–1712):S1711–7. doi:10.1091/mbc.E12-11-0789
63. Kassas N, Tanguy E, Thahouly T, Fouillen L, Heintz D, Chasserot-Golaz S, et al. Comparative characterization of phosphatidic acid sensors and their localization during frustrated phagocytosis. *J Biol Chem* (2017) 292:4266–79. doi:10.1074/jbc.M116.742346
64. Ghosh J, Lun CM, Majeske AJ, Sacchi S, Schrankel CS, Smith LC. Invertebrate immune diversity. *Dev Comp Immunol* (2011) 35:959–74. doi:10.1016/j.dci.2010.12.009
65. Schatz DG, Oettinger MA, Baltimore D. The V(D)J recombination activating gene, RAG-1. *Cell* (1989) 59:1035–48. doi:10.1016/0092-8674(89)90760-5
66. Fugmann SD. The origins of the RAG genes – from transposition to V(D)J recombination. *Semin Immunol* (2010) 22:10–6. doi:10.1016/j.smim.2009.11.004
67. Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An ancient evolutionary origin of the RAG1/2 gene locus. *Proc Natl Acad Sci U S A* (2006) 103:3728–33. doi:10.1073/pnas.0509720103
68. Hencken CG, Li X, Craig NL. Functional characterization of an active Rag-like transposase. *Nat Struct Mol Biol* (2012) 19:834–6. doi:10.1038/nsmb.2338
69. Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, et al. Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell* (2016) 166:102–14. doi:10.1016/j.cell.2016.05.032
70. Carmona LM, Fugmann SD, Schatz DG. Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. *Genes Dev* (2016) 30:909–17. doi:10.1101/gad.278432.116

Conflict of Interest Statement: The authors declare that they have no conflicts of interest and that writing this review was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Smith and Lun. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution and Function of Thioester-Containing Proteins and the Complement System in the Innate Immune Response

Upasana Shokal and Ioannis Eleftherianos*

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

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Simon John Clark,
University of Manchester,
United Kingdom
Lubka T. Roumenina,
INSERM UMRS 1138,
France

*Correspondence:

Ioannis Eleftherianos
ioannise@gwu.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 12 May 2017

Accepted: 16 June 2017

Published: 29 June 2017

Citation:

Shokal U and Eleftherianos I (2017)
Evolution and Function of Thioester-
Containing Proteins and the
Complement System in the Innate
Immune Response.
Front. Immunol. 8:759.
doi: 10.3389/fimmu.2017.00759

The innate immune response is evolutionary conserved among organisms. The complement system forms an important and efficient immune defense mechanism. It consists of plasma proteins that participate in microbial detection, which ultimately results in the production of various molecules with antimicrobial activity. Thioester-containing proteins (TEPs) are a superfamily of secreted effector proteins. In vertebrates, certain TEPs act in the innate immune response by promoting recruitment of immune cells, phagocytosis, and direct lysis of microbial invaders. Insects are excellent models for dissecting the molecular basis of innate immune recognition and response to a wide range of microbial infections. Impressive progress in recent years has generated crucial information on the role of TEPs in the antibacterial and antiparasite response of the tractable model insect *Drosophila melanogaster* and the mosquito malaria vector *Anopheles gambiae*. This knowledge is critical for better understanding the evolution of TEPs and their involvement in the regulation of the host innate immune system.

Keywords: insects, mammals, innate immunity, thioester-containing proteins, complement system, *Drosophila*, mosquito

INTRODUCTION

Innate immunity is a fundamental process for early recognition and subsequent induction of proinflammatory responses against invading pathogens (1). Insects are outstanding models for studying innate immune functions and host–pathogen interactions (2–4). Insects activate a variety of innate immune responses depending upon the type of pathogen they encounter. The cell signaling machinery involved in the insect innate immune response is structurally and functionally similar to innate immune pathways in mammals (5, 6). Previous and recent research involving infections with bacterial and fungal pathogens has led to the identification and characterization of two distinct immune pathways, the toll pathway [similar to mammalian IL-1/TLR pathway (7)] and the Immune deficiency pathway [Imd, similar to mammalian TNF- α R signaling pathway (8)], which regulate NF- κ B transcription factors that control the expression of several antimicrobial peptide (AMP) coding mainly in the fat body tissue (9). In addition, the Janus kinase/signal transducer and activator of transcription (JAK/STAT) and c-Jun N-terminal kinase (JNK) signaling pathways also act in either competing or cooperative modes to modulate the activity of immune effector genes (10, 11).

Insects utilize germ line-encoded receptors known as pathogen recognition receptors (PRRs) to identify distinct pathogen-associated molecular patterns (PAMPs) that are either present on the

surface of microbial pathogens or are released in the host during the infection (12). Insect PRRs are classified into three classes—secreted, endocytic, and signaling (13). A special class of signaling PRRs in the fruit fly *Drosophila melanogaster* is the peptidoglycan recognition proteins (PGRPs) (14). PGRP-SA and PGRP-SD bind to Gram-positive bacteria and activate a protease cascade that induces the toll signaling pathway (15, 16). PGRP-LE and PGRP-LC recognize DAP-type peptidoglycan structures present on the Gram-negative bacteria (17). To identify fungal pathogens, PRRs such as Gram-negative binding protein-3 target the β -(1,3)-glucan structure present on the fungal cell wall (18). Binding of these proteins to their molecular targets results in downstream activation of the NF- κ B signaling pathways Imd and toll (19). In addition to the signaling PRRs, insect genomes also contain secreted recognition molecules such as the thioester-containing proteins (TEPs), named after their active site that functions by forming covalent bonds with specific molecular targets (20). This mini review describes the complement proteins in mammals and the participation of TEPs in the immune response of mosquitoes and flies.

THIOESTER-CONTAINING PROTEINS

Members of the TEPs family have been recognized in primitive Protostomes and in Deuterostomes, ranging from *C. elegans* to mammals. TEPs contain a thioester (TE) motif, GCGEQ, which includes a highly unstable covalent bond between the side groups of cysteine and nearby glutamic acid (21). These proteins remain inactive in the native state due to a shielded environment within the protein, but when they encounter elevated temperature, aqueous conditions, or undergo proteolytic activation; the TE bond becomes active for a very short time (22–24). The active TE motif has the ability to bind to nearby accessible hydroxyl and amine groups that are present on all biological surfaces including pathogens (25). TEPs are classified into two subfamilies—complement factors and alpha-2 macroglobulins (α -2Ms). Once activated, the complement factors produce a small anaphylatoxin fragment lacking the TE motif and a larger fragment that binds to the target as a result of hydrolysis of the TE bond (20). The small anaphylatoxins act as immunoinflammatory stimulators and chemoattractants that recruit macrophages to the infection site. The larger, covalently bound fragment marks the pathogen as foreign and targets it for lysis or phagocytosis. In contrast, the α -2Ms inhibit the protease activity of pathogens *via* a conformational change that traps the attacking protease after linkage with the TE motif within the protein. This conformational change also exposes the receptor-binding domain of the α -2Ms that promotes receptor-mediated endocytosis for clearance of the pathogen through physical interaction with cell surface receptors (26). Hence, both complement factors and α -2Ms serve important functions in recognition as well as clearance of the pathogens from the host. Certain TEPs such as *Drosophila* TEP6, C5 in higher vertebrates, and ovostatin in mammals, contain a mutated TE motif (27). It has been further suggested that the presence of certain TEPs in the host could be an outcome of different environments, selective pressures, and perhaps gene duplications events (28, 29).

Functional characterization of TEPs in model organisms would shed light on their importance and specificity in the host.

COMPLEMENT PROTEINS IN MAMMALS

The complement system is an important effector that functions at the intersection of innate and adaptive immune responses in mammals. The system includes 50 germ line-encoded, circulating, and membrane-bound proteins. The activation of the complement system triggers a protease cascade that ends in opsonization and/or lysis of the pathogen. In addition to being pro-inflammatory, the complement proteins are also involved in homeostatic processes such as removal of dying cells with exposed danger-associated molecular patterns (DAMPs) that consequently generate a sterile inflammatory reaction (30, 31). In certain cases, activation of the complement cascade results in host tissue damage leading to autoimmune and chronic inflammatory diseases (32). Hence, host molecules closely control the activation and regulation of complement system.

The activation of complement system in mammals is regulated through three distinct pathways: the classical pathway, the lectin pathway, and the alternative pathway. Although these pathways have different ligands and receptors, they all converge to produce the same sets of effector molecules (33) (Figure 1A). The initiation of the classical pathway occurs upon binding of the collectin type PRR C1 complex (C1q multimers with inactive serine proteases C1r and C1s) to an antigen–antibody complex, to PAMPs, or to DAMPs (34–36). When C1q binds to PAMPs, a conformational change occurs in C1r and C1s complex, which results in autocatalytic activation of C1r. The activated C1r serine proteases then activate the C1s, which in turn cleave C4 and C2 molecules into the small anaphylatoxin C4a or C2b and the larger C4b or C2a, respectively. This exposes the activated TE within C4b, which binds covalently to the pathogen surface and recruits C2a to form the C4b2a complex. This newly formed complex on the pathogen surface is a C3-convertase that will perpetuate the cascade.

Similar to the classical pathway, the lectin pathway PRRs, either mannan-binding lectin (MBL) or ficolins L/M/H (ficolins-L or ficolins-M or ficolins-H) recognize specific sugars or acetylated moieties on the surfaces of Gram-positive bacteria, Gram-negative bacteria, fungi, protozoans, and viruses (37–39). The lectin pathway PRRs form complex with two MBL-associated serine proteases (MASP)-1 and MASP-2 that are structural homologs of C1r and C1s (40). Thus, MASP-1 and MASP-2 react and cleave C4 and C2 molecules to form the same C3-convertase, as described for the classical pathway. In contrast to the classical and lectin pathways, the alternative pathway does not require pathogen recognition proteins for its activation. Instead, it is initiated through spontaneous generation (also called tick-over mechanism) of short-lived C3(H_2O) by hydrolyzing the TE bond in the C3 molecule. This short-lived molecule binds to factor B in solution, which causes a conformational change in the structure of factor B. This leads to the cleavage of factor B into Ba and Bb fragments by factor D forming the C3(H_2O)Bb complex, which is the alternative pathway version of a C3-convertase, also called fluid-phase C3-convertase.

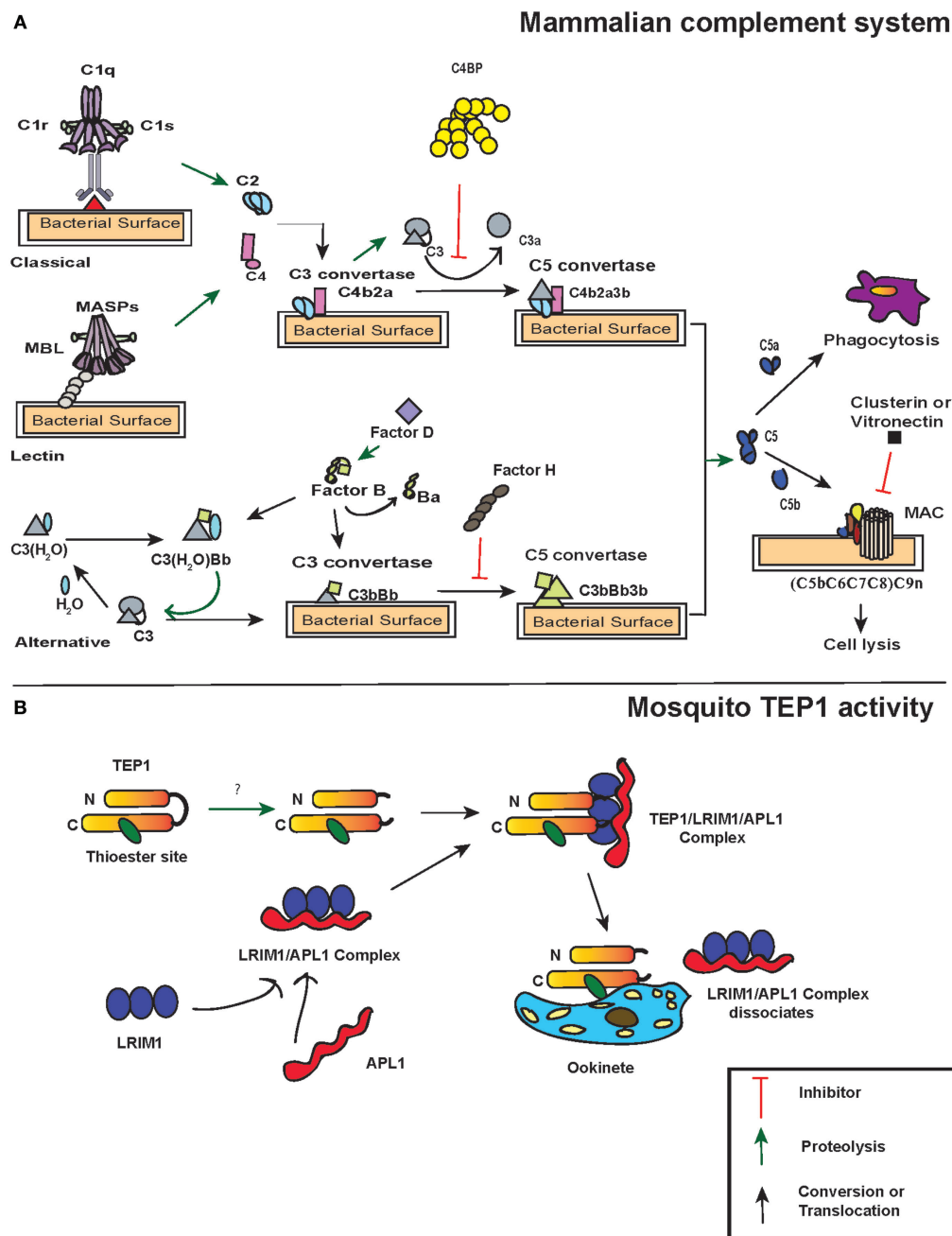


FIGURE 1 | Action of thioester-containing proteins (TEPs) in mammals and mosquitoes. **(A)** Pattern recognition receptors identify the presence of pathogens. In the classical pathway, C1 complex (specifically C1q) recognizes pathogen-associated molecular patterns or danger-associated molecular patterns either through binding to them directly or through binding of antibodies to the foreign antigen. This activates C1r that subsequently leads to C1s activation. Similarly, in the lectin pathway, binding of mannan-binding lectin (MBL) activates MBL-associated serine protease (MASP-1) and MASP-2. Activation of C1 and/or MBL–MASP complex leads to the cleavage of C4 and C2 molecules into C4a, C4b, C2a, and C2b, subsequently forming C3 convertase (C4b2a) that binds to the microbial surface. The newly formed C3 convertases cleave C3 into C3b that also binds to the microbial surface. Bound C3b recruits Factor D that activates Factor B, which results in the formation of C3bBb (C3 convertase of alternative pathway). C3bBb cleaves more C3 and initiates an amplification loop. Additionally, a fluid-phase convertase could also be formed when water associates with C3, forming C3(H₂O). The latter reacts with activated Factor B and thus maintains a low level of complement activation known as tick-over mechanism. The C3 convertases generated from each pathway bind to C3b forming C5 convertase, which cleaves C5 into C5a and C5b. The latter initiates the formation of membrane attacking membrane by recruiting C6, C7, C8, and C9 complement proteins. Certain molecules such as C4-binding protein, Factor H, vitronectin, and clusterin act as regulators of complement proteins. **(B)** TEP1 is constitutively activated in the hemolymph by one or more unknown proteases. The proteolytic cleavage produces two fragments TEP1-N and TEP1-C that remain associated with each other. Two leucine-rich repeats (LRRs) proteins, LRIM1 and APL1, maintain the mature form of TEP1. Upon recognition of the parasite, TEP1 dissociates from the LRR proteins by yet an unknown mechanism and binds to the parasite, which ultimately leads to its destruction. Arrows represent inhibition (red), proteolytic cleavage (green), and conversion or translocation of a molecule (black).

The C3-convertases produced by each of the three pathways generate C5-convertases upon binding C3b to C4b2b in the classical and lectin pathways yielding C4b2b3b. The alternative pathway C3-convertases can cleave many molecules of C3 into C3a and C3b. While most of the C3b is inactivated by hydrolysis, a fraction is able to link covalently to the PAMPs through the TE bond and form C3b2Bb (C5 convertases). The C5-convertases act on C5 and cleave it to C5a and C5b. C5a is released as an anaphylatoxin, and C5b recruits complement factors C6, C7, C8, and C9 that form the membrane-attack complex (MAC) in the cell membrane of the pathogen. While the larger fragment C5b plays a central role in MAC formation, the shorter C5a fragment acts on the endothelial or mast cells and increases the permeability of the blood vessels as well as extravasation of immunoglobulins to the site of inflammation. The activity of C5a causes a septic-shock state called anaphylactic shock and eventually triggers the inflammatory response. Together, these molecules assist in recognition, opsonization, and phagocytosis or lysis of pathogens, and are involved in the activation of adaptive immunity in vertebrates (41, 42) (**Figure 1A**).

The complement factors with TE motifs can also bind self-molecules containing accessible hydroxyl or amine groups on their surface. Therefore, to avoid false activation of the complement cascade in the absence of foreign entities, several complement regulatory proteins are present in mammals. One of the most potent and well-studied regulatory proteins is complement factor H that initiates the decay of the C3-convertase complex by dissociating Bb from C3b (43). Factor H competes with the Bb fragment and binds to C3(H₂O), which results in the dissociation of factor B from the latter. Moreover, it can bind host-specific glycans to prevent complement activation on host surfaces (44). Another regulator is the C4-binding protein (C4BP) that regulates the classic and lectin pathways with similar activities as factor H by targeting C4b and C2a (45). C4BP acts as a decay-accelerating factor and dissociates C2a from the C3-convertases. While these regulators control the formation of C3-convertase, other complement regulators such as clusterin and vitronectin inhibit MAC assembly or C9 insertion into membranes after the formation of C3 convertase complex and activation of the terminal pathway (46, 47).

Although the complement system is extremely efficient in fighting and clearing pathogenic infections, certain bacterial and viral pathogens can evade this immune response (48, 49). They achieve this by escaping the complement action through binding to the complement inhibitors, which target active complement factors that interfere with MAC complex formation and mimic host surfaces (50–52).

TEPs IN INSECTS

Phylogenetic analysis of TEP-coding genes in dipteran insects, other invertebrates, and vertebrate animals has classified them into three subfamilies including complement factors, α -2Ms, and insect TEPs (20) (**Figure 2A**). The complement factor subgroup containing C3, C4, and C5 proteins is the most fast-evolved TEP subfamily. On the other hand, the α -2Ms are present in a larger group of animals other than the two subfamilies, which suggests

their slow evolution due to several functional constraints on the structure of these inhibitors (53). Insect TEPs are highly diverged as well as unstable, and they are more related to the α -2M family than to the TE complement factor group (20) (**Figure 2A**). The presence of multiple TEP homologs in mosquitoes relative to those in *Drosophila* indicates that different adaptations between these insects have led to gene duplication and the generation of more homologs (54). It is currently unknown whether regulators of TEPs, such as homologs of human C4BP or Factor H, in insects exist. Interestingly, mosquitoes can capture Factor H from ingested human blood to escape the deleterious effects of the complement activation system (55). Although there is high structural and functional homology between TEPs and complement proteins, it is unclear whether insect TEPs possess a mechanism of action similar to C3 tick over. Here, we summarize TEPs in mosquitoes and fruit flies.

Mosquito TEPs

Genome sequencing of two mosquito species, *Anopheles gambiae* and *Aedes aegypti*, has contributed toward understanding several molecular mechanisms involved in host immunity. Various components of the complement pathway, specifically, complement-like proteins have been identified in the two genomes. The *A. gambiae* genome contains 19 TEP gene homologs (*AgTep* 1–19), of which four pairs show haplotypic features (*AgTep*1–*AgTep*16, *AgTep*5–*AgTep*17, *AgTep*6–*AgTep*18, and *AgTep*7–*AgTep*19) and hence, represent polymorphic variations rather than distinct genes (54, 56). There are eight *Tep* genes in *A. aegypti* (*AeTep* 1–8) encoding TEP proteins that share 21–39% amino acid similarity to *AgTep*1 (57, 58). In addition, the mosquito TEPs share structural and functional similarities with mammalian α -2Ms (29, 59).

A key immune gene identified through functional studies in *A. gambiae* was *AgTep*1. *AgTep*1 is a constitutively secreted hemolymph protein with a size of 165 kDa (*TEP*1-full) and its cleavage results in the formation of an 80 kDa active fragment (*TEP*1-cut) (60). While the N-terminal region of *AgTep*1 has a hydrophobic signal peptide-like segment as well as a canonical TE motif plus a catalytic histidine residue that is positioned 100 amino acids downstream, the C-terminal region has a cysteine signature. The *TEP*1-cut circulates in the hemolymph in association with two leucine-rich repeats (LRRs) proteins, LRIM1 and APL1C (61, 62). These two LRR proteins act as TEPs regulators and promote pathogen recognition as well as their destruction (**Figure 1B**).

Several studies have shown the functional importance of *AgTep*1 in various processes such as recognition, opsonization, and phagocytosis of certain bacteria. *In vitro* and *in vivo* studies have shown that bacteria are phagocytosed when the C-terminal part of *AgTep*1, also called *AgTep*1-cut, binds to bacteria (63). Moreover, knockdown of *AgTep*1 or culturing of hemocytes with methylamine-treated conditioned medium (prevents autocatalytic fragmentation of the full-length protein into smaller 80 kDa cut fragment) reduced the efficiency of phagocytosis of Gram-negative bacteria by 50–75% (63). Another study also showed decrease in phagocytosis of *Escherichia coli* and *Staphylococcus aureus* after the depletion of *AgTep*1 expression (62). Thus,

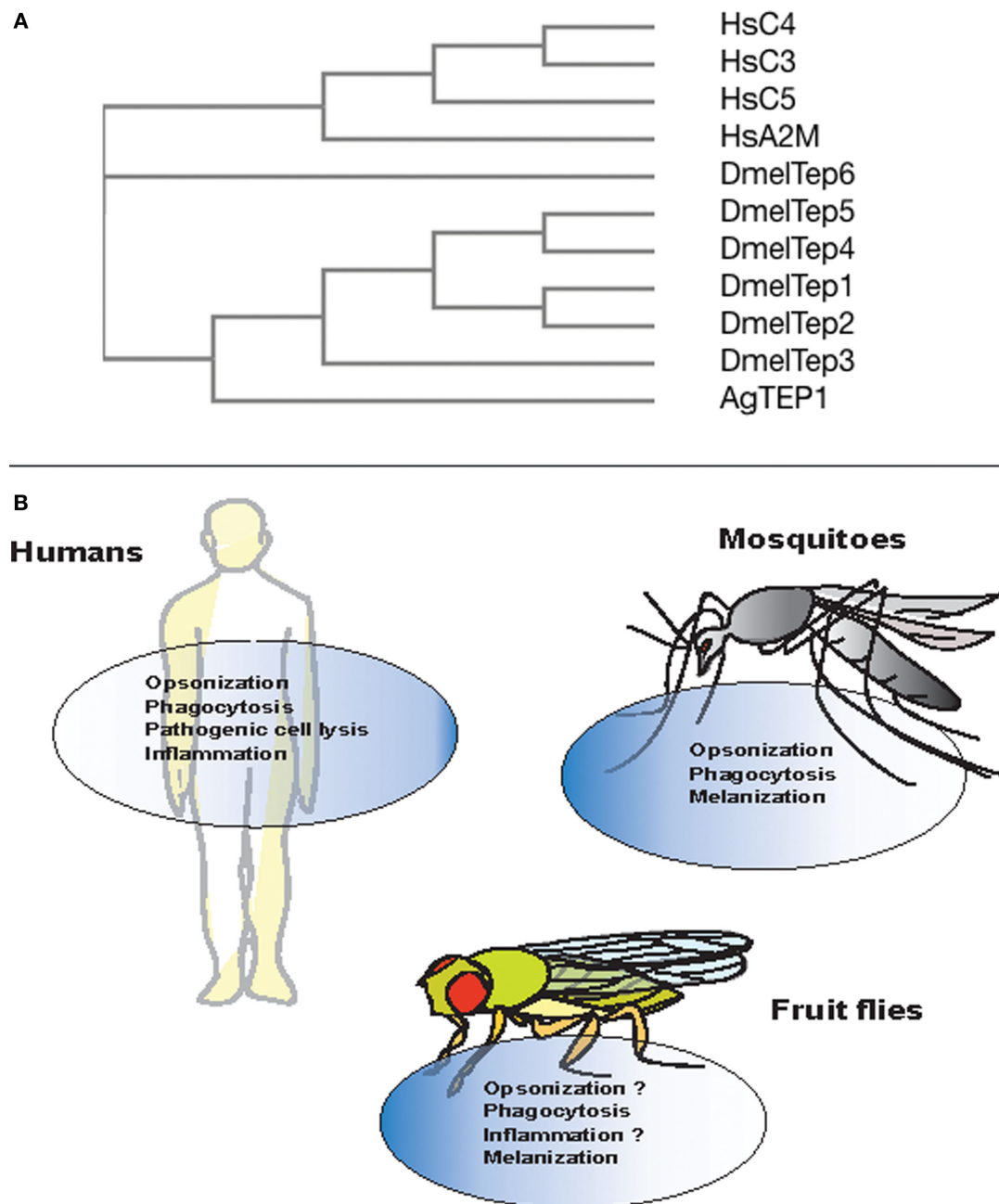


FIGURE 2 | Evolution and conserved function of thioester-containing proteins (TEPs) in mammals, mosquitoes, and fruit flies. **(A)** Phylogenetic tree showing the evolution and similarity between different complement factors, alpha-2 macroglobulin in *Homo sapiens* (Hs), and TEPs in *Anopheles gambiae* (Ae) and *Drosophila melanogaster* (Dm), generated through Clustal-Omega. **(B)** Complement proteins are involved in immune functions such as opsonization, phagocytosis, cell lysis of pathogens, and act as pro-inflammatory molecules. Similarly in mosquitoes, TEP molecules participate in the opsonization, phagocytosis, and melanization of bacterial pathogens and *Plasmodium* parasites. However, with the exception of phagocytosis and melanization, immune anti-pathogen activities of TEPs are yet to be identified in the fly.

AgTEP1 acts as an opsonin and marks targeted bacteria for phagocytosis (**Figure 2B**).

The complement C3-like protein, AgTEP1, is an important molecule in inducing an immune response against *Plasmodium berghei*. The protein binds to the surface of the parasite and triggers its encapsulation by hemocytes, which leads to parasite

death. Two divergent alleles of *AgTep1*-*AgTep1r* and *AgTep1s* are reported (64). While *AgTep1s* is present in most mosquito populations making them susceptible to *P. berghei* infection, the allele *AgTep1r* confers high resistance to the same parasite (65). Moreover, silencing of *AgTep1* also inhibits parasitic lysis and actin polymerization (66). The regulatory molecules LRIM1 and

APL1 are also required for binding of AgTEP1 to the parasites (61). The LRIM1 and APL1 complex not only interacts with AgTEP1 but also interferes with three other TEP proteins, including AgTEP3 (64). Silencing the two genes encoding the LRR proteins results in the conversion of refractory strains into susceptible strains (61, 63).

Although several functional studies have been performed on AgTEPs, only a few studies have characterized the specific function of *A. aegypti* TEPs. The genes *AeTep1–AeTep5* are all constitutively expressed throughout the body of adult mosquitoes. A study on *A. aegypti* has shown a twofold to threefold increase in West Nile virus load after silencing the *AeTep1* and *AeTep2* genes whereas overexpression of *AeTep1* and *AeTep3* resulted in a decrease in viral load (58). Thus, AeTEPs has an important function in the mosquito host defense by limiting viral infection.

The mosquito TEPs have been found to possess conserved function similar to complement factors by binding to bacteria or *Plasmodium* parasites, which results in the phagocytosis of bacteria as well as melanization and lysis of the parasites, respectively (59) (**Figure 2B**). Future studies will focus on investigating the binding specificity of mosquito TEPs and how the binding process leads to parasite lysis at the molecular level.

Fruit Fly TEPs

In insects, TEPs were first discovered in *Drosophila melanogaster* (67). While there is a plethora of information on mosquito TEPs, there are only few studies on the immune function of TEPs in *Drosophila*. The *D. melanogaster* genome contains six TEP homologs (68). TEP1–4 contain a conserved TE motif. *Tep5* may represent a pseudogene as it is found in the genomic sequences but is not expressed (13). *Drosophila* TEPs have a highly conserved region of 30 amino acids that compose the N-terminal of the TE motif, a cysteine signature tail, which is similar to *Anopheles* TEPs. They also have a 60 amino acid hypervariable region, which is structurally similar to mammalian bait region of α -2Ms as well as to the anaphylatoxin domain in vertebrate C3b (67). TEP6 is the only TEP that lacks a functional TE motif, exhibiting a serine instead of cysteine residue. Of the six *Tep* genes, only *Tep2* shows alternative splicing in exon 5 producing five different isoforms. The alternative splicing occurs in the exon region that codes for the hypervariable domain of the TEP2 protein. The alternative splicing may aid in increasing the inhibitor proteases repertoire and augmented diversity of recognition receptors. Flies may have evolved a strategy to encounter distinct pathogens that is analogous to VDJ diversity generated by adaptive immunity in higher vertebrates (69).

Drosophila melanogaster Teps are upregulated in different tissues and participate in immune response and developmental processes. *Teps* are expressed in larval hemocytes, fat body, and in the gut barrier epithelia, whereas, in the case of adults, *Teps* are expressed in the fat body of the head, spermatozoa, and midgut epithelia in the absence of infection (70). Upon bacterial challenge, *Tep1*, *Tep2*, and *Tep4* are upregulated in *D. melanogaster* larvae, whereas only *Tep1*, *Tep2*, *Tep4*, and *Tep6* are upregulated in adults in response to certain bacterial, fungal, or parasitoid infection (67, 70–72). Additionally, *Tep2* and *Tep3* are upregulated against

parasitic infections with the nematode *Heterorhabditis bacteriophora* that contains the mutualistic bacteria *Photorhabdus luminescens* (73). Loss-of-function *tep2* and *tep4* mutants are susceptible to *Pseudomonas ginigivalis* infection whereas loss-of-function *tep3* mutants are susceptible to *H. bacteriophora* infection (74, 75). Another study reported that *tep1–4* mutant flies were slightly resistant to bacterial infection in comparison to wild-type flies (69). Although these studies reported the involvement of fly TEPs in the antibacterial and antiparasitic immune response, the mechanism of TEPs action was not clarified. More recently, it was shown that TEP2, TEP4, and TEP6 has an important regulatory role in the innate immune response of *D. melanogaster* adult flies against the pathogenic bacteria *Photorhabdus* (72, 76). *Tep2*, *Tep4*, and *Tep6* are transcriptionally upregulated in response to *P. luminescens* and *P. asymbiotica* infection. Moreover, transcriptional activation of these genes influences the activation of toll, Imd, JAK/STAT, and JNK signaling and results in differential expression of AMP and stress coding genes. *Tep2* and *Tep4* upregulation also decreases phenoloxidase activity and the melanization response during the early stages of *Photorhabdus* infection. As a result, these effects promote the survival of flies upon infection with pathogenic *Photorhabdus*. This is the first evidence of the involvement of a TEP in the fly antibacterial immune system.

In vitro, *D. melanogaster* TEP2, TEP4, and TEP6 (MCR or macroglobulin-complement related) promote phagocytosis of certain Gram-negative bacteria and fungal pathogens (77). The rate of phagocytosis in *D. melanogaster* S2 cells incubated with *Candida albicans* decreases upon *Mcr* silencing because the MCR protein binds specifically to the fungal surface. Moreover, the rate of *E. coli* and *S. aureus* phagocytosis is reduced after silencing *Tep2* and *Tep4* genes. In addition, inactivation of *Tep2* and *Tep6* significantly impairs the expression of *Eater* gene in adult flies suggesting that TEP2 and TEP6 participate in the phagocytic response against *Photorhabdus* bacteria (72) (**Figure 2B**). This suggests that different TEP molecules are involved in the immune response and probably recognition of different pathogens. It has been suggested previously that JAK/STAT and toll pathways regulate the expression of TEP1, but the mechanisms are poorly understood (67, 78).

CONCLUDING REMARKS AND FUTURE PROSPECTS

Recent efforts have mostly focused on understanding the molecular and genetic mechanisms that regulate the participation of TEPs in interfering with the transmission of eukaryotic parasites and activating innate immune responses against pathogenic infections in insects (76, 79–82). Future studies could potentially examine the tissue-specific patterns of induction of insect *Tep* genes upon infection with different pathogenic and non-pathogenic microorganisms. Tissue-specific profiling of *Tep* gene expression would possible denote their specificity toward certain microbial infections. For example, the upregulation of *Tep* genes in the fat body, gut, or hemocytes upon microbial challenge would indicate their involvement in the insect humoral and/or cellular immune response to microbial invaders. Indeed, complement proteins

are involved in the activation the humoral immune response in invertebrates and vertebrates. In mammals, complement factors are involved in the regulation of humoral immune responses (83). In insects, complement-related factors participate in the upregulation of AMPs against flavivirus infection in the mosquito, *A. aegypti* (84). Recently, it has been proposed that macrocapsules loaded with α -2Ms enhance certain human leukocyte functions, such as the recruitment of leukocytes to the site of inflammation and phagocytosis (85). Although TEP1 is involved in opsonization and phagocytosis of certain bacteria in mosquitoes (20, 26), other TEP molecules, might also participate directly or indirectly in insect cellular immune processes.

In addition in mammals, there is an intricate cross talk between the complement system and the coagulation cascade (86). Within certain hours of pathogenic infection, both of these systems are activated through the activity of serine proteases (87). Likewise, the coagulation system and the phenoloxidase cascade are linked in insects (88). It has been shown that *A. gambiae* TEP1 is essential in the process of melanization of *Plasmodium* parasites (6), and phenoloxidase activity as well as the melanization response are affected in *Drosophila* flies inactivated for *tep2* and *tep4* genes when responding to the pathogen *Photothabdus* (72, 76) (Figure 2B). Future research could concentrate on the identification of the molecular components that facilitate the interaction between complement and coagulation systems in vertebrates and invertebrates.

REFERENCES

- Medzhitov R, Janeway C Jr. Innate immunity. *N Engl J Med* (2000) 343(5):338–44. doi:10.1056/NEJM200008033430506
- Limmer S, Quintin J, Hetru C, Ferrandon D. Virulence on the fly: *Drosophila melanogaster* as a model genetic organism to decipher host-pathogen interactions. *Curr Drug Targets* (2011) 12(7):978–99. doi:10.2174/138945011795677818
- Buchon N, Silverman N, Cherry S. Immunity in *Drosophila melanogaster* – from microbial recognition to whole-organism physiology. *Nat Rev Immunol* (2014) 14(12):796–810. doi:10.1038/nri3763
- Wojda I. Immunity of the greater wax moth *Galleria mellonella*. *Insect Sci* (2017) 24(3):342–57. doi:10.1111/1744-7917.12325
- Vilmos P, Kurucz E. Insect immunity: evolutionary roots of the mammalian innate immune system. *Immunol Lett* (1998) 62(2):59–66. doi:10.1016/S0165-2478(98)00023-6
- Viljakainen L. Evolutionary genetics of insect innate immunity. *Brief Funct Genomics* (2015) 14(6):407–12. doi:10.1093/bfpg/elv002
- Lindsay SA, Wasserman SA. Conventional and non-conventional *Drosophila* toll signaling. *Dev Comp Immunol* (2014) 42(1):16–24. doi:10.1016/j.dci.2013.04.011
- Myllymäki H, Valanne S, Rämet M. The *Drosophila* IMD signaling pathway. *J Immunol* (2014) 192(8):3455–62. doi:10.4049/jimmunol.1303309
- Yi HY, Chowdhury M, Huang YD, Yu XQ. Insect antimicrobial peptides and their applications. *Appl Microbiol Biotechnol* (2014) 98(13):5807–22. doi:10.1007/s00253-014-5792-6
- Zeidler MP, Bausek N. The *Drosophila* JAK-STAT pathway. *JAKSTAT* (2013) 2(3):e25353. doi:10.4161/jkst.25353
- Zhou YY, Li Y, Jiang WQ, Zhou LE. MAPK/JNK signalling: a potential autophagy regulation pathway. *Biosci Rep* (2015) 35(3):e00199. doi:10.1042/BSR20140141
- Pal S, Wu LP. Pattern recognition receptors in the fly: lessons we can learn from the *Drosophila melanogaster* immune system. *Fly* (2009) 3(2):121–9. doi:10.4161/fly.8827
- Stokes BA, Yadav S, Shokal U, Smith LC, Eleftherianos I. Bacterial and fungal pattern recognition receptors in homologous innate signaling pathways of insects and mammals. *Front Microbiol* (2015) 6:19. doi:10.3389/fmicb.2015.00019
- Kurata S. Peptidoglycan recognition proteins in *Drosophila* immunity. *Dev Comp Immunol* (2014) 42(1):36–41. doi:10.1016/j.dci.2013.06.006
- Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D. A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* (2000) 97:13772–7. doi:10.1073/pnas.97.25.13772
- 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
- Takehana A, Yano T, Mita S, Kotani A, Oshima Y, Kurata S. Peptidoglycan recognition protein (PGRP)-LE and PGRP-LC act synergistically in *Drosophila* immunity. *EMBO J* (2004) 23:4690–700. doi:10.1038/sj.emboj.7600466
- Gottar M, Gobert V, Matskevich AA, Reichhart JM, 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
- Ferrandon D, Imler JL, Hetru C, Hoffmann JA. The *Drosophila* systemic immune response: sensing and signaling during bacterial and fungal infections. *Nat Rev Immunol* (2007) 7(11):862–74. doi:10.1038/nri2194
- Blandin S, Levashina EA. Thioester-containing proteins and insect immunity. *Mol Immunol* (2004) 40(12):903–8. doi:10.1016/j.molimm.2003.10.010
- Law SK, Dodds AW. The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci* (1997) 6(2):263–74. doi:10.1002/pro.5560060201
- Salvesen GS, Sayers CA, Barrett AJ. Further characterization of the covalent linking reaction of alpha 2-macroglobulin. *Biochem J* (1981) 195(2):453–61. doi:10.1042/bj1950453
- Sim RB, Sim E. Autolytic fragmentation of complement components C3 and C4 under denaturing conditions, a property shared with alpha 2-macroglobulin. *Biochem J* (1981) 193(1):129–41. doi:10.1042/bj1930129

Complement proteins are involved in the inflammation process and programmed cell death in vertebrates (89, 90). The presence of complement serves a protective function in vertebrates, but complement activation can also be deleterious for the host (91). Deletion in C5a confers resistance and reduced bacteremia shock in mice in response to Gram-negative bacterial infection (92). Identification of TEPs in insects with function analogous to C5a in mammals or relevance to pathophysiological defects in the host offers an exciting and challenging area of future research. In conclusion, future studies on elucidating the molecular mechanisms of interaction of TEPs with specific host physiological processes will undoubtedly shed light on their exact anti-pathogen immune function as well as their evolution in the animal kingdom.

AUTHOR CONTRIBUTIONS

US wrote the paper and IE revised it.

FUNDING

The Eleftherianos laboratory is funded by grants from the National Institutes of Health—National Institute of Allergy and Infectious Diseases (1R01AI110675, 1R56AI110675-01, and 1R21AI109517) and the Columbian College of Arts and Sciences at George Washington University.

24. Sahu A, Kozel TR, Pangburn MK. Specificity of the thioester-containing reactive site of human C3 and its significance to complement activation. *Biochem J* (1994) 302(2):429–36. doi:10.1042/bj3020429
25. Dodds AW, Ren XD, Willis AC, Law SK. The reaction mechanism of the internal thioester in the human complement component C4. *Nature* (1996) 379(6561):177–9. doi:10.1038/379177a0
26. Blandin SA, Marois E, Levashina EA. Antimalarial responses in *Anopheles gambiae*: from a complement-like protein to a complement-like pathway. *Cell Host Microbe* (2008) 3(6):364–74. doi:10.1016/j.chom.2008.05.007
27. Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, et al. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* (2005) 437(7058):505–11. doi:10.1038/nature04005
28. Nonaka M, Yoshizaki F. Primitive complement system of invertebrates. *Immunol Rev* (2004) 198:203–15. doi:10.1111/j.0105-2896.2004.00118.x
29. Nonaka M, Kimura A. Genomic view of the evolution of the complement system. *Immunogenetics* (2006) 58(9):701–13. doi:10.1007/s00251-006-0142-1
30. Rock KL, Latz E, Ontiveros F, Kono H. The sterile inflammatory response. *Ann Rev Immunol* (2010) 28:321–42. doi:10.1146/annurev-immunol-030409-101311
31. Carroll MC, Isenman DE. Regulation of humoral immunity by complement. *Immunity* (2012) 37(2):199–207. doi:10.1016/j.immuni.2012.08.002
32. Dunkelberger JR, Song WC. Complement and its role in innate and adaptive immune responses. *Cell Res* (2010) 20(1):34–50. doi:10.1038/cr.2009.139
33. Dodds AW. Which came first, the lectin/classical pathway or the alternative pathway of complement? *Immunobiology* (2002) 205(4–5):340–54. doi:10.1078/0171-2985-00137
34. Arlaud GJ, Gaboriaud C, Thielens NM, Rossi V, Bersch B, Hernandez JF, et al. Structural biology of C1: dissection of a complex molecular machinery. *Immunol Rev* (2001) 180:136–45. doi:10.1034/j.1600-065X.2001.1800112.x
35. Nayak A, Pednekar L, Reid KB, Kishore U. Complement and non-complement activating functions of C1q: a prototypical innate immune molecule. *Innate Immun* (2012) 18(2):350–63. doi:10.1177/1753425910396252
36. Ramirez-Ortiz ZG, Pendergraft WF III, Prasad A, Byrne MH, Iram T, Blanchette CJ, et al. The scavenger receptor SCARF1 mediates the clearance of apoptotic cells and prevents autoimmunity. *Nat Immunol* (2013) 14(9):917–26. doi:10.1038/ni.2670
37. Degn SE, Thiel S. Humoral pattern recognition and the complement system. *Scand J Immunol* (2013) 78(2):181–93. doi:10.1111/sji.12070
38. Kjaer TR, Thiel S, Andersen GR. Toward a structure-based comprehension of the lectin pathway of complement. *Mol Immunol* (2013) 56(4):413–22. doi:10.1016/j.molimm.2013.05.007
39. Matsushita M. Ficolins in complement activation. *Mol Immunol* (2013) 55(1):22–6. doi:10.1016/j.molimm.2012.08.017
40. Bajic G, Degn SE, Thiel S, Andersen GR. Complement activation, regulation, and molecular basis for complement-related diseases. *EMBO J* (2015) 34(22):2735–57. doi:10.15252/embj.201591881
41. Carroll MC. The complement system in regulation of adaptive immunity. *Nat Immunol* (2004) 5(10):981–6. doi:10.1038/ni1113
42. Morgan BP, Marchbank KJ, Longhi MP, Harris CL, Gallimore AM. Complement: central to innate immunity and bridging to adaptive responses. *Immunol Lett* (2005) 97(2):171–9. doi:10.1016/j.imlet.2004.11.010
43. Ferreira VP, Pangburn MK, Cortes C. Complement control protein factor H: the good, the bad, and the inadequate. *Mol Immunol* (2010) 47(13):2187–97. doi:10.1016/j.molimm.2010.05.007
44. Makou E, Herbert AP, Barlow PN. Functional anatomy of complement factor H. *Biochemistry* (2013) 52(23):3949–62. doi:10.1021/bi4003452
45. Noris M, Remuzzi G. Overview of complement activation and regulation. *Semin Nephrol* (2013) 33(6):479–92. doi:10.1016/j.semnephrol.2013.08.001
46. Tschopp J, French LE. Clusterin: modulation of complement function. *Clin Exp Immunol* (1994) 97(Suppl 2):11–4. doi:10.1111/j.1365-2249.1994.tb06256.x
47. Koch-Brandt C, Morgans C. Clusterin: a role in cell survival in the face of apoptosis? *Prog Mol Subcell Biol* (1996) 16:130–49. doi:10.1007/978-3-642-79850-4_8
48. Serruto D, Rappuoli R, Scarselli M, Gros P, van Strijp JA. Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. *Nat Rev Microbiol* (2010) 8:393–9. doi:10.1038/nrmicro2366
49. Rettig TA, Harbin JN, Harrington A, Dohmen L, Fleming SD. Evasion and interactions of the humoral innate immune response in pathogen invasion, autoimmune disease, and cancer. *Clin Immunol* (2015) 160:244–54. doi:10.1016/j.clim.2015.06.012
50. Zipfel PF, Würzner R, Skerka C. Complement evasion of pathogens: common strategies are shared by diverse organisms. *Mol Immunol* (2007) 44(16):3850–7. doi:10.1016/j.molimm.2007.06.149
51. Zipfel PF, Hallström T, Riesbeck K. Human complement control and complement evasion by pathogenic microbes – tipping the balance. *Mol Immunol* (2013) 56(3):152–60. doi:10.1016/j.molimm.2013.05.222
52. Hovingh ES, van den Broek B, Jongerius I. Hijacking complement regulatory proteins for bacterial immune evasion. *Front Microbiol* (2016) 7:2004. doi:10.3389/fmicb.2016.02004
53. Dodds AW, Law SK. The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and alpha 2-macroglobulin. *Immunol Rev* (1998) 166:15–26. doi:10.1111/j.1600-065X.1998.tb01249.x
54. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, et al. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* (2002) 298(5591):159–65. doi:10.1126/science.1077136
55. Khattab A, Barroso M, Miettinen T, Meri S. *Anopheles* midgut epithelium evades human complement activity by capturing factor H from the blood meal. *PLoS Negl Trop Dis* (2015) 9:e0003513. doi:10.1371/journal.pntd.0003513
56. Christophides GK, Vlachou D, Kafatos FC. Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*. *Immunol Rev* (2004) 198:127–48. doi:10.1111/j.0105-2896.2004.0127.x
57. Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, et al. Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* (2007) 316(5832):1738–43. doi:10.1126/science.1139862
58. Cheng G, Liu L, Wang P, Zhang Y, Zhao YO, Colpitts TM, et al. An in vivo transfection approach elucidates a role for *Aedes aegypti* thioester-containing proteins in flavivirus infection. *PLoS One* (2011) 6:e22786. doi:10.1371/journal.pone.0022786
59. Williams M, Baxter R. The structure and function of thioester-containing proteins in arthropods. *Biophys Rev* (2014) 6(3):261–72. doi:10.1007/s12551-014-0142-6
60. Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* (2001) 104(5):709–18. doi:10.1016/S0092-8674(01)00267-7
61. Fraiture M, Baxter RH, Steinert S, Chelliah Y, Floret C, Quispe-Tintaya W, et al. Two mosquito LRR proteins function as complement control factors in the TEP1-mediated killing of *Plasmodium*. *Cell Host Microbe* (2009) 5(3):273–84. doi:10.1016/j.chom.2009.01.005
62. Moita LF, Wang-Sattler R, Michel K, Zimmermann T, Blandin S, Levashina EA, et al. In vivo identification of novel regulators and conserved pathways of phagocytosis in *A. gambiae*. *Immunity* (2005) 23(1):65–73. doi:10.1016/j.immuni.2005.05.006
63. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK. Leucine-rich repeat protein complex activates mosquito complement in defense against *Plasmodium* parasites. *Science* (2009) 324(5924):258–61. doi:10.1126/science.1171400
64. Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, et al. Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* (2004) 116(5):661–70. doi:10.1016/S0092-8674(04)00173-4
65. Blandin SA, Wang-Sattler R, Lamacchia M, Gagneur J, Lycett G, Ning Y, et al. Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. *Science* (2009) 326(5949):147–50. doi:10.1126/science.1175241
66. Shiao SH, Whitten MM, Zachary D, Hoffmann JA, Levashina EA. Fz2 and cdc42 mediate melanization and actin polymerization but are dispensable for *Plasmodium* killing in the mosquito midgut. *PLoS Pathog* (2006) 2:e133. doi:10.1371/journal.ppat.0020133
67. Lagueux M, Perrodou E, Levashina EA, Capovilla M, Hoffmann JA. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. *Proc Natl Acad Sci U S A* (2000) 97:11427–32. doi:10.1073/pnas.97.21.11427

68. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, et al. The genome sequence of *Drosophila melanogaster*. *Science* (2000) 287(5461):2185–95. doi:10.1126/science.287.5461.2185
69. Bou Aoun R, Hetru C, Troxler L, Doucet D, Ferrandon D, Matt N. Analysis of thioester-containing proteins during the innate immune response of *Drosophila melanogaster*. *J Innate Immun* (2011) 3(1):52–64. doi:10.1159/000321554
70. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* (2007) 450(7167):203–18. doi:10.1038/nature06341
71. Wertheim B, Kraaijeveld AR, Schuster E, Blanc E, Hopkins M, Pletchers SD, et al. Genome-wide gene expression in response to parasitoid attack in *Drosophila*. *Genome Biol* (2005) 6:R94. doi:10.1186/gb-2005-6-11-r94
72. Shokal U, Kopydlowski H, Eleftherianos I. The distinct function of *Tep2* and *Tep6* in the immune defense of *Drosophila melanogaster* against the pathogen *Photographus*. *Virulence* (2017) 1–15. doi:10.1080/21505594.2017.1330240
73. Castillo JC, Creasy T, Kumari P, Shetty A, Shokal U, Tallon LJ, et al. *Drosophila* anti-nematode and antibacterial immune regulators revealed by RNA-Seq. *BMC Genomics* (2015) 16:519. doi:10.1186/s12864-015-1690-2
74. Igboin CO, Tordoff KP, Moeschberger ML, Griffen AL, Leys EJ. *Porphyromonas gingivalis*-host interactions in a *Drosophila melanogaster* model. *Infect Immun* (2011) 79:449–58. doi:10.1128/IAI.00785-10
75. Arefin B, Kucerova L, Dobes P, Markus R, Strnad H, Wang Z, et al. Genome-wide transcriptional analysis of *Drosophila* larvae infected by entomopathogenic nematodes shows involvement of complement, recognition and extracellular matrix proteins. *J Innate Immun* (2014) 6:192–204. doi:10.1159/000353734
76. Shokal U, Eleftherianos I. Thioester-containing protein-4 regulates the *Drosophila* immune signaling and function against the pathogen *Photographus*. *J Innate Immun* (2017) 9(1):83–93. doi:10.1159/000450610
77. Stroschein-Stevenson SL, Foley E, O'Farrell PH, Johnson AD. Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. *PLoS Biol* (2006) 4(1):e4. doi:10.1371/journal.pbio.0040004
78. Mathey-Prevot B, Perrimon N. Mammalian and *Drosophila* blood: JAK of all trades? *Cell* (1998) 92:697–700. doi:10.1016/S0092-8674(00)81396-3
79. Eldering M, Morlais I, van Gemert GJ, van de Vegte-Bolmer M, Graumans W, Siebelink-Stoter R, et al. Variation in susceptibility of African *Plasmodium falciparum* malaria parasites to TEP1 mediated killing in *Anopheles gambiae* mosquitoes. *Sci Rep* (2016) 6:20440. doi:10.1038/srep20440
80. Kamareddine L, Nakhleh J, Osta MA. Functional interaction between apolipoproteins and complement regulate the mosquito immune response to systemic infections. *J Innate Immun* (2016) 8(3):314–26. doi:10.1159/000443883
81. Matetovici I, Caljon G, Van Den Abbeele J. Tsetse fly tolerance to *T. brucei* infection: transcriptome analysis of trypanosome-associated changes in the tsetse fly salivary gland. *BMC Genomics* (2016) 17(1):971. doi:10.1186/s12864-016-3283-0
82. Wang YH, Hu Y, Xing LS, Jiang H, Hu SN, Raikhel AS, et al. A critical role for CLSP2 in the modulation of antifungal immune response in mosquitoes. *PLoS Pathog* (2015) 11(6):e1004931. doi:10.1371/journal.ppat.1004931
83. Carroll MC. Complement and humoral immunity. *Vaccine* (2008) 26(Suppl 8):I28–33. doi:10.1016/j.vaccine.2008.11.022
84. Xiao X, Liu Y, Zhang X, Wang J, Li Z, Pang X, et al. Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing anti-microbial peptides. *PLoS Pathog* (2014) 10:e1004027. doi:10.1371/journal.ppat.1004027
85. Federici Canova D, Pavlov AM, Norling LV, Gobetti T, Brunelleschi S, Le Fauder P, et al. Alpha-2-macroglobulin loaded microcapsules enhance human leukocyte functions and innate immune response. *J Control Release* (2015) 217:284–92. doi:10.1016/j.jconrel.2015.09.021
86. Amara U, Rittirsch D, Flierl M, Bruckner U, Klos A, Gebhard F, et al. Interaction between the coagulation and complement system. *Adv Exp Med Biol* (2008) 632:71–9.
87. Huber-Lang M, Sarma JV, Zetouni FS, Rittirsch D, Neff TA, McGuire SR, et al. Generation of C5a in the absence of C3: a new complement activation pathway. *Nat Med* (2006) 12:682–7. doi:10.1038/nm1419
88. Eleftherianos I, Revenis C. Role and importance of phenoloxidase in insect hemostasis. *J Innate Immun* (2011) 3:28–33. doi:10.1159/000321931
89. Montalto MC, Hart ML, Jordan JE, Wada K, Stahl GL. Role for complement in mediating intestinal nitric oxide synthase-2 and superoxide dismutase expression. *Am J Physiol Gastrointest Liver Physiol* (2003) 285:G197–206. doi:10.1152/ajpgi.00029.2003
90. Ricklin D, Hajishengallis G, Yang K, Lambris JD. Complement: a key system for immune surveillance and homeostasis. *Nat Immunol* (2010) 11:785–97. doi:10.1038/ni.1923
91. Tahtouh M, Croq F, Lefebvre C, Pestel J. Is complement good, bad, or both? New functions of the complement factors associated with inflammation mechanisms in the central nervous system. *Eur Cytokine Netw* (2009) 20:95–100. doi:10.1684/ecn.2009.0157
92. Hollmann TJ, Mueller-Ortiz SL, Braun MC, Wetsel RA. Disruption of the C5a receptor gene increases resistance to acute Gram-negative bacteremia and endotoxic shock: opposing roles of C3a and C5a. *Mol Immunol* (2008) 45:1907–15. doi:10.1016/j.molimm.2007.10.037

Conflict of Interest Statement: 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.

Copyright © 2017 Shokal and Eleftherianos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Immunity in Protochordates: The Tunicate Perspective

Nicola Franchi and Lorian Ballarin*

Department of Biology, University of Padova, Padova, Italy

Tunicates are the closest relatives of vertebrates, and their peculiar phylogenetic position explains the increasing interest toward tunicate immunobiology. They are filter-feeding organisms, and this greatly influences their defense strategies. The majority of the studies on tunicate immunity were carried out in ascidians. The tunic acts as a first barrier against pathogens and parasites. In addition, the oral siphon and the pharynx represent two major, highly vascularized, immune organs, where circulating hemocytes can sense non-self material and trigger immune responses that, usually, lead to inflammation and phagocytosis. Inflammation involves the recruitment of circulating cytotoxic, phenoloxidase (PO)-containing cells in the infected area, where they degranulate as a consequence of non-self recognition and release cytokines, complement factors, and the enzyme PO. The latter, acting on polyphenol substrata, produces cytotoxic quinones, which polymerize to melanin, and reactive oxygen species, which induce oxidative stress. Both the alternative and the lectin pathways of complement activation converge to activate C3: C3a and C3b are involved in the recruitment of hemocytes and in the opsonization of foreign materials, respectively. The interaction of circulating professional phagocytes with potentially pathogenic foreign material can be direct or mediated by opsonins, either complement dependent or complement independent. Together with cytotoxic cells, phagocytes are active in the encapsulation of large materials. Cells involved in immune responses, collectively called immunocytes, represent a large fraction of hemocytes, and the presence of a cross talk between cytotoxic cells and phagocytes, mediated by secreted humoral factors, was reported. Lectins play a pivotal role as pattern-recognition receptors and opsonizing agents. In addition, variable region-containing chitin-binding proteins, identified in the solitary ascidian *Ciona intestinalis*, control the settlement and colonization of bacteria in the gut.

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Stefano Fiorucci,
University of Perugia, Italy
Taruna Madan,
National Institute for Research in
Reproductive Health, India

*Correspondence:

Lorian Ballarin
loriano.ballarin@unipd.it

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 21 March 2017

Accepted: 24 May 2017

Published: 09 June 2017

Citation:

Franchi N and Ballarin L (2017)
Immunity in Protochordates: The
Tunicate Perspective.
Front. Immunol. 8:674.
doi: 10.3389/fimmu.2017.00674

Keywords: tunicates, immune responses, complement, lectins, inflammation, chemical defense

INTRODUCTION

Tunicates or urochordates are marine, filter-feeding invertebrates, members of the phylum Chordata. They owe their name to the tunic that embeds the larval and adult body. Tunicates (ca 3,000 species) include Ascidiacea (benthic and sessile), Thaliacea (pelagic), and Larvacea or Appendicularia (pelagic).

Ascidians have a free-swimming, tadpole-like larva whereas adults have sac-like bodies with two siphons, allowing water flux, and a large branchial basket provided with a ventral endostyle secreting the mucous net required for filtration. They include Phlebobranchia, Aplousobranchia, and Stolidobranchia, previously grouped as Enterogona (Phlebobranchia and Aplousobranchia) and Pleurogona (Stolidobranchia).

Thaliaceans have barrel-like bodies; they include Pyrosomida (colonial), Doliolida (solitary/colonial), and Salpida (solitary/colonial). All Thaliaceans but Doliolida are devoid of larval stages. Larvaceans or appendicularians are similar to ascidian larvae, hence their name: they secrete a gelatinous house containing traps for food particles and use their tail to move water for filtration. Today, Larvaceans are considered a sister group of the remaining tunicates and Thaliaceans as a sister group of Enterogona (**Figure 1**).

Tunicates are the closest relatives to vertebrates (2), and this explains the increasing interest toward this group of animals. Like other invertebrates, tunicates rely only on innate immunity that lacks somatic recombination and long-term immune memory and has a limited array of effector responses.

Ascidians include about 2,300 species and are the most studied tunicates. Accordingly, the majority of the information on tunicate immune responses comes from studies on these organisms. In addition, ascidian innate immune genes did not undergo the expansions reported in other invertebrate deuterostomes, such as amphioxus and sea urchin (3, 4). This review, then, will focus mainly on the ascidian strategies of immune defense. Where

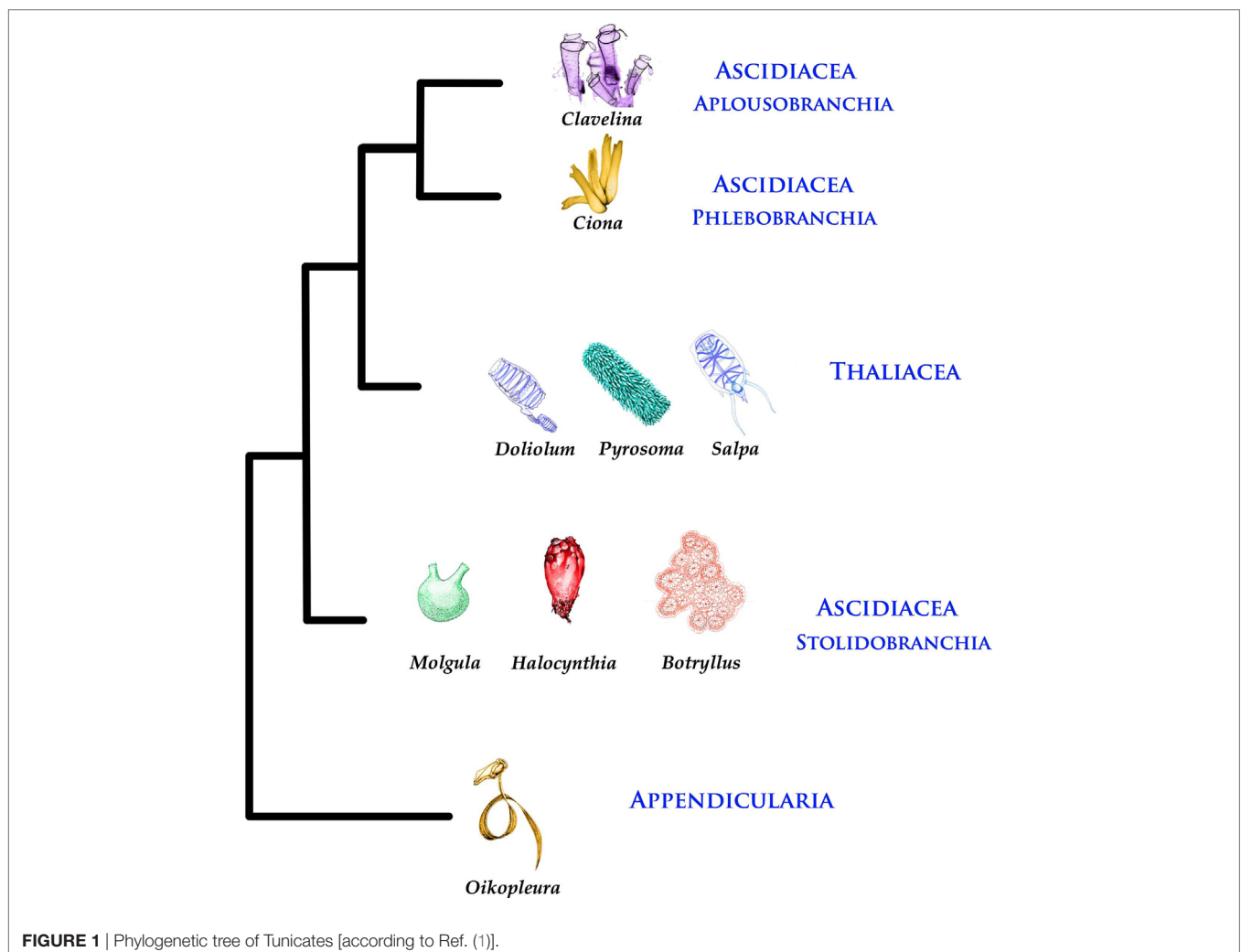
available, information on immune responses of pelagic tunicates will be added.

THE SITES OF IMMUNE RESPONSES

The marine habitat contains 10^5 – 10^6 microbes/ml in the water column and much more in the sediments (5); the amount of viruses is 10 times higher (6). Tunicates, therefore, require an efficient immune system in order to prevent the risk of infections and select appropriate mutualistic bacterial strains for gut colonization (see below). The sites where the ascidian immune system is alerted by the contact with non-self molecules include the tunic, the hemolymph, and the digestive tract.

Tunic

The tunic represents the first outpost against pathogens and parasites and its damage, as in the soft tunic syndrome, can lead the organism to death (7). It is mainly of epidermal origin and resembles the vertebrate connective tissue in consisting of an amorphous matrix containing fibers and interspersed cells (8). The tunic can contain spicules, acting as physical defense against



predators and varying in morphology, size, and mineral content (8). Molecules with antibacterial and anti-inflammatory activity are usually present in the matrix (9, 10). The tunic fibrous components include tunicin, a cellulose-like polysaccharide, collagen, and elastin (8, 11). Intermediate filaments (12) and mucopolysaccharides (9) contribute to the structural integrity of the tunic. The outermost compact layer, known as the cuticle, is continuous with the tunic matrix and frequently presents minute protrusions or spines (8, 13–15).

Tunic cells derive from both the epidermis and the hemocytes that can enter the tunic in response to infections (8, 16, 17). Hemocytes include spreading and round phagocytes, always present, cytotoxic granulocytes, widely found, and other cell types in some particular taxa, such as net cells and cells storing acid or pigments (8, 17–21), all contributing to protect the organism from predators, pathogens, or parasites. Phagocytes ingest foreign cells having entered the tunic (17, 22), and tunic phagocytes are the main effectors of allorecognition in the colonial species *Aplidium yamazii* (23). Granulocytes frequently contain and release antimicrobial peptides (24) and the enzyme phenoloxidase (PO) (25). Bladder cells store acid that, once released, decreases the pH of the tunic, disinfects the wounds, and exerts antifouling activity (17, 26, 27); net cells allow the shrinkage of the tunic in wound areas (17). PO-containing granulocytes can contribute to tunic formation or regeneration *via* degranulation and release of tunichromes, likely fragments of DOPA-containing proteins, that, once oxidized, cross-link tunicin fibers (28, 29). Tunic phagocytes and net cells are present also in Thaliaceans, although their role in defense has been poorly investigated. In pyrosomes, the density of tunic cells is comparable to that of ascidians (30), whereas doliolids and salps have a lower number of cells in their tunic (14, 17, 21). Larvaceans or appendicularians have no tunic, but tunicin is present in their house, secreted by specialized portions of the trunk epithelium (31).

Hemolymph

Ascidians have an open circulatory system and a colorless hemolymph, isotonic with seawater. The beating of a tubular heart guarantees the circulation in blood sinuses and lacunae. It periodically reverses the direction of the peristaltic waves thus inverting the hemolymph flow (8, 13). Circulating hemocyte types differ in morphology and ultrastructure. Various authors proposed unifying classification schemes [Figure 2; Table 1; references therein (32)], but uncertainties and doubts persist on terminology, hemocyte relationships, and differentiation pathways.

Ascidian hemocytes, involved in immune responses (immuno-cytes), represent a relevant fraction of circulating hemocytes (32), synthesize most of the pattern-recognition receptors (Table 2) and actively transcribe genes required for immune defense (60, 61): they include phagocytes and cytotoxic cells. Phagocytes are wandering, spreading cells that actively move toward foreign cells or particles and ingest them. Upon the ingestion of foreign material, phagocytes withdraw their projections and assume a round morphology. Spreading phagocytes can reach 20 μm in length and have a well-defined actin cytoskeleton, with abundance of stress fibers (25). They contain fine cytoplasmic granules, unresolvable under the light microscope, showing positivity for lysosomal enzyme activities (32). Round phagocytes are large

cells (15–20 μm in diameter) with one or more phagosomes containing the ingested material as well as hydrolytic enzymes, lipids, and lipofuscins (32). In the colonial ascidian, *Botryllus schlosseri*, the presence of a static and a mobile population of phagocytes was described: the former adhere to the basal lamina of the peribranchial epithelium and form the ventral islands, on both sides of the endostylar sinus (62).

Cytotoxic cells are granular cells, 10–15 μm in diameter; their cytoplasm is filled with large granules containing the inactive form of PO (34). They frequently constitute the most abundant circulating hemocyte type (32). In most of the studied species, cytotoxic cells assume a typical berry-like morphology after aldehyde fixation and are called morula cells (MCs).

As regards pelagic tunicates, Cima et al. (21) reported the characterization of circulating hemocytes of *Thalia democratica* oozoids: they include phagocytes that contain hydrolytic enzymes in their cytoplasm and can migrate into the tunic. Larvaceans have no hemocytes (21).

Hemocytes containing histamine and heparin inside their granules were observed in both ascidians and Thaliaceans: the molecules can either stabilize the granular content or, when released, modulate the inflammatory reaction by inducing tunic vessel-contraction and inhibition of phagocytosis (21, 63).

Digestive System

The oral and the atrial (cloacal) siphons are preferential ways of entrance of microorganisms. Here, a population of phagocytes is exposed to seawater, adhering to the internal tunic. Such sentinel or guard cells can recognize and ingest foreign particles or cells, thus preventing their entrance in the pharynx or in the atrium (64); they were found also in Thaliaceans (21).

In the solitary ascidian *Ciona intestinalis*, both the endostyle and the gastric epithelium constitutively transcribe genes involved in the inflammatory response triggered by the injection of LPS in the body wall (11, 65, 66), suggesting the importance of the alimentary tract in the recognition and the clearance of non-self material. This assumption is corroborated by the reported transcription of genes for Toll-like receptors (TLRs), mannose-binding lectins (MBLs), and MBL-associated serine proteases (MASPs) in both the stomach and the intestine, in addition to hemocytes, in accordance with the important immunosurveillance role of the alimentary tract (48, 58). In addition, variable region-containing chitin-binding proteins (VCBPs), secreted in the gut lumen and recognizing the surface of Gram (+) and Gram (–) bacteria (see below), probably exert a pivotal function in the maintenance of a stable commensal gut microbial flora. This is consistent with the hypothesis of a role of the immune system in both protecting host tissues from pathogenic attack and supporting the growth of the mutualistic microbiota (67). In *B. schlosseri*, gut epithelial cells are involved in the clearance of neighboring apoptotic cells during the generation change (68).

HUMORAL DEFENSIVE REPERTOIRE

Phenoloxidase

The presence of PO activity in ascidian hemolymph has been widely reported in both solitary and colonial species [references

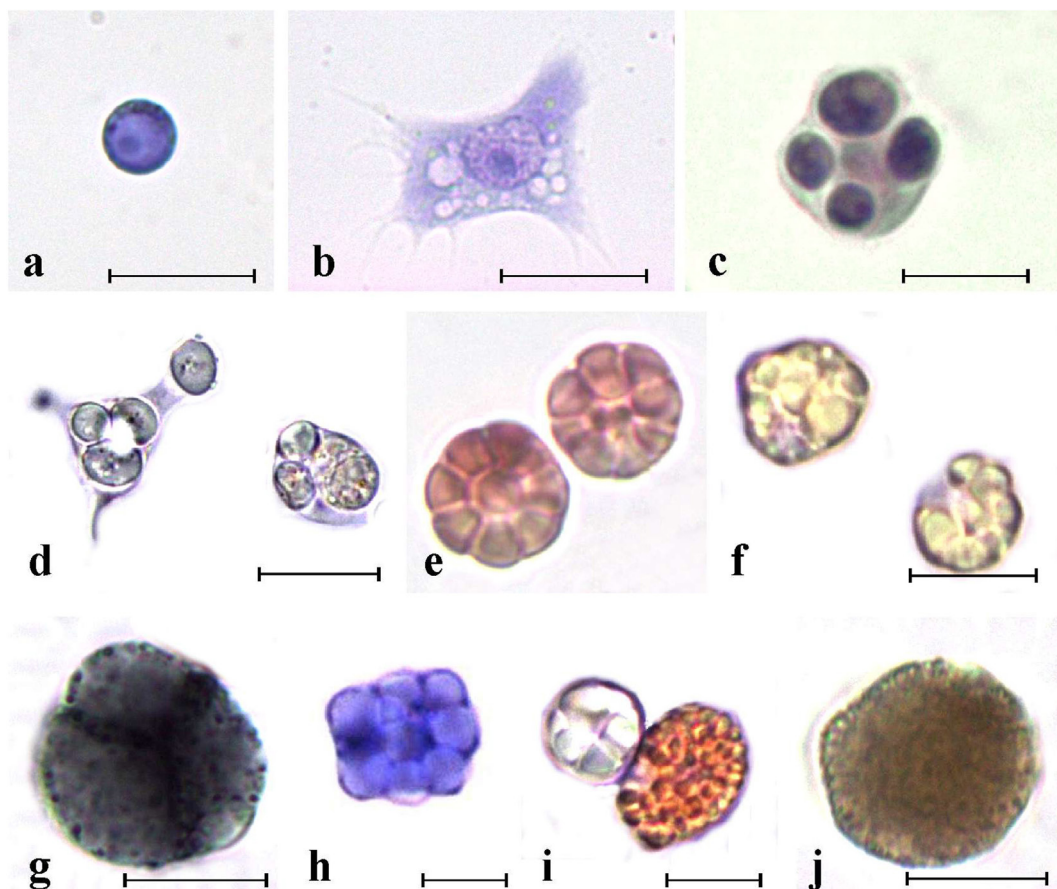


FIGURE 2 | Main ascidian hemocytes. **(A)** Undifferentiated cells. **(B–F)** Immunocytes. **(B,C)** *Botryllus schlosseri* spreading and round phagocytes, respectively; **(D)** *Polyandrocarpa misakiensis* spreading and round phagocyte with ingested yeast cells; **(E)** *B. schlosseri* morula cells (MCs); **(F)** *P. misakiensis* MCs; **(G–J)** storage cells. **(G)** *B. schlosseri* blue pigment cells; **(H)** *P. misakiensis* trophocyte; **(I)** *P. misakiensis* pigment cell and trophocyte; **(J)** *B. schlosseri* nephrocyte. **(A–C,E,H)** aldehyde-fixed cells stained with hematoxylin-eosin; **(D,F,G,I,J)** living cells. Scale bar: 10 µm.

therein (34)]. PO is assumed located as inactive proenzyme (probably, proPO), inside the granules of PO-containing hemocytes and activated by serine proteases once released outside the cells (35, 69). PO-containing hemocytes of *C. intestinalis* store also serine proteases that, once released, are activated by LPS and laminarin shortly before the activation of PO (70, 71). A soluble serine protease is present also in *B. schlosseri* hemolymph (72). This support the idea of an activation of PO mediated by serine proteases, analogous to what is reported in arthropods (73).

Phenoloxidase is involved in cytotoxic responses of ascidians. In colonial botryllid ascidians, the enzyme contributes to the formation of the necrotic spots along the border of contacting, genetically incompatible colonies (34). According to the analysis of nucleotide and predicted amino-acid sequences, ascidian PO shows high similarity with arthropod hemocyanins (74, 75).

Phenoloxidase substrates are likely represented by tunic-chromes, or other phenol-containing peptides, contained inside the hemocyte (mainly MC) granules (29, 35, 76–79). The enzyme produce quinones, that polymerize to melanin, and reactive oxygen species (ROS), that induce oxidative stress and related toxicity in neighboring cells (35).

Lectins

Ascidian immunocytes can synthesize and release humoral lectins with various molecular features and carbohydrate specificities (36, 80–84). Some of them have a clear role in the recognition of foreign molecules or in the modulation of immune responses (36, 65, 85–88). In most cases, they enhance the phagocytosis of microorganisms acting as opsonins (86, 87, 89, 90). Lectins can also trigger the respiratory burst and act as molecules able to influence the behavior of other immunocytes, as in the case of the *Botryllus* rhamnose-binding lectin (BsRBL) (37), or to activate the complement system (46).

A subset of *B. schlosseri* blood cells, probably phagocytes, express an ortholog of the vertebrate CD94 receptor on NK cells, a type II transmembrane protein with a C-type lectin domain (91). A second ortholog in *C. intestinalis* (CiCD94-1) contain a C-type lectin domain without carbohydrate-binding capability: it probably recognizes peptides instead of carbohydrates and is expressed in the same cell type engaged in the production of PO, also recognized by the anti-CiCD94-1 antibody. The fraction of cells positive to the CiCD94-1-1 antisense riboprobe increases after LPS exposure. The anti-CiCD94-1

TABLE 1 | Ascidian main hemocyte and tunic cell categories.

	Cell types (synonyms)	Role	Reference
Hemocytes			
Undifferentiated cells	Hemoblasts (lymphocyte-like cells)	Considered hemocyte precursor cells	(8, 32)
Immunocytes	Phagocytes (hyaline amebocytes, macrophage-like cells, spreading and round phagocytes)	Phagocytosis; encapsulation; synthesis and release of lectins	(8, 13, 32, 33)
	Cytotoxic cells [phenoloxidase (PO)-containing cells, morula cells (MCs), granular amebocytes]	Cell-mediated cytotoxicity; synthesis and release of: cytokines, complement factors, antimicrobial peptides, and collagen	(8, 13, 32, 34, 35)
Storage cells	Pigments cells	Zoid pigmentation	(8, 13, 32)
	Trophocytes	Storage and transport of nutrients	(32)
	Nephrocytes	Storage of uric acid crystals	(8, 32)
	Vanadocytes	Storage of vanadium	(8, 32)
Tunic cells			
Immunocytes	Phagocytes	Ingestion of foreign material having entered the tunic; guard cells (external to the tunic and exposed to the environment in the siphonal areas) controlling the entrance to the pharynx or the atrium	(8, 17, 22, 23)
	Cytotoxic cells (PO-containing cells, MCs, granulocytes)	Cell-mediated cytotoxicity; synthesis and release of: cytokines, complement factors, antimicrobial peptides; crosslinking of tunicin fibers through the oxidation of tunicins by PO	(8, 13, 24, 25, 28, 29)
Bladder cells		Acid storage	(17, 26, 27)
Net cells		Allow the shrinkage of the tunic in wound areas	(17)
Pigment cells		Tunic pigmentation	(17)

TABLE 2 | Ascidian main pattern-recognition receptors.

Name	Location	Role	Reference
Lectins			
Galectins	Granulocytes, phenoloxidase (PO)-containing hemocytes	Mediators of inflammation	(36)
Rhamnose-binding lectins	Phagocytes	Phagocyte activation, opsonins, hemocyte recruitment, PO-containing cell degranulation	(37)
Sialic acid receptor (unknown)	Phagocyte surface	Recognition of sialic acid (do not eat me signal) on the surface of healthy cells	(38)
CD91	Surface of PO-containing cells	Indirect activation of phagocytes	(39)
VCBPs			
VCBP-A, -B, -C	Hemocytes, epithelial cells of stomach and intestine	Opsonins, control of the gut microbiota	(40, 41)
Complement factors			
Mannose-binding lectins	PO-containing hemocytes	Activation of the lectin pathway of complement activation	(42–49)
Ficolins	PO-containing hemocytes	Activation of the lectin pathway of complement activation	(49–52)
C1q	Presumably PO-containing hemocytes	Complement activation by binding to pentraxins	(53–55)
Integrins	Presumably phagocytes	Complement receptor(s)	(32, 46, 56, 57)
Toll-like receptors	Phagocyte surface and endosomes	Recognition of non-self	(48, 58)
CD36	Phagocyte surface	Recognition of oxidized lipids on apoptotic cells	(59)
Phosphatidylserine receptor (unknown)	Presumably phagocyte surface	Recognition of apoptotic cells	(38, 59)

antibody inhibits phagocytosis, suggesting that the interaction of CiCD94-1 with its ligand(s) can indirectly stimulate phagocytes (39, 92), probably through the release of cytokines (see below).

Immunoglobulin (Ig) Domain-Containing Proteins

Despite the lack of orthologs of genes for major histocompatibility complex proteins, T-cell receptors, and Igs, transcripts for

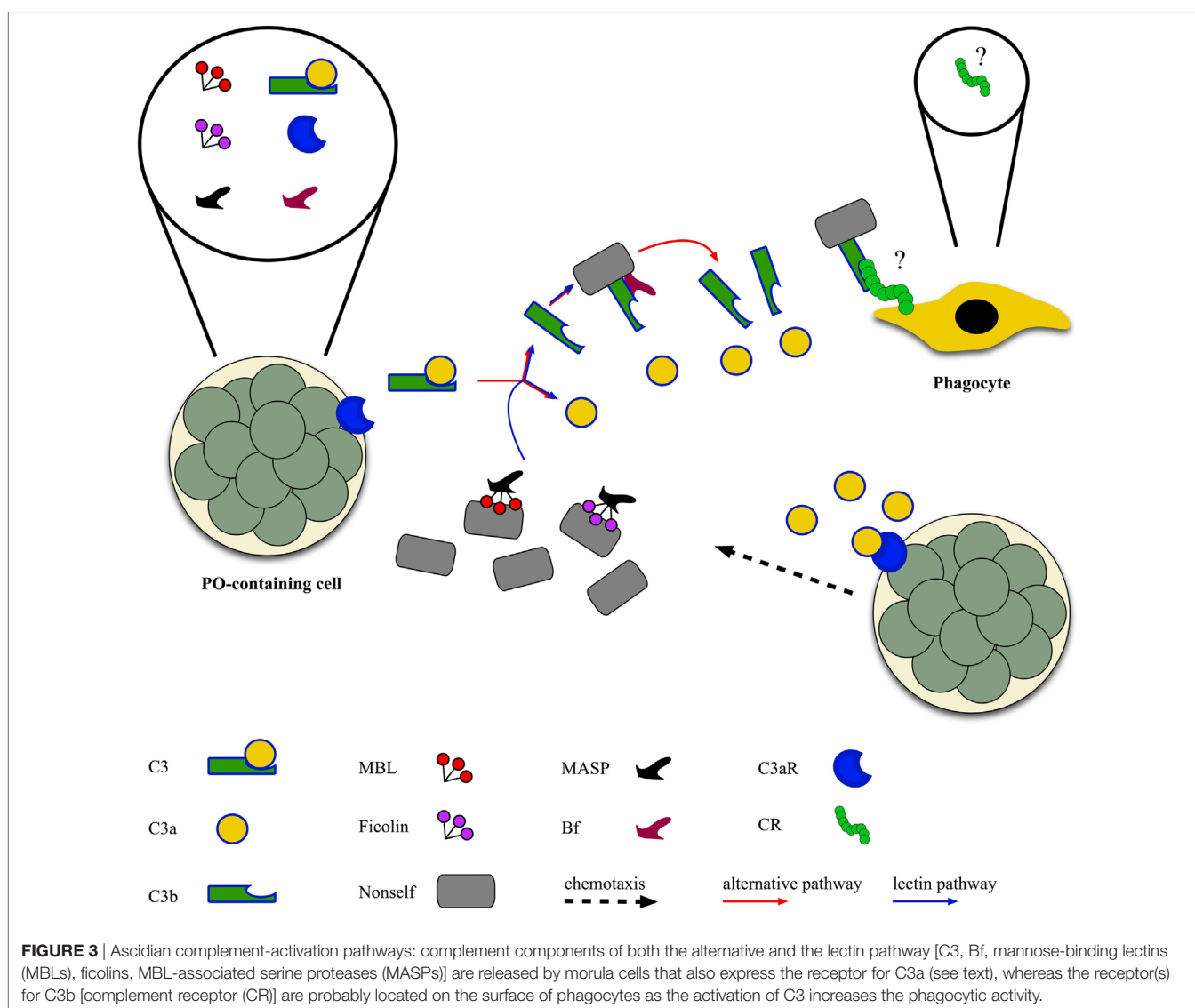
putative molecules with Ig domains were identified in tunicates (93, 94). Three novel genes for VCBPs, containing two N-terminal, variable-type Ig domains, were described in *C. intestinalis*. VCBP-A, -B, and -C are synthesized by epithelial zymogenic cells of the stomach and the intestine, as well as by a fraction of circulating hemocytes (40, 95). VCBPs can bind Gram (+) and Gram (–) bacteria with the variable-type Ig domains and significantly increase microbe phagocytosis by hemocytes, acting as opsonins (40), whereas the chitin-binding domain interacts with the chitin-rich mucus along the intestinal wall, thus influencing the settlement of bacterial communities and the colonization of the intestinal lumen by the microbiota. Indeed, VCBP-C can enhance the *in vitro* production of biofilms by bacteria previously identified in the gut of *Ciona* (41).

Complement System

Both the alternative and lectin complement-activation pathways are present in ascidians [Figure 3; (46, 96)]. Genes for C3 were

identified in all the ascidian species investigated so far (97–100). They are active in the adult (98), and their transcription rate increases after LPS injection in the tunic; similar behavior is reported for the C3-a fragment deriving from the cleavage of C3 in the presence of non-self (101). C3-a can recruit hemocytes to the inflammation site (102) *via* its binding to a G protein-coupled receptor, constitutively expressed in PO-containing hemocytes (103). C3b, the main C3 fragment, can adhere to the microbial surfaces and exert an opsonic role enhancing the recognition and ingestion of bacteria by phagocytes (89, 97, 98, 100, 104).

The transcription of C3 genes occurs in hemocytes, mainly PO-containing hemocytes (97, 100, 101). In *Styela plicata*, hemocytes secrete a protein recognized by anti-C3 antibodies, the concentration of which increases in the culture supernatant after the exposure to non-self molecules (105). In *Pyura stolonifera*, the incubation of hemolymph with LPS induces the release of a chemotactic protein recognized by anti-human C3 antibody (99). In *Halocynthia roretzi*, also cells of the stomach wall transcribe



C3 (97), whereas, in *Ciona*, even ciliated cells bordering the branchial stigmata contain C3 mRNA (101).

Transcripts for Bf, a component of the alternative activation pathway, were identified in various ascidian species (100, 106, 107). Genes for MBLs, C-type lectins members of the collectin family and involved in the lectin pathway of complement activation, are present in the *C. intestinalis* genome (42, 44, 46–48) and over-transcribed during inflammatory reactions (42). Transcripts for MBLs were identified also in other ascidian species (43, 45, 49). In *S. plicata* (108, 109), an increase in the secretion of collectins and in the fraction of hemocytes immunopositive to anti-collectin antibody is observable during inflammatory responses (110). Transcripts for ficolins, also components of the lectin pathway, are present in *H. roretzi* (50), *Botrylloides leachii* (51), and *B. schlosseri* (49, 52). The transcription of *H. roretzi* ficolin 3 gene is significantly impaired in organisms with the soft tunic disease (7). A C-type lectin, interacting with MASP, is involved in the recognition of microbial surfaces and the activation of C3 in *H. roretzi* (111). Transcripts for MASPs were widely described in ascidians (43–46, 48, 49, 55, 96, 104, 112, 113).

C1q-like transcripts were found in *C. intestinalis* (53, 54) and *B. schlosseri* (55). In vertebrates, C1q, a component of the classical activation pathway, can bind pentraxins (mainly C-reactive protein). These molecules were identified in *Ciona* (53) and *Didemnum candidum* (83), suggesting the interaction with pentraxins as the original role of C1q in invertebrate chordates (53). In *B. schlosseri*, the transcription of genes for C1q, MASPs, Bf, and ficolins is upregulated during the allorejection reaction (55); in addition, genes for C3, Bf, ficolin, MASPs, and a putative CR1 are over-transcribed during the recurrent generation changes (113).

As regards complement regulators, in *B. schlosseri*, cDNAs for a putative complement-control protein (CCP), featuring CCP domains, were isolated (114). Genes for α 2-macroglobulin, able to inhibit MASPs, and for various putative molecules with the CCP domain(s), were reported in *C. intestinalis* (44).

C6/C9-like transcripts for proteins containing the membrane-attack complex/perforin domain were described in *C. intestinalis* (44, 46, 47); whether or not a cytolytic pathway is present ascidians, is still a matter of debate.

In *Ciona*, integrin α and β subunits, part of a complement receptor (CR) and showing homology with mammalian CR3 or CR4, are expressed on the surface of hemocytes (46, 56, 57).

Chemical Defense

Ascidians are the source of a great variety of bioactive molecules of potential interest in the sanitary field; some of them have also entered human clinical trials (115). Many compounds act as antiviral or repellents against foulants, predators, and competitors (116–120). Acid substances and metals stored in vacuoles within tunic cells can contribute to additional protection (26, 121, 122). The tunic may host prokaryotes that produce many of the above-reported products (115, 121).

Ascidians produce also molecules with antimicrobial activity (123–126). Most of them are peptides; in many cases, they are synthesized by hemocytes, mostly PO-containing cells. In *H. roretzi*, halocyanines A and B are synthesized by MCs (127), and their cytotoxic activity is likely related to the presence of diphenol

rings that render them substrates for PO. *S. clava* MCs produce clavanins A–D, histidine-rich, α -helix peptides, and clavaspirin (128, 129). In the same species, five styelins, cationic antimicrobial peptides, were identified and isolated from hemocyte lysates (130, 131). In *C. intestinalis*, PO-containing hemocytes synthesize two families of α -helix antimicrobial peptides and the injection of non-self material in the body wall enhances the transcription of the corresponding genes (24, 132–134). Anticancer derivatives were also described (135, 136), and ascidian tunicchromes can exert a cytotoxic activity (28). A gene homologous to mammalian EB1, a protein with tumor suppressing effect, was described in *B. schlosseri* (137).

Cytokines and Cross Talk between Immunocytes

Despite the common opinion that invertebrate cytokines share no homologies with their vertebrate counterparts (138, 139), putative genes for IL1 and TNF receptors were identified in the *Ciona* genome (44, 61). A gene for a TNF α homolog, the transcription of which increases in *Ciona* hemocytes after LPS injection in the body wall, was also cloned (11, 140): it probably exerts a role in recruiting hemocytes to the inflamed area (141). Genes for a putative IL17 receptor and three IL17 homologs were identified in *Ciona* (3, 60, 61): their expression (in hemocytes) is also upregulated after LPS injection in the tunic (142).

In *B. schlosseri*, MCs are the main source of molecules recognized by antibodies raised against mammalian pro-inflammatory cytokines, secreted upon the recognition of foreign molecules (143). They induce phagocytes to synthesize and release BsRBL, with opsonic activity [Figure 4; (144)]. Anti-cytokine antibodies prevent the increase in phagocytosis observed when hemocytes are incubated in the supernatants of hemocytes cultures previously challenged with yeast (*Saccharomyces cerevisiae*) cells (145). In botryllid ascidians, during the allorejection reaction, MCs produce and release molecules immunopositive to anti-IL1 α and anti-TNF α antibodies (25, 100, 146). They are involved in the recruitment of these cells to the ampullae of the contact region (see below), as demonstrated by the inhibition of the MC chemotaxis, induced by cell-free hemolymph from incompatible colonies, in the presence of the above-reported antibodies (146, 147). In *B. schlosseri*, the gene for an IL-17 ortholog is over-transcribed during the generation change: it probably modulates the cellular events occurring during this phase of the colonial life cycle and mediates the cross talk between MCs and phagocytes (113). A cooperation between MCs and phagocytes was reported also in *C. intestinalis* (70).

VARIETY OF CELL-MEDIATED IMMUNE RESPONSES IN ASCIDIANS

Hemocyte Aggregation

Tunicate lack a coagulation system and hemocytes migrate and aggregate to plug the injured sites and prevent hemolymph leakage. Hemocyte aggregation was particularly studied in the solitary ascidian *H. roretzi* (148) where a membrane glycoprotein, active in both phagocytosis and hemocyte aggregation was identified (149). It contains two immunoreceptor tyrosine-based activation

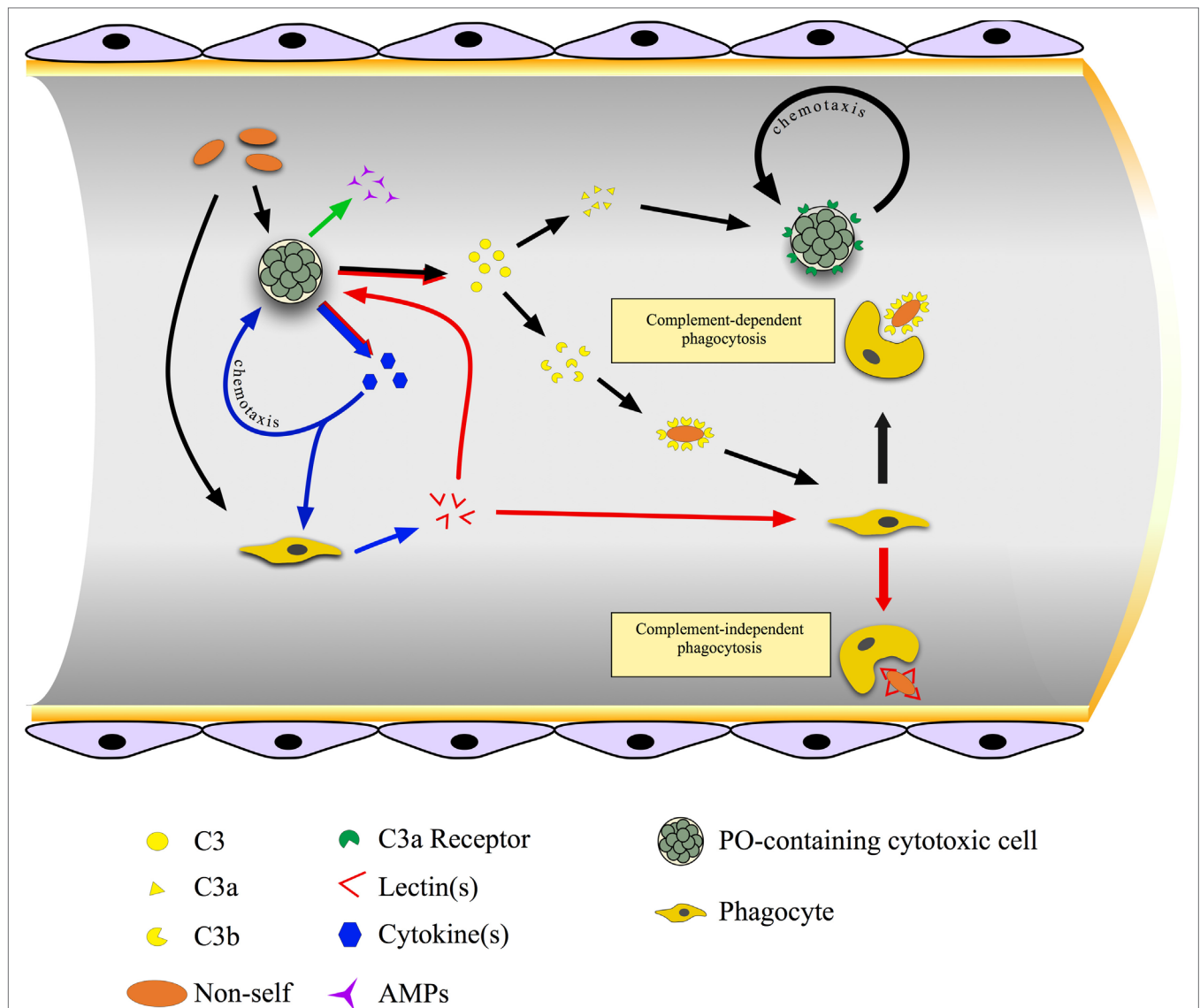


FIGURE 4 | Cross talk between immunocytes in the colonial ascidian *Botryllus schlosseri*. Cytotoxic, phenoloxidase (PO)-containing cells are the first cells to sense non-self material and, as a consequence, they synthesize and release cytokines, antimicrobial peptides, and complement C3. Cytokines act on both morula cells (MCs) themselves, inducing their chemotaxis, and on phagocytes triggering the synthesis and the release of lectins, mainly rhamnose-binding lectin (RBL), that bind carbohydrates on the microbial surfaces and exert a complement-independent opsonic role. C3 is cleaved to C3a, which cooperates in recruiting MCs, and C3b, which interacts with the microbial surface and acts as opsonin.

motifs (ITAMs) and associates with phosphorylated and unphosphorylated proteins, strongly suggesting its involvement in triggering signal transduction pathways (150). Further analyses demonstrated that, during hemocyte aggregation, it induces gene transcription through the activation of phosphatidylinositol-3 kinase (PI3K) and cytosolic calcium rise (151).

Endocytosis

In ascidians, the ingestion of foreign materials occurs through either macropinocytosis or phagocytosis. In both cases, integrins and molecules containing the Arg-Gly-Asp (RGD) motif (e.g., fibronectin or fibrinogen) are involved (25). Pattern-recognition receptors allow the direct interaction of circulating

professional phagocytes with potentially pathogenic foreign material. As an alternative, they recognize opsonins covering the microbial surfaces and enhancing phagocytosis. Opsonin-mediated phagocytosis can be either complement-dependent or complement-independent (Figure 4). A transient rise in cytosolic Ca^{2+} concentration is required for the ingestion, whereas a sustained increase lowers the extent of phagocytosis (25). The interaction of phagocytes with non-self particles triggers a respiratory burst, with the activation of both a membrane oxidase and an inducible nitric oxide (NO) synthase that leads to the production of ROS and reactive nitrogen species with microbicidal activity (152).

As for receptors involved in endocytosis, in *C. intestinalis*, two TLR genes were identified (60) and fully characterized: the

corresponding proteins have cytoplasmic TIR, transmembrane, and extracellular LRR domains and are located in both the plasma membrane and the endosome membrane of phagocytes (58). In addition, *Ciona* also possesses a rich repertoire of transcripts of genes involved in signal transduction, including those for proteins with immunoreceptor tyrosine-based inhibition motifs and ITAMs, MyD88, IL1 receptor-associated kinase, TNF receptor-associated factor, nuclear factor κ B (NF- κ B), and inhibitor of κ B (44, 53, 60). In the colonial *B. schlosseri*, TLRs are present on the surface and the interior of phagocytes (25). Here, the signal transduction pathways triggered by non-self recognition, include the activation of trimeric G-proteins, protein kinase A, protein kinase C, PI3K, mitogen-activated protein kinases (MAPKs), and NF- κ B (25, 153, 154).

Phagocytosis of apoptotic cells is a common event in botryllid ascidians, where cyclical generation of new zooids by budding occurs, and old zooids are periodically resorbed (155). Generation change or take-over implies massive apoptosis in the tissues of old zooids and the clearance of dying cells by professional and occasional phagocytes (68, 156–158). Phagocytes recognize phosphatidylserine and the lack of sialic acid on the surface of effete cells and corpses (38, 59) and avidly ingest them: because of the sudden increase of oxygen consumption and the related oxidative stress, they undergo phagocytosis-induced apoptosis and are, in turn, ingested by other phagocytes (159). Clearance of dying cells requires also the presence of CD36, a scavenger receptor able to recognize oxidized lipids, on the phagocyte surface (59); a putative CD36 ortholog was identified in the *Ciona* genome (44). In *B. schlosseri*, the clearance of apoptotic cells by phagocytes is necessary for the completion of the take-over and the progression of bud development (160). The opposite is also true: buds are required for the clearance of cell corpses as they recycle the nutrients deriving from their digestion by phagocytes (161, 162).

Encapsulation

Foreign material too large to be ingested by phagocytosis is usually encapsulated by circulating hemocytes. The formation of multi-layered capsules was observed around parasitic crustaceans, and both phagocytes and cytotoxic MCs can be involved in capsule formation (33). In *C. intestinalis*, intratunical injection of mammalian erythrocytes or non-self molecules results in massive recruitment of hemocytes to the inoculum site and capsule formation (11).

In *B. scalaris*, unlike other botryllid ascidians (see below), encapsulation plays a pivotal role in allorecognition. Here, the circulatory systems fuse during allerejection and blood exchange begins. Phagocytes crowd inside the fused vessels and stimulate the aggregation of hemocytes into large clusters, finally encapsulated by other phagocytes, so to plug the lumen of the vessels and interrupt the hemolymph flow in a few minutes (163).

Cytotoxicity

A Ca^{2+} -dependent cytotoxic activity against mammalian erythrocytes or tumor cells, inhibited by sphingomyelin, was described in *C. intestinalis* and *S. plicata* (164–166). In *C. intestinalis*, cytotoxicity against mammalian cells requires the activity of the

enzyme phospholipase A2, modulated by lectins with specificity for galactosides (167). A cytotoxic reaction, called *contact reaction*, occurs in allogeneic or xenogeneic combinations of hemocytes from various solitary ascidians (168). In *B. schlosseri*, cytotoxicity can be observed *in vitro* by exposing hemocytes to non-self molecules or cell-free hemolymph of incompatible colonies (79). In all the above cases, cytotoxicity is consequent to the release of active PO in the medium upon degranulation of PO-containing hemocytes and the oxidation of polyphenol substrata, leading to the production of toxic quinones and ROS (34). In *B. schlosseri*, NO is also involved in the induction of cell death (146). The production of NO by hemocytes, after their exposure to either LPS or zymosan, was reported also in *S. plicata* and *Phallusia nigra* (169, 170).

Inflammation

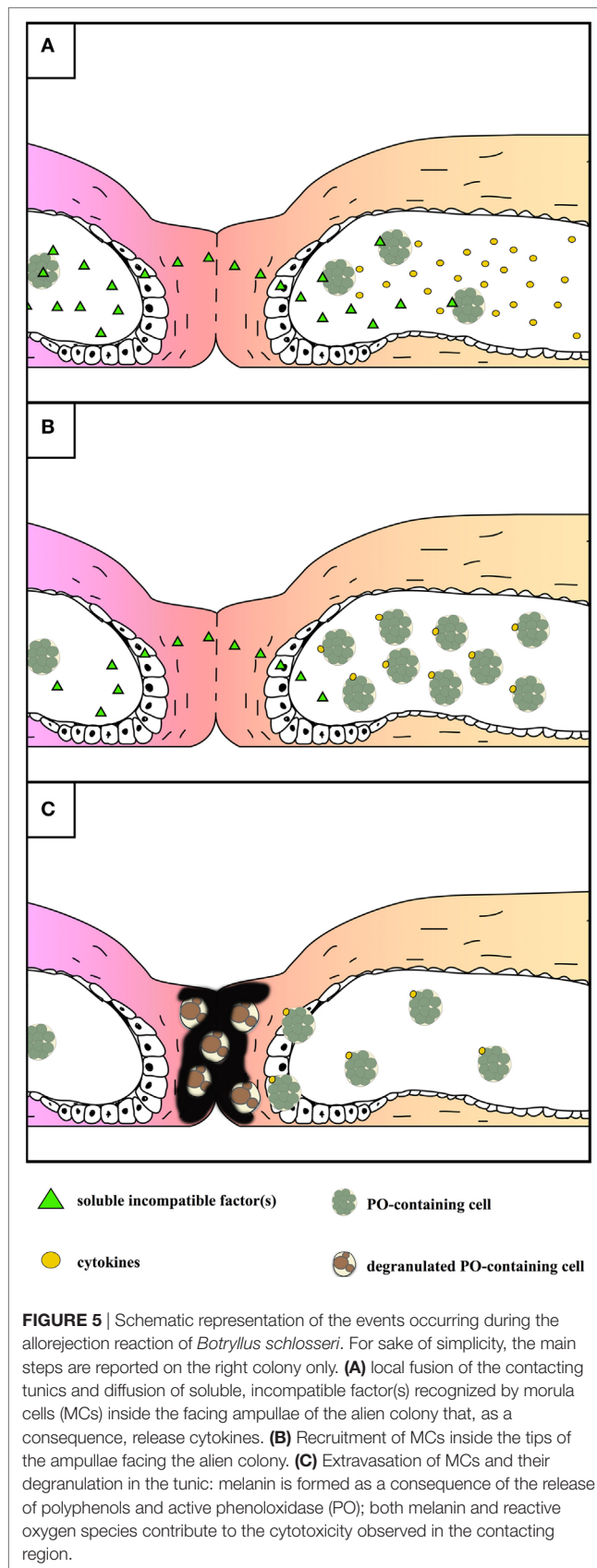
Inflammation is characterized by the recruitment of circulating hemocytes, extravasation, cell degranulation, induction of cytotoxicity, and phagocytosis (or encapsulation) of the foreign material. Inflammation-related cytotoxicity requires the recruitment of PO-containing hemocytes and the release of active PO in the infected area (142, 171, 172). It was particularly studied in *C. intestinalis*, after the injection of foreign material in the tunic (11). Circulating hemocytes of treated animals increase the transcription of genes involved in the recognition of non-self and tissue repair (11, 173–175).

Inflammation in Tissue Transplantation

Tissue transplantation represents a cause of inflammation. In solitary species, higher recruitments of hemocytes occur in the case of allografts with respect to autografts, leading to allograft rejection. The latter is more rapid in primed animals, having previously received (and rejected) a similar graft (176–178). Graft rejection relies on PO-containing hemocytes reaching the inflamed area and the induction of cytotoxicity (179). In *C. intestinalis*, the products of a polymorphic gene, structurally similar to a vertebrate CR and containing CCP domains, were proposed as individuality markers. They are synthesized by hemocytes, with various splice variants and high interindividual variability (180, 181).

Inflammation in Allorecognition

In botryllid ascidians, inflammatory events are the consequence of allorecognition between incompatible colonies, probably to prevent the risk of somatic/germ cell parasitism in genetically unrelated colonies (182, 183). In *Botryllus primigenus* and *B. schlosseri*, a highly polymorphic fusibility/histocompatibility (Fu/HC) gene with codominant alleles controls the outcome of the colony contact (184, 185). When colonies share no alleles at the Fu/HC locus, partial fusion of the facing tunics occurs as well as the leakage of soluble histocompatibility factors, recognized by MCs (186). Activated MCs release chemotactic cytokines able to recruit other MCs in the peripheral blind endings of the tunic vasculature (ampullae) of the contact region (147), from which they enter the tunic and degranulate, thus releasing the enzyme PO and its polyphenol substrata. A series of melanic cytotoxic foci, called points of rejection, appear along the contact border as a result of cytotoxicity [Figure 5; (34, 79, 146, 187)]. Rejecting colonies of *B. schlosseri* increase the transcription rate of various



immune-relevant genes (52, 55, 188). A change in the growth direction of contacting colonies occurs after the allorecognition reaction (189). MCs are involved in the allorecognition reaction also in *Botrylloides simodensis*, *Botrylloides fuscus*, *Botrylloides violaceus*, *B. leachi* (35), and *Didemnum perlucidum* (190). The unusual growth of facing ampullae during allorecognition was reported in *B. leachi* (35, 191).

An intense inflammatory reaction is observed when incompatible colonies of ovoviviparous botryllid ascidians are brought into contact at their cut surfaces (192, 193), whereas fusion of tunics and hemolymph vessels always occurs in the case of viviparous species. This suggests that, in the latter case, hemocytes have lost their ability of allorecognition, probably to avoid immune attacks toward the brooded embryos that share only one Fu/HC allele with the mother colony (194). In support of the above hypothesis, the PO activity of the hemolysate of viviparous species is lower than that of ovoviviparous ones (195, 196).

When *Botryllus* colonies share at least one allele at the Fu/HC locus, contacting colonies can fuse and form a single chimeric colony (197). However, in the case of a single shared allele, the resorption of one of the chimeric partner occurs within 30 days from the temporary fusion (198). Even in this case, MCs are directly involved as they infiltrate the tissues of the loser colony, together with phagocytes. The resorption phenomenon can be induced by the injection of enriched populations of MCs in the vasculature of recipient colonies and shares many similarities with the take-over, including apoptosis in zooid tissues, clearance of dying cells by phagocytes, and modulation by IL17 (113).

In *B. schlosseri*, the ampullar epithelium and hemocytes express genes for proteins involved in allorecognition, although uncertainties on the identity of the allorecognition gene still persist (199–204).

ROLE IN DEVELOPMENT?

Many invertebrate molecules have a role in both development and immunity. The best example is the *Drosophila* Toll receptor, required for the establishment of dorsal–ventral polarity early in development and switching to an immune role in adult flies (205). In Tunicates, various genes, involved in adult immune responses, are transcribed also during embryonic, larval, and asexual development, and this opens interesting perspectives on their role in development.

In *B. villosa*, the analysis of the transcriptome revealed the expression of immune-related genes in both the larval and juvenile development (43). In addition, MASPs are probably involved in the activation of metamorphosis (206).

In the larva of the ascidian *Ascidia callosa*, tunichrome, the putative substrate of PO, is required for tunic morphogenesis (207).

In *C. intestinalis*, a C3-like gene is transcribed during early development: it codifies a protein that, probably, does not exert a typical C3 role (208). Orthologous genes of C6 and C1 are also active during the embryonic stage (54). In addition, the gene for *CiCd94-1* is transcribed in larval papillae, in cells of the larval nervous system, and in the coronet cells, the probable precursors

of neural crest cells, with a role in modeling the nervous system during development (39). Furthermore, swimming larvae transcribe a gene for a CiTNF α -like protein (141), and PO gene expression is modulated in early and larval development (209). In the same organism, very low transcription levels of VCBP genes can be detected before the tailbud stage. From the larval stage onward, their mRNAs are located in gut primordia, with different distributions in defined territories, suggesting a role of VCBPs in the functional compartmentalization of the developing intestine (95, 210). VCBP mRNAs are translated after metamorphosis, with different timing of appearance and distribution (41). The transcription of VCBP genes in juveniles is differentially modulated by Gram (+) and Gram (–) bacteria, fitting the idea of their role in mediating the onset of the microbial gut colonization (95, 210).

An increase in the transcription of several immune-related genes occurs also during the whole body regeneration of *B. leachii* (51). In addition, signaling pathways, such as those involving MAPK and the NF- κ B/Rel family members, are required in the formation of the larval notochord (211) and in the budding process of botryllid ascidians (212).

FUTURE PERSPECTIVES

Tunicates, and ascidians in particular, are simple chordates that represent valuable models for the study of the innate immune responses and the evolutionary events that occurred in the course of invertebrate–vertebrate transition, leading to the appearance of lymphocytes and receptor diversification through somatic recombination. The progressive availability of new sequenced

transcriptomes and genomes from tunicates will enable researchers to dissect the genetic and molecular processes associated with immune responses, clarify the regulatory pathways and the diversity of pattern-recognition receptors involved in immune responses, and compare them with what known in vertebrates. Ascidians offer also the possibility to study some particular aspects of the immune responses, such as the evolutionary importance of the polymorphism found in Fu/HC and other immune genes and its relationships with pathogen threats, the molecular basis of the priming phenomenon, the evolution of the complement system, and the role of lectins as immunomodulatory molecules. In addition, the possibility of synthesizing the gene products once the gene sequences are known, can render available a quantity of bioactive molecules, involved in chemical defense, testable as antimicrobial, antiviral, or anticancer compounds. Last, but not least, research on hemocytes will contribute to disentangle the unresolved aspects of hemocyte ontogeny and differentiation pathways and better elucidate their role in tunicate biology.

AUTHOR CONTRIBUTIONS

LB set up the work plan. LB and NF equally contributed to the text of the review.

FUNDING

This work was supported by the University of Padova (DOR 2016).

REFERENCES

- Swalla BJ, Smith AB. Deciphering deuterostome phylogeny: molecular, morphological and paleontological perspectives. *Philos Trans R Soc* (2008) 363B:1557–68. doi:10.1098/rstb.2007.2246
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* (2006) 439:965–8. doi:10.1038/nature04336
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso AW, et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* (2002) 298:2157–67. doi:10.1126/science.1080049
- Voskoboinik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, Koh W, et al. The genome sequence of the colonial chordate, *Botryllus schlosseri*. *Elife* (2013) 2:e00569. doi:10.7554/eLife.00569
- Whitman WB, Coleman DC, Wiebe WJ. Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* (1998) 95:6578–83. doi:10.1073/pnas.95.12.6578
- Suttle CA. Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol* (2007) 5:801–12. doi:10.1038/nrmicro1750
- Cha IS, Segovia del Castillo C, Nho SW, Hikima J, Aoki T, Jung TS. Innate immune response in the hemolymph of an ascidian, *Halocynthia roretzi*, showing soft tunic syndrome, using label-free quantitative proteomics. *Dev Comp Immunol* (2011) 35:809–16. doi:10.1016/j.dci.2011.01.011
- Burighel P, Cloney RA. Urochordata: Ascidiacea. In: Harrison FW, Ruppert EE, editors. *Microscopic Anatomy of Invertebrates*. (Vol. 15), New York: Wiley-Liss Inc. (1997). p. 221–347.
- Xu CX, Jin H, Chung YS, Shin JY, Woo MA, Lee KH, et al. Chondroitin sulfate extracted from the *Styela clava* tunic suppresses TNF- α -induced expression of inflammatory factors, VCAM-1 and iNOS by blocking Akt/NF- κ B signal in JB6 cells. *Cancer Lett* (2008) 264:93–100. doi:10.1016/j.canlet.2008.01.022
- Christomelba DC, Ananthan G. Chitosan derived from the tunic of ascidian *Phallusia nigra* (Savigny, 1816) showing antibacterial activities and its characterization. *Int J Med Med Sci* (2016) 8:91–5. doi:10.5897/IJMMMS2014.1087
- Parrinello N, Cammarata M, Parrinello D, Vizzini A. Inflammatory response of the ascidian *Ciona intestinalis*. In: Ballarin L, Cammarata M, editors. *Lessons in Immunity: from Single-Cell Organisms to Mammals*. London: Elsevier (2016). p. 177–92.
- Di Bella MA, Carbone MC, D'Amato M, Alessandro R, De Leo G. The identification and localization of two intermediate filament proteins in the tunic of *Styela plicata* (Tunicata, Styelidae). *Tissue Cell* (2009) 41:381–9. doi:10.1016/j.tice.2009.04.001
- Goodbody I. The physiology of ascidians. *Adv Mar Biol* (1974) 12:1–149. doi:10.1016/S0065-2881(08)60457-5
- Hirose E, Kimura S, Itoh T, Nishikawa J. Tunic morphology and cellulosic components of pyrosomas, doliolids, and salps (Thaliacea, Urochordata). *Biol Bull* (1999) 196:113–20. doi:10.2307/1543173
- Di Bella MA, Cassarà G, Russo D, De Leo G. Cellular components and tunic architecture of the solitary ascidian *Styela canopus* (Stolidobranchiata, Styelidae). *Tissue Cell* (1998) 30:352–9. doi:10.1016/S0040-8166(98)80048-7
- Di Bella MA, Carbone MC, De Leo G. Aspects of cell production in mantle tissue of *Ciona intestinalis* L. (Tunicata, Ascidiacea). *Micron* (2005) 36:477–81. doi:10.1016/j.micron.2005.01.007
- Hirose E. Ascidian tunic cells: morphology and functional diversity of free cells outside the epidermis. *Invertebr Biol* (2009) 128:83–96. doi:10.1111/j.1744-7410.2008.00153.x
- Endean R. The test of the ascidian, *Phallusia mammillata*. *Q J Micr Sci* (1961) 102:107–17.
- Hirose E, Saito Y, Watanabe H. Tunic cell morphology and classification in botryllid ascidians. *Zoolog Sci* (1991) 8:951–8.
- Hirose E. Pigmentation and acid storage in the tunic: protective functions of the tunic cells in the tropical ascidian *Phallusia nigra*. *Invertebr Biol* (1999) 118:414–22. doi:10.2307/3227010
- Cima F, Caicci F, Sordino P. The haemocytes of the salp *Thalia democratica* (Tunicata, Thaliacea): an ultrastructural and histochemical study in the oozoid. *Acta Zool (Stockh)* (2014) 95:375–91. doi:10.1111/azo.12034

22. Hirose E, Ishii T, Saito Y, Taneda Y. Phagocytic activity of tunic cells in the colonial ascidian *Aplidium yamazii* (Polyclinidae Aplousobranchia). *Zool Sci* (1994) 11:203–8.
23. Ishii T, Hirose E, Taneda Y. Tunic phagocytes are involved in allorejection reaction in the colonial tunicate *Aplidium yamazii* (Polyclinidae, Ascidiacea). *Biol Bull* (2008) 214:145–52. doi:10.2307/25066671
24. Di Bella MA, Fedders H, De Leo G, Leippe M. Localization of antimicrobial peptides in the tunic of *Ciona intestinalis* (Ascidiacea, Tunicata) and their involvement in local inflammatory-like reactions. *Results Immunol* (2011) 1:70–5. doi:10.1016/j.rinim.2011.09.001
25. Ballarin L. Immunobiology of compound ascidians, with particular reference to *Botryllus schlosseri*: state of art. *Invert Surviv J* (2008) 5:54–74.
26. Hirose E. Acid containers and cellular networks in the ascidian tunic with special remarks on ascidian phylogeny. *Zool Sci* (2001) 18:723–31. doi:10.2108/zsj.18.723
27. Hirose E, Yamashiro H, Mori Y. Properties of tunic acid in the ascidian *Phallusia nigra* (Asciidiidae, Phlebobranchia). *Zool Sci* (2001) 18:309–14. doi:10.2108/zsj.18.309
28. Cai M, Sugumaran M, Robinson WE. The crosslinking and antimicrobial properties of tunichrome. *Comp Biochem Physiol* (2008) 151B:110–7. doi:10.1016/j.cbpb.2008.06.004
29. Sugumaran M, Robinson WE. Structure, biosynthesis and possible function of tunichromes and related compounds. *Comp Biochem Physiol* (2012) 163B:1–25. doi:10.1016/j.cbpb.2012.05.005
30. Hirose E, Ohshima C, Nishikawa J. Tunic cells in pyrosomes (Thaliacea, Urochordata): cell morphology, distribution, and motility. *Invertebr Biol* (2001) 120:386–93. doi:10.1111/j.1744-7410.2001.tb00047.x
31. Kimura S, Ohshima C, Hirose E, Nishikawa J, Itoh T. Cellulose in the house of the appendicularian *Oikopleura rufescens*. *Protoplasma* (2001) 186:24–33. doi:10.1007/BF01276931
32. Cima F, Franchi N, Ballarin L. Origin and function of tunicate hemocytes. In: Malagoli D, editor. *The Evolution of the Immune System*. London: Elsevier (2016). p. 29–49.
33. Wright RK, Cooper EL. Inflammatory reactions of the protochordata. *Am Zool* (1983) 23:205–11. doi:10.1093/icb/23.1.205
34. Ballarin L. Ascidian cytotoxic cells: state of the art and research perspectives. *Invert Surviv J* (2012) 9:1–6.
35. Franchi N, Ballarin L. Cytotoxic cells of compound ascidians. In: Ballarin L, Cammarata M, editors. *Lessons in Immunity: from Single-Cell Organisms to Mammals*. London: Elsevier (2016). p. 193–203.
36. Vizzini A, Parrinello D, Sanfratello MA, Salerno G, Cammarata M, Parrinello N. Inducible galectins are expressed in the inflamed pharynx of the ascidian *Ciona intestinalis*. *Fish Shellfish Immunol* (2012) 32:101–9. doi:10.1016/j.fsi.2011.10.028
37. Franchi N, Schiavon F, Carletto M, Gasparini F, Bertoloni G, Tosatto SCE, et al. Immune roles of a rhamnose-binding lectin in the colonial ascidian *Botryllus schlosseri*. *Immunobiology* (2011) 216:725–36. doi:10.1016/j.imbio.2010.10.011
38. Cima F, Manni L, Basso G, Fortunato E, Accordi B, Schiavon F, et al. Hovering between death and life: natural apoptosis in the blastogenetic cycle of the colonial ascidian *Botryllus schlosseri*. *Dev Comp Immunol* (2010) 34:272–85. doi:10.1016/j.dci.2009.10.005
39. Zucchetti I, Marino R, Pinto MR, Lambris JD, Du Pasquier L, De Santis R. CiCD94-1, an ascidian multipurpose C-type lectin-like receptor expressed in *Ciona intestinalis* hemocytes and larval neural structures. *J Immunol* (2008) 162:387–91. doi:10.1111/j.1432-0436.2007.00214.x
40. Dishaw LJ, Giacomelli S, Melillo D, Zucchetti I, Haire RN, Natale L, et al. A role for variable region-containing chitin-binding proteins (VCBPs) in host gut-bacteria interactions. *Proc Nat Acad Sci U S A* (2011) 108:16747–52. doi:10.1073/pnas.1109687108
41. Dishaw LJ, Leigh B, Cannon JP, Liberti A, Mueller MG, Skapura DP, et al. Gut immunity in a protochordate involves a secreted immunoglobulin-type mediator binding host chitin and bacteria. *Nat Commun* (2016) 7:10617. doi:10.1038/ncomms10617
42. Bonura A, Vizzini A, Salerno G, Parrinello N, Longo V, Colombo P. Isolation and expression of a novel MBL-like collectin cDNA enhanced by LPS injection in the body wall of the ascidian *Ciona intestinalis*. *Mol Immunol* (2009) 46:2389–94. doi:10.1016/j.molimm.2009.04.035
43. Davidson B, Swalla BJ. A molecular analysis of ascidian metamorphosis reveals activation of an innate immune response. *Development* (2002) 129:4739–51.
44. Azumi K, De Santis R, De Tomaso AW, Rigoutsos I, Yoshizaki F, Pinto MR, et al. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: “waiting for Godot”. *Immunogenetics* (2003) 55:570–81. doi:10.1007/s00251-003-0606-5
45. Vasta GR, Quesenberry MS, Ahmed H, O’Leary N. C-type lectins and galectins mediate innate and adaptive immune functions: their roles in the complement activation pathway. *Dev Comp Immunol* (1999) 23:401–20. doi:10.1016/S0145-305X(99)00020-8
46. Fujita T, Endo Y, Nonaka M. Primitive complement system – recognition and activation. *Mol Immunol* (2004) 41:103–11. doi:10.1016/j.molimm.2004.03.026
47. Wakoh T, Ikeda M, Uchino R, Azumi K, Nonaka M, Kohara Y, et al. Identification of transcripts expressed preferentially in hemocytes of *Ciona intestinalis* that can be used as molecular markers. *DNA Res* (2004) 11:345–52. doi:10.1093/dnares/11.5.345
48. Skjoedt MO, Palarasah Y, Rasmussen K, Vitved L, Salomonsen J, Kliem A, et al. Two mannose-binding lectin homologues and an MBL-associated serine protease are expressed in the gut epithelia of the urochordate species *Ciona intestinalis*. *Dev Comp Immunol* (2010) 34:59–68. doi:10.1016/j.dci.2009.08.004
49. Franchi N, Ballarin L. Morula cells as key hemocytes of the lectin pathway of complement activation in the colonial tunicate *Botryllus schlosseri*. *Fish Shellfish Immunol* (2017) 63:157–64. doi:10.1016/j.fsi.2017.02.003
50. Kenjo A, Takahashi M, Matsushita M, Endo Y, Nakata M, Mizuuchi T, et al. Cloning and characterization of novel ficolins from the solitary ascidian *Halocynthia roretzi*. *J Biol Chem* (2001) 276:19959–65. doi:10.1074/jbc.M011723200
51. Rinkevich Y, Douek J, Haber O, Rinkevich B, Reshef R. Urochordate whole body regeneration inaugurates a diverse innate immune signaling profile. *Dev Biol* (2007) 312:131–46. doi:10.1016/j.ydbio.2007.09.005
52. Oren M, Escande ML, Paz G, Fishelson Z, Rinkevich B. Urochordate histoincompatible interactions activate vertebrate-like coagulation system components. *PLoS One* (2008) 3:e3123. doi:10.1371/journal.pone.0003123
53. Iwanaga S, Lee BL. Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol* (2005) 38:128–50.
54. Azumi K, Sabau SV, Fujie M, Usami T, Koyanagi R, Kawashima T, et al. Gene expression profile during the life cycle of the urochordate *Ciona intestinalis*. *Dev Biol* (2007) 308:572–82. doi:10.1016/j.ydbio.2007.05.022
55. Oren M, Paz G, Douek J, Rosner A, Or Amar K, Rinkevich B. Marine invertebrates cross phyla comparisons reveal highly conserved immune machinery. *Immunobiology* (2013) 218:484–95. doi:10.1016/j.imbio.2012.06.004
56. Miyazawa S, Azumi K, Nonaka M. Cloning and characterization of integrin α subunits from the solitary ascidian *Halocynthia roretzi*. *J Immunol* (2001) 166:1710–5. doi:10.4049/jimmunol.166.3.1710
57. Miyazawa S, Nonaka M. Characterization of novel ascidian β integrins as primitive complement receptor subunits. *Immunogenetics* (2004) 55:836–44. doi:10.1007/s00251-004-0651-8
58. Sasaki N, Ogasawara M, Sekiguchi T, Kusumoto S, Satake H. Toll-like receptors of the ascidian *Ciona intestinalis*: prototypes with hybrid functionalities of vertebrate toll-like receptors. *J Biol Chem* (2009) 284:27336–43. doi:10.1074/jbc.M109.032433
59. Cima F, Basso G, Ballarin L. Apoptosis and phosphatidylserine-mediated recognition during the take-over phase of the colonial life-cycle in the ascidian *Botryllus schlosseri*. *Cell Tissue Res* (2003) 312:369–76. doi:10.1007/s00441-003-0738-9
60. Shida K, Terajima D, Uchino R, Ikawa S, Ikeda M, Asano K, et al. Hemocytes of *Ciona intestinalis* express multiple genes involved in innate immune host defense. *Biochem Biophys Res Commun* (2003) 302:207–18. doi:10.1016/S0006-291X(03)00113-X
61. Terajima D, Yamada S, Uchino R, Ikawa S, Ikeda M, Shida K, et al. Identification and sequence of seventy-nine new transcripts expressed in hemocytes of *Ciona intestinalis*, three of which may be involved in characteristic cell-cell communication. *DNA Res* (2003) 10:203–12. doi:10.1093/dnares/10.5.203
62. Lauzon RJ, Brown C, Kerr L, Tiozzo S. Phagocyte dynamics in a highly regenerative urochordate: insights into development and host defense. *Dev Biol* (2013) 374:357–73. doi:10.1016/j.ydbio.2012.11.006

63. García-García E, Gómez-González NE, Meseguer J, García-Ayala A, Mulero V. Histamine regulates the inflammatory response of the tunicate *Styela plicata*. *Dev Comp Immunol* (2014) 46:382–91. doi:10.1016/j.dci.2014.05.017
64. Cima F, Ballarin L, Gasparini F, Burighel P. External amebocytes guard the pharynx entry in a tunicate (Ascidacea). *Dev Comp Immunol* (2006) 30:463–72. doi:10.1016/j.dci.2005.07.004
65. Parrinello D, Sanfratello MA, Vizzini A, Parrinello N, Cammarata M. *Ciona intestinalis* galectin (Cigals-a and Cigals-b) genes are differentially expressed in endostyle zones and challenged by LPS. *Fish Shellfish Immunol* (2015) 42:171–6. doi:10.1016/j.fsi.2014.10.026
66. Parrinello D, Sanfratello MA, Vizzini A, Tetasecca L, Parrinello N, Cammarata M. The *Ciona intestinalis* immune-related galectin genes (Cigals-a and Cigals-b) are expressed by the gastric epithelium. *Fish Shellfish Immunol* (2017) 62:24–30. doi:10.1016/j.fsi.2016.12.027
67. Dishaw LJ, Cannon JP, Litman GW, Parker W. Immune-directed support of rich microbial communities in the gut has ancient roots. *Dev Comp Immunol* (2014) 47:36–51. doi:10.1016/j.dci.2014.06.011
68. Tiozzo S, Ballarin L, Burighel P, Zaniolo G. Programmed cell death in vegetative development: apoptosis during the colonial life cycle of the ascidian *Botryllus schlosseri*. *Tissue Cell* (2006) 38:193–201. doi:10.1016/j.tice.2006.02.003
69. Vizzini A, Parrinello D, Sanfratello MA, Trapani MR, Mangano V, Parrinello N, et al. Upregulated transcription of phenoloxidase genes in the pharynx and endostyle of *Ciona intestinalis* in response to LPS. *J Invertebr Pathol* (2015) 126:6–11. doi:10.1016/j.jip.2015.01.009
70. Smith VJ, Peddie CM. Cell cooperation during host defense in the solitary tunicate *Ciona intestinalis* (L). *Biol Bull* (1992) 183:211–9. doi:10.2307/1542208
71. Jackson AD, Smith VJ. LPS-sensitive protease activity in the blood cells of the solitary ascidian, *Ciona intestinalis* (L). *Comp Biochem Physiol* (1993) 106B:505–12.
72. Müller WE, Pancer Z, Rinkevich B. Molecular cloning and localization of a novel serine protease from the colonial tunicate *Botryllus schlosseri*. *Mol Mar Biol Biotechnol* (1994) 3:70–7.
73. Cerenius L, Söderhäll K. The prophenoloxidase-activating system in invertebrates. *Immunol Rev* (2004) 198:116–26. doi:10.1111/j.0105-2896.2004.00116.x
74. Immesberger A, Burmester T. Putative phenoloxidase in the tunicate *Ciona intestinalis* and the origin of the arthropod hemocyanin superfamily. *J Comp Physiol* (2004) 174B:169–80. doi:10.1007/s00360-003-0402-4
75. Ballarin L, Franchi N, Schiavon F, Tosatto SC, Mičetić I, Kawamura K. Looking for putative phenoloxidases of compound ascidians: haemocyanin-like proteins in *Polyandrocarpa misakiensis* and *Botryllus schlosseri*. *Dev Comp Immunol* (2012) 38:232–42. doi:10.1016/j.dci.2012.05.008
76. Taylor SW, Ross MM, Waite JH. Novel 3,4-di- and 3,4,5-trihydroxyphenylalanine-containing polypeptides from the blood cells of the ascidian *Ascidia ceratodes* and *Molgula manhattensis*. *Arch Biochem Biophys* (1995) 324:228–40. doi:10.1006/abbi.1995.0035
77. Tincu JA, Taylor SW. Tunichrome sp-1: new pentapeptide tunichrome from the hemocytes of *Styela plicata*. *J Nat Prod* (2002) 65:377–8. doi:10.1021/np010352z
78. Rao MR, Faulkner DJ. Botryllamides E-H, four new tyrosine derivatives from the ascidian *Botrylloides tyreum*. *J Nat Prod* (2004) 67:1064–6. doi:10.1021/np0499618
79. Franchi N, Cima F, Ballarin L. Insights on cytotoxic cells of the colonial ascidian *Botryllus schlosseri*. *Invert Surviv J* (2015) 12:109–17.
80. Coombe DR, Ey PL, Jenkin CR. Ascidian haemagglutinins: incidence in various species, binding specificities and preliminary characterisation of selected agglutinins. *Comp Biochem Physiol* (1984) 77B:811–9.
81. Parrinello N. Humoral and cellular lectins of ascidians. *J Mar Biotechnol* (1995) 3:29–34.
82. Vasta GR, Quesenberry MS, Ahmed H, O'Leary N. Lectins from tunicates: structure-function relationships in innate immunity. *Adv Exp Med Biol* (2001) 484:275–87. doi:10.1007/978-1-4615-1291-2_26
83. Quesenberry MS, Ahmed H, Elola MT, O'Leary N, Vasta GR. Diverse lectin repertoires in tunicates mediate broad recognition and effector innate immune responses. *Integr Comp Biol* (2003) 43:323–30. doi:10.1093/icb/43.2.323
84. Ballarin L, Cammarata M, Franchi N, Parrinello N. Routes in innate immunity evolution: galectins and rhamnose-binding lectins in ascidians. In: Kim S-W, editor. *Marine Protein and Peptides. Biological Activities and Applications*. Chichester: Wiley-Blackwell (2013). p. 185–205.
85. Coombe DR, Ey PL, Jenkin CR. Particle recognition by haemocytes from the colonial ascidian *Botrylloides leachi*: evidence that the B. leachi HA-2 is opsonic. *J Comp Physiol* (1984) 154:509–21. doi:10.1007/BF02515156
86. Kelly KL, Cooper EL, Raftos DA. Purification and characterization of a humoral opsonin from the solitary urochordate *Styela clava*. *Comp Biochem Physiol* (1992) 103B:749–53.
87. Pearce S, Newton RA, Nair SV, Raftos DA. Humoral opsonins of tunicate, *Pyura stolonifera*. *Dev Comp Immunol* (2001) 25:377–85. doi:10.1016/S0145-305X(01)00011-8
88. Parrinello N, Arizza V, Cammarata M, Giamrita FT, Pergolizzi M, Vazzana M, et al. Inducible lectins with galectin properties and human IL1 α epitopes opsonize yeast during the inflammatory response of the ascidian *Ciona intestinalis*. *Cell Tissue Res* (2007) 329:379–90. doi:10.1007/s00441-007-0415-5
89. Azumi K, Ishimoto R, Fujita T, Nonaka M, Yokosawa H. Opsonin-independent and -dependent phagocytosis in the ascidian *Halocynthia roretzi*: galactose-specific lectin and complement C3 function as target-dependent opsonins. *Zoolog Sci* (2000) 17:625–32. doi:10.2108/zsj.17.625
90. Gasparini F, Franchi N, Spolaore B, Ballarin L. Novel rhamnose-binding lectins from the colonial ascidian *Botryllus schlosseri*. *Dev Comp Immunol* (2008) 32:1177–91. doi:10.1016/j.dci.2008.03.006
91. Khatun K, Becker M, Rinkevich B, Bosch TCG. Urochordates and the origin of natural killer cells: identification of a CD94/NKR-P1-related receptor in blood cells of *Botryllus*. *Proc Natl Acad Sci U S A* (2003) 100:622–7. doi:10.1073/pnas.0234104100
92. Parrinello N. Focusing on *Ciona intestinalis* (Tunicata) innate immune system. Evolutionary implications. *Invert Surviv J* (2009) 6:S46–57.
93. Pancer Z, Cooper L, Müller WEG. A tunicate (*Botryllus schlosseri*) cDNA reveals similarity to vertebrate antigen receptor. *Immunogenetics* (1996) 45:69–72. doi:10.1007/s002510050169
94. Pancer Z, Diehl-Seifert B, Rinkevich B, Müller WEG. A novel tunicate (*Botryllus schlosseri*) putative C-type lectin features an immunoglobulin domain. *DNA Cell Biol* (1997) 16:801–6.
95. Liberti A, Melillo D, Zucchetti I, Natale L, Dishaw LJ, Litman GW, et al. Expression of *Ciona intestinalis* variable region-containing chitin-binding proteins during development of the gastrointestinal tract and their role in host-microbe interactions. *PLoS One* (2014) 9:e94984. doi:10.1371/journal.pone.0094984
96. Fujita T. Evolution of the lectin-complement pathway and its role in innate immunity. *Nat Rev Immunol* (2002) 2:346–53. doi:10.1038/nri800
97. Nonaka M, Azumi K, Ji X, Namikawa-Yamada C, Sasaki M, Saiga H, et al. Opsonic complement component C3 in the solitary ascidian *Halocynthia roretzi*. *J Immunol* (1999) 162:387–91.
98. Marino R, Kimura Y, DeSantis R, Lambris JD, Pinto MR. Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian *Ciona intestinalis*. *Immunogenetics* (2002) 53:1055–64. doi:10.1007/s00251-001-0421-9
99. Raftos DA, Robbins J, Newton RA, Nair SV. A complement component C3a-like stimulates chemotaxis by hemocytes from an invertebrate chordate – the tunicate, *Pyura stolonifera*. *Comp Biochem Physiol* (2003) 134A:377–86. doi:10.1016/S1095-6433(02)00287-8
100. Franchi N, Ballarin L. Preliminary characterization of complement in a colonial tunicate: C3, Bf and inhibition of C3 opsonic activity by compstatin. *Dev Comp Immunol* (2014) 46:430–8. doi:10.1016/j.dci.2014.05.014
101. Giacomelli S, Melillo D, Lambris JD, Pinto MR. Immune competence of the *Ciona intestinalis* pharynx: complement system-mediate activity. *Fish Shellfish Immunol* (2012) 33:946–52. doi:10.1016/j.fsi.2012.08.003
102. Pinto MR, Chinnici CM, Kimura Y, Melillo D, Marino R, Spruce LA, et al. C1C3-1a-mediated chemotaxis in the deuterostome invertebrate *Ciona intestinalis* (Urochordata). *J Immunol* (2003) 171:5521–8. doi:10.4049/jimmunol.171.10.5521
103. Melillo D, Sfyrroera G, De Santis R, Graziano R, Marino R, Lambris JD, et al. First identification of a chemotactic receptor in an invertebrate species: structural and functional characterization of *Ciona intestinalis* C3a receptor. *J Immunol* (2006) 177:4132–40. doi:10.4049/jimmunol.177.6.4132
104. Nonaka M, Azumi K. Opsonic complement system of the solitary ascidian *Halocynthia roretzi*. *Dev Comp Immunol* (1999) 23:421–7. doi:10.1016/S0145-305X(99)00021-X

105. Raftos DA, Fabbro M, Nair SV. Exocytosis of a complement component C3-like protein by tunicate hemocytes. *Dev Comp Immunol* (2004) 28:181–90. doi:10.1016/S0145-305X(03)00136-8
106. Ji X, Namikawa-Yamada M, Nakanishi M, Sasaki M, Nonaka M. Molecular cloning of complement factor B from a solitary ascidian: unique combination of domains implicating ancient exon shuffling. *Immunopharmacology* (2000) 49:43. doi:10.1016/S0162-3109(00)80122-8
107. Yoshizaki FY, Ikawa S, Satake M, Satoh N, Nonaka M. Structure and the evolutionary implication of the triplicated complement factor B genes of a urochordate ascidian, *Ciona intestinalis*. *Immunogenetics* (2005) 56:930–42. doi:10.1007/s00251-004-0752-4
108. Nair SV, Pearce S, Green PL, Mahajan D, Newton RA, Raftos DA. A collectin-like protein from tunicates. *Comp Biochem Physiol* (2000) 125B:279–89. doi:10.1016/S0305-0491(99)00180-7
109. Green P, Luty A, Nair S, Radford J, Raftos D. A second form of collagenous lectin from the tunicate, *Styela plicata*. *Comp Biochem Physiol* (2006) 144B:343–50. doi:10.1016/j.cbpb.2006.03.011
110. Green PL, Nair SV, Raftos DA. Secretion of a collectin-like protein in tunicates is enhanced during inflammatory responses. *Dev Comp Immunol* (2003) 27:3–9. doi:10.1016/S0145-305X(02)00067-8
111. Sekine H, Kenjo A, Azumi K, Ohi G, Takahashi M, Kasukawa R, et al. An ancient lectin-dependent complement system in an ascidian: novel lectin isolated from the plasma of the solitary ascidian, *Halocynthia roretzi*. *J Immunol* (2001) 167:4504–10. doi:10.4049/jimmunol.167.8.4504
112. Ji X, Azumi K, Sasaki M, Nonaka M. Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian *Halocynthia roretzi*. *Proc Natl Acad Sci U S A* (1997) 94:6340–5. doi:10.1073/pnas.94.12.6340
113. Corey DM, Rosental B, Kowarsky M, Sinha R, Ishizuka KJ, Palmeri KJ, et al. Developmental cell death programs license cytotoxic cells to eliminate histocompatible partners. *Proc Natl Acad Sci U S A* (2016) 113:6520–5. doi:10.1073/pnas.1606276113
114. Pancer Z, Gershon H, Rinkevich B. Cloning of a urochordate cDNA featuring mammalian short consensus repeats (SCR) of complement-control protein superfamily. *Comp Biochem Physiol* (1995) 111B:625–32. doi:10.1016/0305-0491(95)00025-4
115. Cooper EL, Yao D. Diving for drugs: tunicate anticancer compounds. *Drug Discov Today* (2012) 17:636–48. doi:10.1016/j.drudis.2012.02.006
116. Davis AR, Bremner JB. Potential antifouling natural products from ascidians: a review. In: Fingerman M, Nagabhushanam R, Thompson M-F, editors. *Marine Biotechnology, Volume 3: Biofilms, Bioadhesion, Corrosion and Biofouling*. Boca Raton, FL: CRC Press (1999). p. 259–308.
117. Pisut DP, Pawlik JR. Anti-predatory chemical defenses of ascidians: secondary metabolites or inorganic acids? *J Exp Mar Bio Ecol* (2002) 270:203–14. doi:10.1016/S0022-0981(02)00023-0
118. Bryan PJ, McClintock JB, Slattery M, Rittschof DP. A comparative study of the non-acidic chemically mediated antifouling properties of three sympatric species of ascidians associated with seagrass habitats. *Biofouling* (2003) 19:235–45. doi:10.1080/0892701031000085222
119. Núñez-Pons L, Forestieri R, Nieto RM, Varela M, Nappo M, Rodríguez J, et al. Chemical defenses of tunicates of the genus *Aplidium* from the Weddell Sea (Antarctica). *Polar Biol* (2010) 33:1319–29. doi:10.1007/s00300-010-0819-7
120. Mayzel B, Haber M, Ilan M. Chemical defense against fouling in the solitary ascidian *Phallusia nigra*. *Biol Bull* (2014) 227:232–41. doi:10.1086/BBLv227n3p232
121. Martínez-García M, Díaz-Valdés M, Ramos-Esplá A, Salvador N, Lopez P, Larriba E, et al. Cytotoxicity of the ascidian *Cystodites dellechiaiei* against tumor cells and study of the involvement of associated microbiota in the production of cytotoxic compounds. *Mar Drugs* (2007) 5:52–70. doi:10.3390/md503052
122. Odate S, Pawlik JR. The role of vanadium in the chemical defense of the solitary tunicate, *Phallusia nigra*. *J Chem Ecol* (2007) 33:643–54. doi:10.1007/s10886-007-9251-z
123. Lippert H, Brinkmeyer R, Mülhaupt T, Iken K. Antimicrobial activity in sub-Arctic marine invertebrates. *Polar Biol* (2003) 26:591–600. doi:10.1007/s00300-003-0525-9
124. Selegim MHR, Lira SP, Kossuga MH, Batista T, Berlink RGS, Hajdu E, et al. Antibiotic, cytotoxic and enzyme inhibitory activity of crude extracts from Brazilian marine invertebrates. *Rev Bras Farmacogn* (2007) 17:287–318. doi:10.1590/S0102-695X2007000300002
125. Tadesse M, Gulliksen B, Strøm MB, Styrvold OB, Haug T. Screening for antibacterial and antifungal activities in marine benthic invertebrates from northern Norway. *J Invertebr Pathol* (2008) 99:286–93. doi:10.1016/j.jip.2008.06.009
126. Lu Y, Zhuang Y, Liu J. Mining antimicrobial peptides from small open reading frames in *Ciona intestinalis*. *J Pept Sci* (2014) 20:25–9. doi:10.1002/psc.2584
127. Azumi K, Yokosawa H, Ishii S. Halocyamines: novel antimicrobial tetrapeptide-like substances isolated from the hemocytes of the solitary ascidian *Halocynthia roretzi*. *Biochemistry* (1990) 29:159–65. doi:10.1021/bi00453a021
128. Lee IH, Zhao C, Nguen T, Menzel L, Waring AJ, Sherman MA, et al. Clavaspilin, an antibacterial and haemolytic peptide from *Styela clava*. *J Pept Res* (2001) 58:445–56. doi:10.1034/j.1399-3011.2001.10975.x
129. Menzel LP, Lee IH, Sjostrand B, Lehrer RI. Immunolocalization of clavansins in *Styela clava* hemocytes. *Dev Comp Immunol* (2002) 26:505–15. doi:10.1016/S0145-305X(02)00010-1
130. Zhao C, Liaw L, Lee IH, Lehrer RI. cDNA cloning of three cecropin-like antimicrobial peptides (Styelins) from the tunicate *Styela clava*. *FEBS Lett* (1997) 412:144–8. doi:10.1016/S0014-5793(97)00769-2
131. Lehrer RI, Tincu JA, Taylor SW, Menzel LP, Waring AJ. Natural peptide antibiotics from tunicates: structures, functions and potential uses. *Integr Comp Biol* (2003) 43:313–22. doi:10.1093/icb/43.2.313
132. Fedders H, Leippe M. A reverse search for antimicrobial peptides in *Ciona intestinalis*: identification of a gene family expressed in hemocytes and evaluation of activity. *Dev Comp Immunol* (2008) 32:286–98. doi:10.1016/j.dci.2007.06.003
133. Fedders H, Michalek M, Grötzinger J, Leippe M. An exceptional salt-tolerant antimicrobial peptide derived from a novel gene family of haemocytes of the marine invertebrate *Ciona intestinalis*. *Biochem J* (2008) 416:65–75. doi:10.1042/BJ20080398
134. Jena P, Mishra B, Leippe M, Hasilik A, Griffiths G, Sonawane A. Membrane-active antimicrobial peptides and human placental lysosomal extracts are highly active against mycobacteria. *Peptides* (2011) 32:881–7. doi:10.1016/j.peptides.2011.03.002
135. Cotellet N, Moreau S, Cotellet P, Cateau JP, Bernier JL, Hénichart JP. Generation of free radicals by simple preglylated hydroquinone derivatives, natural antitumor agents from the marine urochordate *Aplidium californicum*. *Chem Res Toxicol* (1991) 4:300–5. doi:10.1021/tx00021a007
136. Ogi T, Taira J, Margiastuti P, Ueda K. Cytotoxic metabolites from the Okinawan ascidian *Diplosoma virens*. *Molecules* (2008) 13:595–602. doi:10.3390/molecules13030595
137. Pancer Z, Cooper L, Müller WEG. A urochordate putative homolog of human EB1, the protein which binds APC. *Cancer Lett* (1996) 109:155–60. doi:10.1016/S0304-3835(96)04440-0
138. Beschin A, Bilej M, Torrelee E, De Baetselier P. On the existence of cytokines in invertebrates. *Cell Mol Life Sci* (2001) 58:801–14. doi:10.1007/PL00000901
139. Beschin A, Bilej M, Magez S, Lucas R, De Baetselier P. Functional convergence of invertebrate and vertebrate cytokine-like molecules based on a similar lectin-like activity. *Prog Mol Subcell Biol* (2004) 34:145–63. doi:10.1007/978-3-642-18670-7_6
140. Parrinello N, Vizzini A, Arizza V, Salerno G, Parrinello D, Cammarata M, et al. Enhanced expression of a cloned and sequenced *Ciona intestinalis* TNF α -like (CiTNF α) gene during the LPS-induced inflammatory response. *Cell Tissue Res* (2008) 334:305–17. doi:10.1007/s00441-008-0695-4
141. Parrinello N, Vizzini A, Salerno G, Sanfratello MA, Cammarata M, Arizza V, et al. Inflamed adult pharynx tissues and swimming larva of *Ciona intestinalis* share CiTNF α -producing cells. *Cell Tissue Res* (2010) 341:299–311. doi:10.1007/s00441-010-0993-5
142. Vizzini A, Di Falco F, Parrinello D, Sanfratello MA, Mazzarella C, Parrinello N, et al. *Ciona 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
143. Ballarin L, Franchini A, Ottaviani E, Sabbadin A. Morula cells as the major immunomodulatory hemocytes in ascidians: evidences from the colonial species *Botryllus schlosseri*. *Biol Bull* (2001) 201:59–64. doi:10.2307/1543526
144. Menin A, Ballarin L. Immunomodulatory molecules in the compound ascidian *Botryllus schlosseri*: evidence from conditioned media. *J Invertebr Pathol* (2008) 99:275–80. doi:10.1016/j.jip.2008.08.001

145. Menin A, Del Favero M, Cima F, Ballarin L. Release of phagocytosis-stimulating factor(s) by morula cells in a colonial ascidian. *Mar Biol* (2005) 148:225–30. doi:10.1007/s00227-005-0081-7
146. Cima F, Sabbadin A, Ballarin L. Cellular aspects of allorecognition in the compound ascidian *Botryllus schlosseri*. *Dev Comp Immunol* (2004) 28:881–9. doi:10.1016/j.dci.2004.02.001
147. Cima F, Sabbadin A, Zaniolo G, Ballarin L. Colony specificity and chemotaxis in the compound ascidian *Botryllus schlosseri*. *Comp Biochem Physiol* (2006) 145A:376–82. doi:10.1016/j.cbpa.2006.07.017
148. Takahashi H, Azumi K, Yokosawa H. Hemocyte aggregation in the solitary ascidian *Halocynthia roretzi*: plasma factors, magnesium ion, and Met-Lys-bradykinin induce the aggregation. *Biol Bull* (1994) 186:247–53. doi:10.2307/1542270
149. Takahashi H, Azumi K, Yokosawa H. A novel membrane glycoprotein involved in ascidian hemocyte aggregation and phagocytosis. *Eur J Biochem* (1995) 233:778–83. doi:10.1111/j.1432-1033.1995.778_3.x
150. Takahashi H, Ishikawa G, Ueki K, Azumi K, Yokosawa H. Cloning and tyrosine phosphorylation of a novel invertebrate immunocyte protein containing immunoreceptor tyrosine-based activation motifs. *J Biol Chem* (1997) 272:32006–10. doi:10.1074/jbc.272.51.32006
151. Azumi K, Sasaki T, Okochi K, Yamasaki S, Saito T, Takayama H, et al. Differential display analysis reveals the expression of glutathione S-transferase ω and novel genes through an ITAM-containing receptor in ascidian immunocytes. *Immunogenetics* (2005) 27:444–52. doi:10.1007/s00251-005-0003-3
152. Cima F, Ballarin L, Sabbadin A. New data on phagocytes and phagocytosis in the compound ascidian *Botryllus schlosseri*. *Ital J Zool* (1996) 63:357–64. doi:10.1080/11250009609356159
153. Ishikawa G, Azumi K, Yokosawa H. Involvement of tyrosine kinase and phosphatidylinositol 3-kinase in phagocytosis by ascidian hemocytes. *Comp Biochem Physiol* (2000) 125A:351–7. doi:10.1016/S1095-6433(00)00165-3
154. Franchi N, Schiavon F, Betti M, Canesi L, Ballarin L. Insight on signal transduction pathways involved in phagocytosis in the colonial ascidian *Botryllus schlosseri*. *J Invertebr Pathol* (2013) 112:260–6. doi:10.1016/j.jip.2012.12.001
155. Manni L, Zaniolo G, Cima F, Burighel P, Ballarin L. *Botryllus schlosseri*: a model ascidian for the study of asexual reproduction. *Dev Dyn* (2007) 236:335–52. doi:10.1002/dvdy.21037
156. Lauzon RJ, Ishizuka KJ, Weissman IL. A cyclical, developmentally-regulated death phenomenon in a colonial urochordate. *Dev Dyn* (1992) 194:71–83. doi:10.1002/aja.1001940109
157. Lauzon RJ, Patton CW, Weissman IL. A morphological and immunohistochemical study of programmed cell death in *Botryllus schlosseri* (Tunicata, Ascidiacea). *Cell Tissue Res* (1993) 272:115–27. doi:10.1007/BF00323577
158. Ballarin L, Burighel P, Cima F. A tale of death and life: natural apoptosis in the colonial ascidian *Botryllus schlosseri* (Urochordata Ascidiacea). *Curr Pharm Des* (2008) 14:138–47. doi:10.2174/138161208783378798
159. Franchi N, Ballin F, Manni L, Schiavon F, Basso G, Ballarin L. Recurrent phagocytosis-induced apoptosis in the ciclica generation change of the compound ascidian *Botryllus schlosseri*. *Dev Comp Immunol* (2016) 62:8–16. doi:10.1016/j.dci.2016.04.011
160. Voskoboinik A, Rinkevich B, Weiss A, Moiseeva E, Reznick AZ. Macrophage involvement for successful degeneration of apoptotic organs in the colonial urochordate *Botryllus schlosseri*. *J Exp Biol* (2004) 207:2409–16. doi:10.1242/jeb.01045
161. Lauzon RJ, Shizuka KJ, Weissman IL. Cyclical generation and degeneration of organs in a colonial urochordate involves crosstalk between old and new: a model for development and regeneration. *Dev Biol* (2002) 249:333–48. doi:10.1006/dbio.2002.0772
162. Lauzon RJ, Kidder SJ, Long P. Suppression of programmed cell death regulates the cyclical degeneration of organs in a colonial urochordate. *Dev Biol* (2007) 301:92–105. doi:10.1016/j.ydbio.2006.08.055
163. Shirae M, Hirose E, Saito Y. Behavior of hemocytes in the allorecognition reaction in two compound ascidians *Botryllus scalaris* and *Symplesma reptans*. *Biol Bull* (1999) 197:188–97. doi:10.2307/1542614
164. Parrinello N, Cammarata M, Lipari L, Arizza V. Sphingomyelin inhibition of *Ciona intestinalis* (Tunicata) cytotoxic hemocytes assayed against sheep erythrocytes. *Dev Comp Immunol* (1995) 19:31–41. doi:10.1016/0145-305X(94)00046-I
165. Lipari L, Cammarata M, Arizza V, Parrinello D. Cytotoxic activity of *Styela plicata* hemocytes against mammalian cell targets: I. Properties of the in vitro reaction against erythrocytes. *Anim Biol* (1995) 4:131–7.
166. Cammarata M, Candore G, Arizza V, Caruso C, Parrinello N. Cytotoxic activity of *Styela plicata* hemocytes against mammalian cell targets: II properties of the in vitro reaction against human tumour cell lines. *Anim Biol* (1995) 4:139–44.
167. Arizza V, Parrinello D, Cammarata M, Vazzana M, Vizzini A, Giaramita FT, et al. A lytic mechanism based on soluble phospholipases A2 (sPLA2) and b-galactoside specific lectins is exerted by *Ciona intestinalis* (ascidian) unilocular refractile hemocytes against K562 cell line and mammalian erythrocytes. *Fish Shellfish Immunol* (2011) 30:1014–23. doi:10.1016/j.fsi.2011.01.022
168. Fuke M. “Contact reactions” between xenogeneic or allogeneic coelomic cells of solitary ascidians. *Biol Bull* (1980) 158:304–15. doi:10.2307/1540857
169. De Barros CM, De Carvalho DR, Andrade LR, Pavao MSG, Allodi S. Nitric oxide production by hemocytes of the ascidian *Styela plicata*. *Cell Tissue Res* (2009) 338:117–28. doi:10.1007/s00441-009-0851-5
170. De Barros CM, Emrich LC, de A Mello A, Da Fonseca RN, Allodi S. Regulation of nitric-oxide production in hemocytes of the ascidian *Phallusia nigra*. *Nitric Oxide* (2014) 38:26–36. doi:10.1016/j.niox.2014.02.007
171. Cammarata M, Arizza V, Cianciolo C, Parrinello D, Vazzana M, Vizzini A, et al. The phenoloxidase system is activated during the tunic inflammatory reaction of *Ciona intestinalis*. *Cell Tissue Res* (2008) 333:481–92. doi:10.1007/s00441-008-0649-x
172. Trapani MR, Sanfratello MA, Mangano V, Parrinello D, Vizzini A, Cammarata M. Phenoloxidases of different sizes are modulated by LPS inoculation into *Ciona intestinalis* tunic and pharynx. *Invert Surviv J* (2015) 12:75–81.
173. Vizzini A, Parrinello D, Sanfratello MA, Mangano V, Parrinello N, Cammarata M. *Ciona intestinalis* peroxinectin is a novel component of the peroxidase-cyclooxygenase superfamily upregulated by LPS. *Dev Comp Immunol* (2013) 41:59–67. doi:10.1016/j.dci.2013.03.015
174. Vizzini A, Bonura A, Longo V, Sanfratello MA, Parrinello D, Cammarata M, et al. Isolation of a novel LPS-induced component of the ML superfamily in *Ciona intestinalis*. *Dev Comp Immunol* (2015) 53:70–8. doi:10.1016/j.dci.2015.06.018
175. Vizzini A, Di Falco F, Parrinello D, Sanfratello MA, Cammarata M. Transforming growth factor β (TGF- β) gene expression is induced in the inflammatory reaction of *Ciona intestinalis*. *Dev Comp Immunol* (2016) 55:102–10. doi:10.1016/j.dci.2015.10.013
176. Reddy AL, Bryan B, Hildemann WH. Integumentary allograft versus autograft reactions in *Ciona intestinalis*: a protochordate species of solitary tunicata. *Immunogenetics* (1975) 7:584–90.
177. Raftos DA, Tait NN, Briscoe DA. Allograft rejection and alloimmune memory in the solitary urochordate *Styela plicata*. *Dev Comp Immunol* (1987) 11:343–51. doi:10.1016/0145-305X(87)90078-4
178. Raftos DA, Tait NN, Briscoe DA. Cellular basis of allograft rejection in the solitary urochordate *Styela plicata*. *Dev Comp Immunol* (1987) 11:713–25. doi:10.1016/0145-305X(87)90078-4
179. Parrinello N. Cytotoxic activity of tunicate hemocytes. In: Rinkevich B, Muller WEG, editors. *Invertebrate Immunology*. Berlin: Springer-Verlag (1996). p. 190–217.
180. Kürn U, Sommer F, Hemmrich G, Bosch TCG, Kalthurin K. Allorecognition in urochordates: identification of a highly variable complement receptor-like protein expressed in follicle cells of *Ciona*. *Dev Comp Immunol* (2007) 31:360–71. doi:10.1016/j.dci.2006.06.008
181. Sommer F, Awazu S, Anton-Exleben F, Jiang D, Klimovich AV, Klimovich BV, et al. Blood system formation in the urochordate *Ciona intestinalis* requires the variable receptor vCRL1. *Mol Biol Evol* (2012) 29:3081–93. doi:10.1093/molbev/mss120
182. Laird DJ, De Tomaso AW, Weissman IL. Stem cells are units of natural selection in a colonial ascidian. *Cell* (2005) 123:1351–60. doi:10.1016/j.cell.2005.10.026
183. De Tomaso AW. Allorecognition polymorphism versus parasitic stem cells. *Trends Genet* (2006) 22:485–90. doi:10.1016/j.tig.2006.07.001
184. Sabbadin A. Le basi genetiche della capacità di fusione fra colonie in *Botryllus schlosseri* (Ascidiacea). *Rend Accad Naz Lincei* (1962) 32:1021–35.
185. Oka H. Colony specificity in compound ascidians. The genetic control of fusibility. In: Yukawa H, editor. *Profiles of Japanese Science and Scientists*. Tokyo: Kodansha (1970). p. 196–206.
186. Rinkevich B, Tartakover S, Gershon H. Contribution of morula cells to allogeneic responses in the colonial ascidian *Botryllus schlosseri*. *Mar Biol* (1998) 131:227–36. doi:10.1007/s002270050315

187. Rinkevich B. Aspects of the incompatibility nature in botryllid ascidian. *Anim Biol* (1992) 1:17–28.
188. Oren M, Douek J, Fishelson Z, Rinkevich B. Identification of immune-relevant genes in histoincompatible rejecting colonies of the tunicate *Botryllus schlosseri*. *Dev Comp Immunol* (2007) 31:889–902. doi:10.1016/j.dci.2006.12.009
189. Rinkevich B, Weissman IL. Retreat growth in the ascidian *Botryllus schlosseri*: a consequence of nonself recognition. In: Grosberg RK, Hedgecock D, Nelson K, editors. *Invertebrate Historecognition*. New York: Plenum Press (1988). p. 93–109.
190. Dias GM, Yokohama Q. Spatial competition induces the mobilization of morula cells in the colonial ascidian *Didemnum perlucidum* (Tunicata: Didemnidae). *Invertebr Biol* (2011) 130:186–92. doi:10.1111/j.1744-7410.2011.00222.x
191. Rinkevich B, Lilker-Levav T, Goren M. Allorecognition/xenorecognition responses in *Botrylloides* (Ascidacea) subpopulations from the Mediterranean coast of Israel. *J Exp Zool* (1994) 270:302–13. doi:10.1002/jez.1402700309
192. Hirose E, Saito Y, Watanabe H. Surgical fusion between incompatible colonies of the compound ascidian *Botrylloides fuscus*. *Dev Comp Immunol* (1994) 18:287–94. doi:10.1016/S0145-305X(94)90354-9
193. Saito Y, Hirose E, Watanabe H. Allorecognition in compound ascidians. *Int J Dev Biol* (1994) 38:237–47.
194. Hirose E. Colonial allorecognition, hemolytic rejection, and viviparity in botryllid ascidians. *Zoolog Sci* (2003) 20:387–94. doi:10.2108/zsj.20.387
195. Shirae M, Saito Y. A comparison of hemocytes and their phenoloxidase activity among botryllid ascidians. *Zoolog Sci* (2000) 17:881–91. doi:10.2108/zsj.17.881
196. Okuyama M, Saito Y, Hirose E. Fusion between incompatible colonies of a viviparous ascidian *Botrylloides lentus*. *Invertebr Biol* (2002) 121:163–9. doi:10.1111/j.1744-7410.2002.tb00057.x
197. Katow H, Watanabe H. Fine structure of fusion reaction in the compound ascidian *Botryllus primigenus* Oka. *Dev Biol* (1980) 76:1–14. doi:10.1016/0012-1606(80)90358-9
198. Rinkevich B, Weissman IL. Allogeneic resorption in colonial protochordates: consequences of non-self recognition. *Dev Comp Immunol* (1992) 16:275–86. doi:10.1016/0145-305X(92)90002-T
199. McKittrick TR, Muscat CC, Pierce JD, Bhattacharya D, De Tomaso AW. Allorecognition in a basal chordate consists of independent activating and inhibitory pathways. *Immunity* (2011) 34:616–26. doi:10.1016/j.immuni.2011.01.019
200. Rinkevich B, Douek J, Rabinowitz C, Paz G. The candidate Fu/HC gene in *Botryllus schlosseri* (Urochordata) and ascidians' historecognition – an oxymoron? *Dev Comp Immunol* (2012) 36:718–27. doi:10.1016/j.dci.2011.10.015
201. Nydam ML, Netuschil N, Sanders E, Langenbacher A, Lewis DD, Taketa DA, et al. The candidate histocompatibility locus of a basal chordate encodes two highly polymorphic proteins. *PLoS One* (2013) 8:e65980. doi:10.1371/journal.pone.0065980
202. Voskoboinik A, Newman AM, Corey DM, Sahoo D, Pushkarev D, Neff NF, et al. Identification of a colonial chordate histocompatibility gene. *Science* (2013) 341:384–7. doi:10.1126/science.1238036
203. Taketa DA, De Tomaso AW. *Botryllus schlosseri* allorecognition: tackling the enigma. *Dev Comp Immunol* (2015) 48:254–65. doi:10.1016/j.dci.2014.03.014
204. Taketa DA, Nydam ML, Langenbacher AD, Rodriguez D, Sanders E, De Tomaso AW. Molecular evolution and *in vitro* characterization of *Botryllus* histocompatibility factor. *Immunogenetics* (2015) 67:605–23. doi:10.1007/s00251-015-0870-1
205. Lemaitre B, Nicolas E, Michaut L, Reichhart JM, Hoffmann JA. The dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* (1996) 86:973–83. doi:10.1016/S0092-8674(00)80172-5
206. Roberts B, Davidson B, MacMaster G, Lockhart V, Ma E, Wallace SS, et al. A complement response may activate metamorphosis in the ascidian *Boltenia villosa*. *Dev Genes Evol* (2007) 217:449–58. doi:10.1007/s00427-007-0157-0
207. Robinson WE, Kustin K, Cloney RA. The influence of tunichrome and other reducing compounds on tunic and fin formation in embryonic *Ascidia callosa* Stimpson. *J Exp Zool* (1986) 237:63–72. doi:10.1002/jez.1402370110
208. Hibino T, Nonaka M. A novel third complement component C3 gene of *Ciona intestinalis* expressed in the endoderm at the early developmental stages. *Invert Surv J* (2013) 10:29–37.
209. Parrinello D, Sanfratello MA, Vizzini A, Cammarata M. The expression of an immune-related phenoloxidase gene is modulated in *Ciona intestinalis* ovary, test cells, embryos and larva. *J Exp Zool* (2015) 342B:141–51. doi:10.1002/jez.b.22613
210. 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
211. Kawai N, Takahashi H, Nishida H, Yokosawa H. Regulation of NF-kappaB/Rel by IkappaB is essential for ascidian notochord formation. *Dev Biol* (2005) 277:80–91. doi:10.1016/j.ydbio.2004.09.007
212. Rosner A, Alfassi G, Moiseeva E, Paz G, Rabinowitz C, Lapidot Z, et al. The involvement of three signal transduction pathways in botryllid ascidian astogeny, as revealed by expression patterns of representative genes. *Int J Dev Biol* (2014) 58:677–392. doi:10.1387/ijdb.140114ar

Conflict of Interest Statement: 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.

Copyright © 2017 Franchi and Ballarin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Dscam1 in Pancrustacean Immunity: Current Status and a Look to the Future

Sophie A. O. Armitage^{1*†}, Joachim Kurtz^{1*†}, Daniela Brites^{2,3†}, Yuemei Dong^{4†}, Louis Du Pasquier^{3†} and Han-Ching Wang^{5†}

¹Institute for Evolution and Biodiversity, University of Münster, Münster, Germany, ²Tuberculosis Research Unit, Swiss Tropical and Public Health Institute, Basel, Switzerland, ³Zoological Institute, University of Basel, Basel, Switzerland, ⁴Department of Molecular Microbiology and Immunology, Bloomberg School of Public Health, John Hopkins University, Baltimore, MD, United States, ⁵Department of Biotechnology and Bioindustry Sciences, College of Bioscience and Biotechnology, National Cheng Kung University, Tainan, Taiwan

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, United States

Reviewed by:

Humberto Lanz-Mendoza,
Instituto Nacional de Salud Pública,
Mexico
Jonathan P. Rast,
Sunnybrook Research Institute,
Canada

*Correspondence:

Sophie A. O. Armitage
sophie.armitage@uni-muenster.de;
Joachim Kurtz
joachim.kurtz@uni-muenster.de

[†]These authors are first co-authors.

[‡]Authors in alphabetical order.

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 28 February 2017

Accepted: 19 May 2017

Published: 09 June 2017

Citation:

Armitage SAO, Kurtz J, Brites D,
Dong Y, Du Pasquier L and
Wang H-C (2017) *Dscam1* in
Pancrustacean Immunity: Current
Status and a Look to the Future.
Front. Immunol. 8:662.
doi: 10.3389/fimmu.2017.00662

The *Down syndrome cell adhesion molecule 1* (*Dscam1*) gene is an extraordinary example of diversity: by combining alternatively spliced exons, thousands of isoforms can be produced from just one gene. So far, such diversity in this gene has only been found in insects and crustaceans, and its essential part in neural wiring has been well-characterized for *Drosophila melanogaster*. Ten years ago evidence from *D. melanogaster* showed that the *Dscam1* gene is involved in insect immune defense and work on *Anopheles gambiae* indicated that it is a hypervariable immune receptor. These exciting findings showed that *via* processes of somatic diversification insects have the possibility to produce unexpected immune molecule diversity, and it was hypothesized that *Dscam1* could provide the mechanistic underpinnings of specific immune responses. Since these first publications the quest to understand the function of this gene has uncovered fascinating insights from insects and crustaceans. However, we are still far from a complete understanding of how *Dscam1* functions in relation to parasites and pathogens and its full relevance for the immune system. In this Hypothesis and Theory article, we first briefly introduce *Dscam1* and what we know so far about how it might function in immunity. By focusing on seven questions, we then share our sometimes contrasting thoughts on what the evidence tells us so far, what essential experiments remain to be done, and the future prospects, with the aim to provide a multiangled view on what this fascinating gene has to do with immune defense.

Keywords: alternative splicing, crustaceans, isoform diversity, immunoglobulin domain, innate immunity, insects

INTRODUCTION

Dscam1: Mutually Exclusive Alternative Splicing Generates Isoform Diversity

There are few genes that encode for such extreme molecular diversity as *Dscam1*, the insect and crustacean homolog of the human Down syndrome cell adhesion molecule (DSCAM) (1, 2). Although *Dscam1* shares homology with DSCAM, only *Dscam1* has evolved the possibility to produce a profusion of protein isoforms. Since *Dscam1* was discovered as a cell surface hypervariable axon

guidance receptor in *D. melanogaster* (2), our knowledge of its functions in the nervous system and the immune system as well as its evolution across insects and crustaceans (i.e., the subgroup of the arthropods that is called Pancrustacea) has expanded extensively [reviewed in Ref. (3–10)], yet we are still far from a complete understanding of how *Dscam1* reacts and responds to parasites and pathogens.

Dscam1 [synonymous with *Dscam*, *Dscam-hypervariable* (*Dscam-hv*) and species-specific notations, e.g., the shrimp *Litopenaeus vannamei* *Dscam1* has been named *LvDscam*] has a complex gene structure whereby clusters of alternative exons encode for different immunoglobulins (Igs) domains (Figure 1). As an example, in *D. melanogaster* exons 4, 6, and 9 have numerous alternative sequences (2). Exon 4 has evolved 12 alternative variants, exon 6 has 48 [of which 47 are transcribed (11–13)], and exon 9 has 33 variants (Figure 1A). The number of alternative variants is not conserved across species, but the existence of multiple variants within three exon clusters is consistent across all pancrustaceans studied to date. However, in species other than *D. melanogaster*, the orthologous exon clusters sometimes have different numbering [e.g., exons 4, 6, and 10 in *Anopheles gambiae* (14)] because of differing positions of exon–exon boundaries. The pre-mRNA undergoes mutually exclusive alternative

splicing, so that each mRNA contains only one of the possible variants from each of the three alternative exon clusters (Figure 1B). Across species, the alternatively spliced exons code for the N-terminal halves of Ig2 and Ig3 and the whole of Ig7 (Figure 1C). These Ig domains are located in the extracellular portion of the protein. Mutually exclusive alternative splicing of the exons encoding the extracellular region, could potentially lead to the production of $12 \times 48 \times 33 = 19,008$ gene isoforms (18,612 if the non-transcribed exon in cluster 6 is excluded). If exon 17, which has two alternatively spliced variants and encodes the transmembrane region of the protein, and exons 19 and 23, which can be contained within or skipped from the cytoplasmic region of the protein (15), are included in the isoform diversity calculation, the estimate increases to just under 150,000 gene isoforms. This is an incredible amount of diversity to be expressed by just one gene.

Involvement in the Nervous System

Our knowledge about *Dscam1*'s function in the nervous system comes predominantly from research on *D. melanogaster*, where it has been extensively reviewed [e.g., Ref. (3–5, 9, 16, 18, 19)]. The diverse extracellular domains of the *Dscam1* protein facilitate its function as a molecular surface code, which enables neurites to

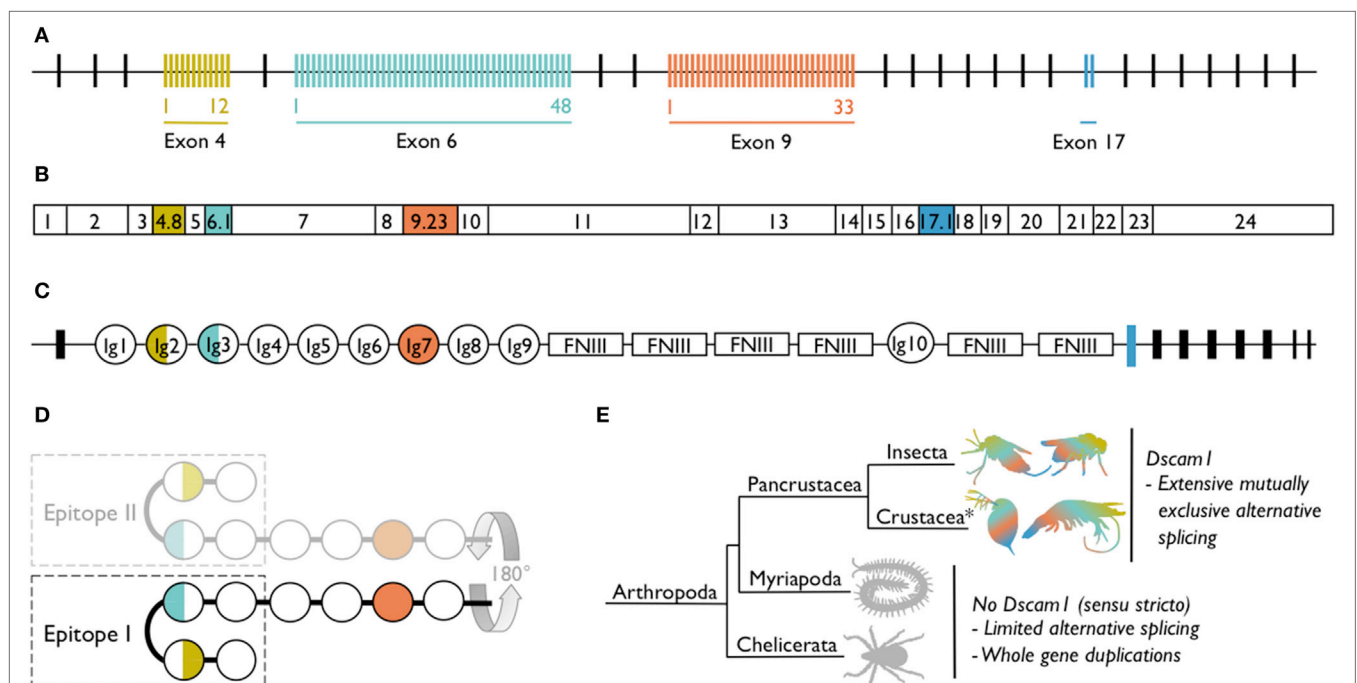


FIGURE 1 | *Dscam1* in *Drosophila melanogaster* and known occurrence of *Dscam1* in arthropods. **(A)** *D. melanogaster* *Dscam1* genomic DNA structure contains 20 constant exons (black lines). Four exon clusters contain variable numbers of alternative exons (colored lines): exon 4 contains 12, exon 6 contains 48, exon 9 contains 33, and exon 17 contains 2 variants. **(B)** *Dscam1* mRNA contains every constant exon (white boxes), but through the process of mutually exclusive alternative splicing, only one of each of the alternative exons is present in each mRNA; one exon combination for *D. melanogaster* is illustrated. **(C)** *Dscam1* protein structure, where Ig indicates an immunoglobulin domain and FNIII indicates a fibronectin type III domain. The alternatively spliced exons encode the N-terminal halves of Ig2 and Ig3, all of Ig7, and the transmembrane domain. **(D)** Ig1 to Ig4 form a horseshoe configuration (24). Epitope I is one side of the horseshoe and in the nervous system engages in homophilic binding with identical *Dscam1* isoforms coded for by the identical exon 4, 6, and 9 variants; the other side of the horseshoe, epitope II, has been proposed to bind to non-*Dscam1* ligands, i.e., pathogen-related ligands. [(A–D) after (16)]. **(E)** *Dscam1* as illustrated in (A–C) has, to date, only been found in pancrustaceans. Myriapods and chelicerates have diversified the *Dscam* gene family via other routes. *Crustacea is considered a paraphyletic group containing the hexapods; phylogeny follows Legg et al. (17).

tell self from non-self, thus avoiding neuronal self-connectivity (5). Homophilic binding, i.e., binding between identical isoforms and subsequent repulsion, is the key to *Dscam1*'s function in self-recognition. In brief, expression analyses have estimated that individual cells produce a much reduced portion of the total number of possible isoforms, i.e., in the order of tens of isoforms, and that cells produce a suite of isoforms that are different to their neighboring cells (11). These two points are important because they make it highly likely that sister neurites from the same neuron will express identical *Dscam1* isoforms that will in turn differ from the neighboring cells; identical isoforms will bind to each other, but not at all, or only weakly, to non-identical isoforms (20, 21). Once identical isoforms have interacted, the protein's endodomain (cytoplasmic tail) converts isoform recognition into repulsion between the sister neurites and promotes self-avoidance (22). In the nervous system the identity of each isoform, i.e., the combination of exons 4, 6, and 9 that encode it, does not matter, but it is essential that neighboring neurons express different isoforms from one another (23). Non-self recognition is thereby effectively a game of probabilities, where the number of potential *Dscam1* isoforms and the number of and stochasticity with which cells express *Dscam1* isoforms determine the rules. In contrast to the nervous system, if *Dscam1* diversity affords the immune system the ability to discriminate between different pathogens, it is hypothesized that the identity of individual isoforms does matter.

One Protein, Two Roles?

Elucidation of the protein structure of *D. melanogaster* *Dscam1* hinted at how one protein might function in both the nervous system and the immune system. *D. melanogaster* *Dscam1* Ig1 to Ig4 form a horseshoe configuration, with an independent interaction surface on either side of the horseshoe (24). Some of the amino acids encoded for by exons 4 and 6 can be found on either one (epitope I) or the other (epitope II; **Figure 1D**) interaction surfaces. By swapping peptide segments, Meijers et al. (24) found that it is epitope I that engages in homophilic binding specificity, whereas epitope II was hypothesized to bind to non-*Dscam1* ligands, i.e., heterophilic binding. It has not to date been demonstrated empirically that epitope II binds to non-*Dscam1* ligands, but it has been hypothesized that the ligands could be antigens, thereby affording *Dscam1* an immune receptor function. It was then discovered that upon homophilic binding, Ig5–Ig8 also form a turn in the protein but in the opposite direction to Ig1–Ig4, which means that Ig1–Ig8 together make up a serpentine or “S” shape, binding homophilically in an antiparallel manner (25). Similar to Ig2 and Ig3, it is not known whether Ig7 is involved in heterophilic binding.

Dscam Diversity in Arthropods

Taking a broader phylogenetic perspective, diversity is a common theme in the *Dscam* gene family. Although to date only pancrustaceans have been found to share *Dscam1* and its extreme somatic diversification, species from two other arthropod taxa, chelicerates (e.g., ticks), and myriapods (e.g., centipedes) have evolved diversity *via* whole gene duplication and some degree of alternative splicing (**Figure 1E**). For example, *Dscam* gene duplication

in the centipede *Strigamia maritima* genome is estimated to have led to 60–80 *Dscam* genes and in the tick *Ixodes scapularis* to between 13 and 27 (26, 27). To date, there is no evidence of arrays of duplicated exons in chelicerate or myriapod *Dscam* Ig2 or Ig3; however, mutually exclusive alternative splicing does occur in the exons encoding for Ig7 in at least one *S. maritima* *Dscam* gene, and duplicated exons coding for Ig7 and Ig8 were found in four *I. scapularis* genes (26). Furthermore, two *Dscam* gene subfamilies have also recently been uncovered in the Chinese scorpion *Mesobuthus martensii*; the genes are shorter than *Dscam1* but contain Ig domains that correspond to Ig7 or Ig7 and Ig8, as well as multiple tandem exon arrays (28).

Dscam1 and Immune Defense in Pancrustaceans

Pancrustaceans do not have the same mechanisms for acquired (adaptive) immune defenses as vertebrates, i.e., somatic generation of receptor diversity by V(D)J joining of antibody genes followed by clonal selection of antigen-specific lymphocytes (29), which underlie immunological memory. They instead rely on the evolution of diverse innate immune defenses, which share a number of conserved features with the innate defenses of vertebrates (27, 30). Nonetheless, some pancrustaceans and other non-vertebrates show evidence of a phenomenon similar to immune memory, termed “immune priming” (31), and they can also somatically generate a limited amount of receptor diversity by alternative splicing [e.g., Ref. (32)], albeit that this diversity is many orders of magnitude lower than in vertebrates.

The link between *Dscam1* and pancrustacean immunity has been extensively reviewed in the last few years and we refer readers to the following reviews for more details (6–8, 10). Here we briefly describe evidence linking *Dscam1* to the immune-function hypotheses that have been proposed. Early studies on *Dscam1* in *D. melanogaster* (12) hypothesized that it may function as a signaling receptor or coreceptor during phagocytosis and potentially as an opsonin [i.e., bind to the surface of a pathogen, facilitating its phagocytosis (29)]. Following this, work on *A. gambiae* also suggested the hypothesis that *Dscam1* could act as a hypervariable pattern-recognition receptor for the immune system (14). Consistent with *Dscam1* playing a role as an opsonin, a shorter soluble *Dscam1* protein was found in S2 cell line-conditioned medium and also in haemolymph serum (12). Furthermore, *Dscam1* in the shrimps *L. vannamei* and *Penaeus monodon* and the Chinese mitten crab *Eriocheir sinensis* lacks the transmembrane domain and cytoplasmic tail and has been suggested to be secreted directly into the haemolymph (33–35). It was also shown that recombinantly expressed *Dscam1* protein binds to pathogens [(12), (36) this article has since been retracted], (37)].

Reducing *Dscam1* *via* RNA interference (RNAi), or mutation, or antibody blocking of *Dscam1* function, lead to the reduced phagocytosis of dead bacteria (12, 14), and to the hypothesis that *Dscam1* acts as a phagocytosis receptor. The membrane-bound protein has been hypothesized to interact directly with the bacteria or it could interact with an opsonizing *Dscam1* that has already bound to a pathogen (6, 38). Dong et al. (14) also

showed that after *A. gambiae* infection with bacteria, a fungus, or protozoan parasites, *Dscam1* exon 4 produces distinct mRNA splice variants in response to each antigen (exons 6 and 10 were not tested). *Dscam1* has generated intense interest in the field of pancrustacean (ecological) immunology largely because it was hypothesized that the somatic diversity generated by this gene has a function in the recognition of diverse parasite and pathogen antigens. However, we are far from understanding whether, and if so how, this might be the case.

Our Aims

In this Hypothesis and Theory article, we bring together the ideas of six researchers who have contributed toward our current knowledge on *Dscam1* in immune defense. Through our responses to seven questions, we discuss different perspectives and hypotheses on what the evidence tells us so far and our ideas for future progress on this controversial topic.

Question 1. Looking Back at 10 years of Research on a Potential Immune Function of *Dscam1*, Do You Think That *Dscam1* Has a Role in Pancrustacean Immunity?

Daniela Brites and Louis Du Pasquier

Yes. The main reason why we think that *Dscam1* has an immunological role is the fact that the diversity of its repertoires (splice variants) that are expressed in cells of the nervous system and in cells involved in immunity (fat body cells and equivalent and hemocytes) are different. *Dscam1* might, therefore, fulfill functions that are specific to each of these systems, rather than have a single general purpose, which would have been suggested by identical repertoires in both tissues. Differences in exon expression patterns between the nervous and immune systems have been observed both in *Drosophila* and in *Daphnia* (12, 39). There is a lower diversity in the immune system than in the nervous system. This argues that the two repertoires are under different selection pressures and/or constraints. This restrictive evaluation may look a little provocative, but indeed the role of *Dscam1* in immunity is still mysterious. As many reviews written recently point out, the situation remains unclear, with pros and cons (6, 7, 10, 40). There are too many contradictory reports concerning: (1) *Dscam1* expression, whether monitored at the RNA level by PCR or at the protein level in binding assays or Western blots (up, down regulation or no change following stimulation); (2) the immunological specificity of its isoforms and the amplification of selected isoforms that has not been reproduced or convincingly demonstrated following exposure to parasites or other antigens; (3) its role as a phagocytic receptor; (4) the mode of signaling suggested by the composition of its cytoplasmic segment. The situation is complicated by the fact that *Dscam1* in pancrustaceans is not encoded in a uniform way (26). There are major differences in gene numbers and in types of alternative splicing from chelicerates to pancrustaceans (28). Even within pancrustaceans there could be room for differences in the mode of expression (e.g., importance of the soluble form) resulting in modulations of *Dscam1* role in immunity.

Yuemei Dong

Dscam1 in insects was first characterized as a highly diverse axon guidance molecule in the neuron system of fruit flies (2, 4, 5). In the past decade, the studies of *Dscam1* in mosquitoes and other pancrustaceans have established it as an essential hyper-variable pattern recognition receptor (PRR) of the innate immune system, mainly contributed by the extraordinary splice form generation at the molecular level (12, 14, 36, 37, 39).

Han-Ching Wang

As a crustacean immunologist, I think there is now considerable evidence to suggest that *Dscam1* might be involved in immunity against non-self molecules in long-lived crustaceans such as shrimp. *Dscam1* shows a typical fast (2–6 h) non-specific immune response to pathogen-associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS) and beta-1,3-glucan (41, 42), but unlike most innate immune factors, *Dscam1* is not always induced immediately after immune stimulation. Instead, viruses and bacteria usually take more than 24 h to induce elevated *Dscam1* levels (37, 43, 44). In crayfish, this increased expression usually reaches a maximum after 5 days and then falls back to baseline levels (44). However, overall expression levels are not the only indication of *Dscam1*'s role in immunity, and it now appears that the correct combination of *Dscam1* isoforms might be more important. For instance, we have found that some of the pathogen-induced *Dscam1* isoforms induced after challenge with a particular pathogen show significantly greater binding ability to that same pathogen (37). We have also found that the haemolymph taken from “super-survivor” crayfish within 1 month of white spot syndrome virus (WSSV) challenge can provide protection to other animals against the same virus (44). Furthermore, when the *Dscam1* in this haemolymph is blocked, this protection is lost (44). At the very least, it therefore seems that crayfish *Dscam1* shows an ability to support an extended, specific anti-virus immune response.

Sophie A. O. Armitage and Joachim Kurtz

It is clear that *Dscam1* is involved in immune defense in some contexts in some insects and crustaceans. However, is difficult with the current data to determine exactly what this role is, how important it is, and the generality of its importance (7, 10). For example, immune gene expression (total or alternatively spliced variants) after exposure to a pathogen or parasite shows varied results across studies [reviewed in Ref. (7, 10)], and gene knock-down can reduce survival after infection (14), but it can also have no effect on survival (40). Furthermore, some host–pathogen interactions seem to provide more convincing evidence [e.g., *A. gambiae* and *Plasmodium* spp. (14, 45)] than others (40) for a role of *Dscam1* in immunity. *Dscam1* might not necessarily play an important role in all taxa, but instead be an “add-on” to immunity, where described phenomena are the side effects of e.g., altered hemocyte behavior. It is also worth bearing in mind that only a tiny fraction of the extremely speciose Pancrustacea have been examined to date. We do not know whether *Dscam1* publishing bias exists, in terms of an under-representation of “negative results,” but should unpublished data be sitting on someone's hard drive it could be helpful to share this information to unravel the

conditions under which *Dscam1* does or does not respond in an immunological context.

Question 2. *Dscam1* Was Hypothesized to Produce the Large Number of Variable Receptors Needed for Specificity in the Immune Response, in Some Ways Analogous to Antibodies in Vertebrates. Do You Think That *Dscam1* Is Indeed the Equivalent of Antibodies As Specific Immune Receptors?

Daniela Brites and Louis Du Pasquier

No. We think that there is a lot of confusion around the *Dscam1* analogy with antibodies. With respect to *Dscam1*'s somatically acquired diversity we think that the analogy with antibody diversity has been over-emphasized even though a warning had been formulated at the very beginning (46). One has had the tendency to compare apple and oranges. To be analogous to antibodies, *Dscam1* isoforms, specific to a pathogen epitope, should be secreted by some clones of uncommitted hemocytes that resulted from the stimulation of a precursor cell expressing the relevant *Dscam1* specificity. This is what happens in the adaptive immune system of vertebrates where specifically stimulated uncommitted B lymphocytes, the DNA of which has been somatically modified to encode a single receptor specificity per cell, proliferate (i.e., generating a clone) and differentiate into secreting plasma-cells that release large amounts of one antibody. Today, so far, nothing of the above applies to *Dscam1*. Since *Dscam1* variability is produced at the RNA level, it is not inheritable in the progenies of cells, would those cells where splicing occurred divide. But anyway there are so far no reported specifically induced proliferative responses of *Dscam1* producing cells. Unlike what has been proposed (47) there is no clonal amplification of the cells producing *Dscam1*. In addition adult flies do not produce new cells from the hematopoietic organs. However, one should be careful not to generalize from a single species or stage. In fact in the light of the interesting recent observations of transdetermination and proliferation of hemocyte lineages reported in *D. melanogaster* larvae, it might become interesting to follow *Dscam1* expression on those cells even though no information on their clonality is available (48). One speaks carefully of "demand adapted increase in hemocyte proliferation." Increases in cell numbers in parasitized flies have been reported as being due to proliferative response but an increase from 0 to 1,000 cells in 6 h cannot be due to a simple proliferation. There is here something new to investigate. There have been many examples of induction of *Dscam1* gene expression after some "antigenic" exposure (Membrane form? Soluble form? This is not always specified). In addition, significant increases in *Dscam1* diversity were observed in parasite-exposed mosquitoes (49). Increasing diversity after immunization does not make *Dscam1* a likely analog of antibodies. Indeed, following immunization one would rather see the amplification of one or two useful variants with some specificity like in antibody responses. Increasing diversity means lowering the concentration of single isoforms and therefore offering minimal chances for profiting from a special binding property. But we may simply not understand the mode of action of *Dscam1*. Since apparently a single cell expresses more than one *Dscam1* isotype (a profound difference compared to uncommitted lymphocytes) the best that

can be produced is a "shot gun" of unrelated *Dscam1* molecules (8). This leads us to an issue that has been often neglected: the concentration of each isoform either on the cell surface in the hemocyte population, or in the biological fluids. If one assumes that one variant of *Dscam1* has a better avidity for its ligand, how can this advantage be exploited? It is difficult to imagine the utility of a single variant diluted in the middle of thousands of other forms, so there is a need for selection and amplification steps. Those are difficult to conceive in a system where diversification is due to mutually exclusive alternative splicing and without specific cell proliferation. Therefore, *Dscam1* diversity might have a function other than being a repertoire of antigen reactive molecules.

Yuemei Dong

Lacking in vertebrate antibodies, insects rely on relatively small numbers of PRRs to combat various pathogens during their complex life cycles, which for a long time lead researchers to believe that the immune system in the invertebrates is not as sophisticated as its counterpart in the vertebrates (50–53). The genetic expansion of *Dscam1* and its ability to generate enormous pathogen specific receptors through immune responsive alternative splicing have equipped insects with a similar level of complexity at the molecular level, and thereby generate astounding analogs to antibodies. The rapid progress of *Dscam1* research in immunity has marked its role and importance in insect immunity, a groundbreaking contribution that blurs the classical strict clarification between innate and adaptive immunity (12, 14, 39, 53). Innate immunity used to be defined as being dependent on germ line encoded receptors, rather than recombination of somatically expressed antibodies, therefore *Dscam1*'s role in immunity fits the strict definition of innate immunity as it is germ line encoded, but not in the sense of the definition as *Dscam1* produces immune responsive splice forms. The vast diversity of the antibody system is clearly adaptive, hypothetically *Dscam1* also seems adaptive when considering it can produce tens of thousands potential splice variants.

Han-Ching Wang

The three hallmarks of acquired immunity are immune diversity, immune specificity, and immune memory (51). Mammalian antibody-based immune systems have all of these abilities; however, although there is evidence to suggest that *Dscam1* is also able to support all of these functions, *Dscam1* would have to provide these functionalities *via* different mechanisms than those used by antibodies. *Dscam1* is capable of immune diversity and immune specificity through alternative RNA splicing (2, 14, 33, 35, 39, 44, 45). However, after pathogen challenge, we still do not know whether immune cells in pancrustaceans are somehow able to actively design the particular alternative exons that show the ability to bind to the pathogen, or alternatively, whether populations of pathogen-induced specific *Dscam1* isoforms are created either through positive selection or by the same kind of negative selection mechanism that is used in vertebrate adaptive immune systems. Another curious similarity between *Dscam1* and antibodies is that whereas antibody diversity/specificity is achieved by combinations of three gene segments, V(D)J, *Dscam1* hyper-variability is achieved *via* three variable exon regions, Ig2/Ig3/Ig7. Furthermore, since most *Dscam1* studies have so far focused

primarily on particular variants rather than the whole Ig2/Ig3/Ig7 *Dscam1* combination, we might therefore be underestimating the potential immune specificity of *Dscam1*. Maintenance of the appropriate *Dscam1* populations is another problem, and immune memory in pancrustaceans is still an open question. In antibody-based mammalian immune systems, memory is achieved by somatic changes, but in pancrustaceans, to date, no convincing model of immune memory has yet been established.

Joachim Kurtz

My answer depends on what we mean when we say “equivalent.” The original view that *Dscam1* might function like antibodies [e.g., Ref. (38, 47, 50)] was probably a bit too optimistic, since several crucial elements could as of yet not be demonstrated and are maybe unlikely to exist: there seems to be no clonal amplification of the cells that produce the “right” isoforms, and maybe no receptor for *Dscam1* that could serve a similar role as the Fc receptor for a hypothetical opsonin-like function of *Dscam1*. Having said this, we should still be aware that being “equivalent” does not mean that everything has to be similar, and if we search for an equivalent system to produce somatically diversified receptors, then *Dscam1* is still “alive and kicking.” It is reasonable to assume that some form of somatic diversification is needed to produce a sufficiently large pathogen receptor repertoire that would be needed for specificity in discrimination among a large number of potential antigens. As of yet, we have only very limited evidence that pancrustaceans are actually able to achieve such a very high level of specificity in their pathogen and parasite defenses [e.g., Ref. (54)]. But if they are able, *Dscam1* is currently the only system known for pancrustaceans that could at least theoretically provide the needed receptor diversity. However, critical tests of the involvement of *Dscam1* in the specificity of immune reactions are still lacking.

Sophie A. O. Armitage

Through combinatorial diversification of vertebrate variable, diversity and joining gene segments (V(D)J) millions of combinations can be produced, and this number is in the order of billions of antibody molecules as a result of junctional diversification and somatic hypermutation (55). *Dscam1*, on the other hand, shows many orders of magnitude less diversity than vertebrate antibodies. Through ultra-deep sequencing of *D. melanogaster Dscam1* mRNA using next generation sequencing, Sun et al. (13) detected 18,496 of the possible 19,008 isoform combinations for exons 4, 6, and 9. However, *D. melanogaster* fat body and hemocytes do not express the full range of particularly the exon 9 cluster (11, 12), which could considerably reduce the total isoform estimation. Therefore, in addition to the above responses to this question, in terms of variation, *Dscam1* does not produce diversity that is equivalent to that produced by antibodies as specific immune receptors.

Question 3. Next to Specificity, Remembering Is a Key Aspect of Immune Memory. Could This Be Achieved with *Dscam1*?

Daniela Brites and Louis Du Pasquier

No, according to our conservative concept of memory! Memory, in an immunological sense, demands clonal amplification and

storage of specialized cells. It implies a reactivation of those cells after the initial antibody response has been down regulated (anamnestic response). This does not happen in any invertebrate and more specifically it does not happen in the *Dscam1* case. However, if some soluble form with specificity persists, the protection that it may confer can persist: it will be called memory by some but not by classical immunologists who see then a persisting on-going response or the long survival of a protecting agent and not the proper “recall” that characterizes memory responses.

Yuemei Dong

Evolution might have taken different routes to achieve functional similarities with *Dscam1* splice clouds in the invertebrates’ and antibodies in the vertebrates’ immunity. Given the two major features of adaptive immunity, immune specificity and memory, much of the work about *Dscam1*’s role in adaptive-resembling immunity was focused on addressing the pathogen recognition diversities and specificities. Quite some studies have shown that past infections influence insects’ humoral and cellular immune system thereby protecting the host from the second and the following pathogenic infections (53, 56–62). So-called immune priming or trained immunity has now been demonstrated in a wide range of pancrustacean species. However, with the currently available data, it still remains to be demonstrated that recognition specificities mediated by *Dscam1* splice variant repertoires have memory.

Han-Ching Wang

Although there is increasing evidence to suggest that immune memory occurs in pancrustaceans, the underlying molecular mechanism is still an open question. In pancrustaceans, it has long been clear that somatically generated immune factors, such as lectins and proPO-related proteins, could not account for immune specificity or immune memory. When *Dscam1* was discovered, it seemed to have great potential to support these special immune responses. Initially, however, most *Dscam1* studies were performed in short-lived pancrustaceans, which made it difficult to investigate its role in immune memory. In a recent, as of yet unpublished study, we challenged 200–300 long-lived crayfish twice with WSSV, with the second challenge made 14 days after the first. We then used gene cloning to determine the expressed combinations of Ig2–Ig3 in the *Dscam1* populations in collected hemocyte samples. In the crayfish that survived both challenges, some *Dscam1* isoforms with particular Ig2–Ig3 combinations showed a good binding affinity with WSSV. Furthermore, these isoforms appeared after the first challenge and they increased in quantity after the second challenge. This result is consistent with the idea that there might be meaningful selection and maintained expression of particular *Dscam1* exons. Unfortunately, there were also complications: first, the expression pattern was not seen in every surviving crayfish, and second, each surviving crayfish produced different *Dscam1* isoforms.

Sophie A. O. Armitage and Joachim Kurtz

We here consider a phenomenological definition of immune memory, which has been called “immune priming” in invertebrates and can be described as “the ability of an immune system to store or simply use the information on a previously

encountered antigen or parasite, upon secondary exposure” (31), rather than considering a mechanistic definition invoking the acquired immune system. Since a review (7), where we discussed the absence of empirical data on the hypothesis that *Dscam1* is involved in immune priming, there are at least two published studies on this topic (63, 64). The latter study found no change in *Dscam1* gene expression in a transgenerational immune priming study. However, Fu et al. (63) found that shrimp that had fed on bacterial spores harboring a WSSV protein, and then received siRNA to knockdown *Dscam1*, were less phagocytically active and had lower survival after a subsequent exposure to WSSV compared to shrimp that had also been primed with WSSV protein but did not receive *Dscam1* siRNA. This study would support the hypothesis that *Dscam1* has some involvement in immune priming, but we note that the expression of individual splice variants was not tested. If we imagine the hypothesis that *Dscam1* splice-variants are specific for a particular pathogen [this was not tested by Fu et al. (63)] it is difficult to conceive the mechanism by which *Dscam1* could “remember” aspects of previously encountered antigens. Variability in the *Dscam1* isoforms comes from somatically generated mRNA *via* mutually exclusive alternative splicing, therefore there would need to be some mechanism by which splicing patterns can be reproduced. Alternatively the variation in *Dscam1* may not be important for immune priming, it is just the presence or absence (reduction) of the protein that affects the phenomenon.

Question 4. What Alternatives Are There to an Antibody-Like Function of *Dscam1* in Pancrustacean Immune Systems?

Daniela Brites and Louis Du Pasquier

To sum up, *Dscam1* diversity as a whole seems to be the selected feature (diversity for diversity's sake) and we see it best exploited in the nervous system i.e., to specify cell identity. “Thus, the *Dscam1* repertoire of each cell is different from those of its neighbors, providing a potential mechanism for generating unique cell identity in the nervous system and elsewhere” (11). We therefore suggest that in a manner analogous to what it does in the nervous system, *Dscam1* on hemocytes might specify hemocyte identity, using homologous interactions in the way proposed for neurons (20) (see below paragraph 7 a suggestion for a method). Other functions could be inferred from understanding better the signaling capacities of the molecule.

Han-Ching Wang

In crustaceans, a pathogen can induce “antibody-like” *Dscam1* isoforms that show specific binding ability to the invading pathogen. In shrimp, but not yet in crayfish, we have also observed “super *Dscam1* isoforms” that have a wider binding ability to a range of bacteria and viruses (37). We have also seen that while a whole intact pathogen takes ~24 h to induce *Dscam1* expression (37, 43, 44), challenge with pathogen-associated molecular patterns (PAMPs), such as lipopolysaccharides (LPS) and beta-1-3-glucan and peptidoglycan (PG), induce *Dscam1* expression within just a few hours, after which *Dscam1* expression

levels then decline (41, 42). These findings suggest that, as with other innate, non-hypervariable, crustacean immune factors, *Dscam1* can also be triggered even without any antigen-specific recognition. Taking all of these results together, it is tempting to propose that it might be the “super *Dscam1* isoforms” that are responsible for this rapid, non-specific immune response (37). A corollary of this proposal is that *Dscam1* might therefore be regulated by at least two molecular mechanisms: one involved in the regulation of *Dscam1* expression; the other involved in the regulation of alternative splicing to generate specific *Dscam1* isoforms.

Joachim Kurtz

It is important to note that one of the “beauties” of the antibody system lies within the fact that antibodies are at the same time specific receptors and powerful effectors, such that the specificity of recognition is directly linked to the defensive function. However, this does not need to be the case for other immune molecules and provides the alternative that in the case of *Dscam1*, there might well be a function as a pathogen receptor (the studies demonstrating binding to pathogens support this view), but not as an effector (the mixed results regarding expression changes upon infection and the relatively low expression level of *Dscam1* in the immune system suggest this). The receptor function could be somewhat similar to the role of *Dscam1* in the nervous system. Hemocytes interact with one another when they encapsulate a pathogen or close a wound, while such interactions could in the absence of an insult be blocked by *Dscam1*, just as *Dscam1* homophilic binding blocks neuronal self-interactions. In this context it is intriguing that the parts of *Dscam1* that are responsible for homophilic interactions differ from the potentially pathogen-binding parts (24), so that both functions could co-occur. More generally, neuro-immunological feedbacks could be involved and link the neuronal function of *Dscam1* to its immune function. Such feedbacks are for example known for the regulation of immune genes by the internal clock (65) and it would be worth exploring an immune regulation role for *Dscam1*.

Sophie A. O. Armitage

Dscam1 has been proposed to act as a hypervariable PRR, a co-receptor during phagocytosis and an opsonin. As mentioned above, in addition to, or instead of, directly interacting with antigens, cell surface expressed *Dscam1* might be important for host cell–cell interactions, be these from hemocyte to hemocyte, or hemocyte to fat body/nervous system/other cell. If these interactions were in the form of homophilic binding, and if each cell has a restricted repertoire of isoforms, then the frequency of *Dscam1* homophilic binding between different cells would likely be low. Furthermore, if *Dscam1* interacts with pathogens, one could hypothesize that it also interacts with non-pathogenic microbiota found within the host. There are indications that *Dscam1* influences microbiota, more specifically bacteria, in *A. gambiae*: knockdown of *Dscam1* increased bacteria in the haemolymph (14) and overexpression of a particular *Dscam1* variant reduced bacteria in the gut (45). In contrast, in the small brown planthopper, *Laodelphax striatellus*, the titer of an extracellular symbiotic

bacterium was unaffected by *Dscam1* knockdown, and the titers of an endosymbiotic bacterium, *Wolbachia*, and the rice stripe virus were even decreased after knockdown (66). It is not clear why intracellular passengers would be affected by the knockdown of a cell adhesion molecule on the surface of the cell, is it a direct effect of the reduction in *Dscam1* or does knockdown negatively affect the host cells or their behavior in some way, so reducing survival for intracellular passengers? These are speculations, but it will be interesting to see whether other host–microbe interactions are influenced by *Dscam1*.

Question 5. What Is the Meaning of *Dscam1* Genetic Diversity?

Daniela Brites and Louis Du Pasquier

Comparative analysis of *Dscam1* in different arthropod groups has shown that two mechanisms of generating *Dscam1* diversity have evolved independently; massive whole-gene duplications in basal arthropods and the refined mutually exclusive alternative splicing of duplicated exons in the *Dscam1* of pancrustaceans. The ability to generate *Dscam1* diversity seems thus to have been positively selected in the evolutionary history of arthropods. Perhaps because diversity provided means of specifying cell identity (e.g., in hemocytes which are important mediators of embryonic development). We still know very little about *Dscam1* in basal arthropods, however, the evolution of pancrustacean *Dscam1* is well studied. We can conclude that in contrast to the constitutively expressed domains of *Dscam1* which are highly conserved, the alternative domains encoded by the alternative exons are highly diverse across pancrustaceans. If providing cell identity has been the most important driver of *Dscam1* diversity and that already happened in the most recent common ancestor of pancrustaceans, why would each group of pancrustaceans have evolved its own set of alternative exons? Could that be driven by an additional role in immunity? Then there is the question of *Dscam1*'s polymorphism within species, and what we can learn from it. In *Dscam1* polymorphism can be understood *sensu lato* both as the variants generated within an individual *via* alternative splicing of duplicated exons of *Dscam1*, and as polymorphism at the population level caused by mutations accumulated in orthologous exons in different individuals. The first source of polymorphism we have touched upon already and we would briefly like to mention what we have learnt from studying *Dscam1*'s allelic polymorphism. The regions of the variable domains that are not involved in the homophilic binding of the molecule (so-called epitope II) are more diverse (at the population level) than the regions involved in homophilic binding. Why are they more diverse? Could these variants be important for antigen recognition? Population genetic tests did not provide solid evidence supporting that these variants are maintained in the population because of antigen recognition, however the power of these analyses was low (67). The question of whether epitope II could be involved in binding to antigens therefore still remains open and should be tested experimentally.

Han-Ching Wang

Although *Dscam* is a ubiquitous protein that can be found in various animal species, such as mammals, fish, mollusks and

arthropods, I would like to discuss its genetic diversity solely in terms of arthropod *Dscams*. Curiously the ancestral hyper-variable *Dscam1* gene is only found in the pancrustaceans, while other arthropods have non-hypervariable *Dscam*-like genes (68). This situation presumably arose due to independent gene duplication and diversification events that in turn would be driven by their adaptive value in the evolution of the *Dscam1* gene family during Arthropoda evolution (68). It is very likely that this evolutionary pressure depended on the functional requirements of the arthropod's nervous system and/or its putative immune system, and in this content, it is important to note that the genetic diversity of *Dscam1* depends on both its extracellular region and its intracellular region. The hypervariable *Dscam1* extracellular region is used for axonal guidance during neuronal development and also provides a mechanism that might, at least potentially, be used for pathogen recognition (4). But the intracellular *Dscam1* cytoplasmic tails also show an interesting divergence: for instance, although there is a high homology between insect *Dscam1*s, the crustaceans have evolved quite differently, with variable cytoplasmic tails in shrimp (34), and a unique tail-less form of *Dscam1* in shrimp and crab (33–35). Furthermore, the secreted insect *Dscam1* is generated from membrane-bound *Dscam1* by a shedding process (47), whereas the more long-lived crustaceans express the tail-less *Dscam1* directly through alternative splicing (33–35). The way that this tail-less *Dscam1* is directly expressed bears a thought-provoking resemblance to the way that secreted IgM antibodies are expressed in mammals. While this might simply be a coincidence, it might also be a form of convergent evolution that reflects the importance that immune memory should have to a long-lived arthropod (i.e., a crustacean) as opposed to arthropods with shorter lifespans (e.g., most insects).

Joachim Kurtz

Generally, genetic diversity can come in different flavors: as diversity in the population (i.e., polymorphism) and as diversity within each individual. Accordingly, these different types of diversity could have different meanings: diversity in the population could have arisen from the processes of gene duplication and mutation and could be maintained by negative frequency-dependent selection, for example by parasites. Diversity in the individual might further be increased by somatic diversification processes, such as alternative splicing in the case of *Dscam1*. Its meaning could be diversity just for its own sake, which seems to be what is going on for *Dscam1* in the nervous system, so as to enable neuron self/non-self discrimination. Alternatively, its meaning could be to produce immune repertoire diversity so as to recognize diverse parasitic antigens. It is interesting to compare with other systems [for review see Ref. (31)], where immune diversity sometimes stems from massive diversification in the germ-line, such as in the case of V region-containing chitin binding proteins (VCBPs) in amphioxus, while it mainly comes from somatic diversification processes in other systems, such as the vertebrate antibodies and maybe the mollusks' fibrinogen-related proteins (FREPs) and the Sp185/333 proteins of sea urchins. For *Dscam1*, it is still difficult to say which of these potential “meanings” of genetic diversity is most relevant.

Sophie A. O. Armitage

Diversity in *Dscam1*, and more generally in the *Dscam* gene family, operates at different levels. Starting with a broader perspective, *Dscam1* paralogs have been described for insects (39, 68, 69) and at least one crustacean (27), showing that diversity exists at the level of whole gene duplications. For example in addition to *Dscam1*, the *D. melanogaster* genome contains *Dscam2*, *Dscam3*, and *Dscam4*, of which only *Dscam2* has (two) alternatively spliced exons in Ig7 (70). Narrowing our perspective to just the *Dscam1* gene, diversity is found across orthologs in terms the number of alternatively spliced exons that have evolved within each of the alternatively spliced exon cassettes found in a species. For example, from the lower diversity *Dscam1* in *Daphnia magna* [8, 24, and 17 alternatively spliced exons in Ig2, Ig3, and Ig7, respectively (39)] to higher diversity in *Anopheles gambiae* [14, 30, and 38, respectively (71)]. Reconstructing the evolutionary history of alternatively spliced exons across pancrustacean species with confidence proved difficult, probably because of the relatively short exons and long evolutionary timescale studies (68). It was possible to infer orthologs of most of the Ig2 and Ig7 variants between comparatively closely related species, i.e., *D. melanogaster* and *D. mojavensis*; but this was more difficult for Ig3, indicating more duplication or deletion events and resulting in a faster accumulation of diversity in this cluster of exons compared to Ig2 and Ig7 (68). In contrast, the amino acid sequences of the non-alternatively spliced regions in *Dscam1* orthologs show greater conservation (12). To zoom into the last level, we know relatively little about within-species diversity in terms of polymorphisms in the conserved or alternatively spliced regions of *Dscam1* [but see Ref. (67, 72)]. It has been hypothesized that because diversity within individuals is generated somatically, that one might not expect to find strong signatures of selection in the alternatively spliced exons (7, 10).

Question 6. What Was the Main Factor Driving the Evolution of Diversity in *Dscam1* – The Nervous or Immune System or Even Something Else?

Han-Ching Wang

It is interesting to note that, just like *Dscam*, a number of immune factors/receptors also play an important role in the neuronal system, and in fact there is increasing evidence that both systems share several mechanisms and have similar physical properties. Currently, however, it is still too early to say whether *Dscam1* diversity evolved dependently or independently of the nervous system because work in *Dscam1* neuroimmunology is still in its infancy in insects and has not even begun in long-lived crustaceans. Even so, based on current knowledge, I tend to believe that the diversity in *Dscam1* must on some level be driven by immune-related evolutionary pressure. First, at least in *D. melanogaster*, the ways that *Dscam1* alternative exons are used in neural cells and immune cells are different (12), suggesting that the regulation and exon selection of *Dscam1* alternative splicing may be mediated by different mechanisms. Second, in shrimp, our experimental data showed that recombinant *Dscam1* proteins containing various Ig2/Ig3 combinations bound more strongly to natural shrimp pathogens (such as *Vibrio harveyi* and WSSV) than to other bacteria (*Escherichia coli* and *Staphylococcus aureus*) (37). This

suggested that host–parasite coevolution may have occurred in a way that contributed to *Dscam1*'s hypervariability in immunity.

Joachim Kurtz

We can only speculate here, but when we consider that outside of the Pancrustacea, *Dscam*'s function seems to be only in the brain, then it is more likely that *Dscam*'s role in the nervous system pre-dates its function in the immune system. So let us assume there was an ancient function of *Dscam* in the brain, what could have driven the evolution of diversity? It is not unlikely that there was negative frequency-dependent selection, because a rare isoform has the advantage that it offers a higher value for the function to discriminate neurons. For a rare isoform, few other neurons will express the same isoform. However, a novel isoform also bears the risk of potentially harmful self-reactivity with any other pattern in the organism, leading to selection against self-reactive isoforms, i.e., self-reacting *Dscam* isoforms would be purged from the isoform “pool.” As a result, but still predating any immune function, we could imagine that with *Dscam* a molecular system has evolved that represents “non-self.” This could then have been a preadaptation (i.e., an evolutionary “exaptation”) for a system that would allow for non-self recognition also outside of the nervous system, i.e., a potential pathogen recognition system could have emerged. This way, an immune function might have followed from a more ancient nervous system function. Once there, selection pressures from the immune system would kick in and lead to further diversification. And finally, there is yet another possible initial driving factor for the evolution of diversity: to enable histocompatibility reactions within the species, i.e., allorecognition [see, e.g., Ref. (73, 74)], which for example explains the diversity at the *fuhc* locus in the ascidian *Botryllus*, where the *fester* gene also shows quite extensive alternative splicing [(75); for review see Ref. (76)]. Allorecognition systems have likely evolved in taxa where chimerism is a relevant problem, such as colonial invertebrates, where there is in particular the risk of germ-line parasitism. It would thus be interesting to find out whether or not chimerism might have played a role in those arthropods that initially diversified *Dscam1*.

Sophie A. O. Armitage

This question is difficult to answer with our current knowledge. Considerable data exists describing the function of *Dscam1* in the nervous system of *D. melanogaster* [reviewed in Ref. (5, 9)]. *Dscam1* mRNA is expressed in the brain of other species of Pancrustacea [e.g., *Daphnia* (39)], but our knowledge of how *Dscam1* functions in the nervous system of these species is less well understood. Studies focusing on the function of *Dscam1* in basal pancrustacean species might help to elucidate the selection pressure that maintains current diversity. Perhaps diversity in Ig7, which can also be found in non-*Dscam1* genes, initially evolved in response to different cues to those that resulted in diversity in Ig2 and Ig3? As detailed above, the *Dscam* gene family in arthropods is highly diverse; what were the selection pressures that lead to diversification not only of *Dscam1*, but also of the *Dscam* gene family in general? Was this the same selection pressure? We know that some of the non-highly diversified *Dscam* genes function in the nervous system [e.g., Ref. (69, 70, 77)], do these genes

also play immune roles? Do taxa that are evolutionarily basal to arthropods, e.g., Onychophora and Tardigrada, have *Dscam* homologs, if so are they diversified and what are the functions of these genes?

Question 7. What Are the Future Perspectives for Studies on *Dscam1* in Immunity? (Including What Essential Experiments or Approaches Are Missing That Would Help Our Understanding of *Dscam1* in Immunity?)

Daniela Brites and Louis Du Pasquier

(1) Make more reagents. Raise more monoclonal antibodies to follow and play with expression in different species. (2) Repertoire analysis. How does restriction of *Dscam1* isoform per single cell work? How stable is it? Similarly to what happens in neurons, does *Dscam1* splicing vary overtime in one immune cell [e.g., *Daphnia* (39)]? Use NGS for complete repertoire analysis over time after antigenic stimulation of all hemocytes including the especially interesting subsets recently discovered (78). (3) Cellular assays. Try plaque forming cell assays or ELISPOT assays to see whether there is real secretion vs. shedding by hemocytes or other cells. The *D. melanogaster* S2 cells that have been studied (48) produce perhaps a reduced repertoire of 15–50 different *Dscam1* molecular categories/cells but are far from being uncommitted. Proliferating in artificial conditions *in vitro*, S2 cells are not the equivalent of *in vivo* lymphocytes, but at least they are derived from the macrophage-like cell type of *D. melanogaster* and divide every 24 h at 26–28°C. This might still provide an *in vitro* model for understanding *Dscam1* signaling, stability of expression, and properties of progeny cells within the hematopoietic tissues. (4) Signaling. What are the consequences of ligand/receptor interactions? The signaling pathways that are known for *Dscam1* are still controversial (11). A possible relationship with the cytoskeleton has been suggested, which could be compatible with a role in phagocytosis and/or in cell movement. *Dscam1* mutants should help elucidating this aspect. How does *Dscam1* induction (upregulation) work in the fat body and hemocytes? Is it *via* direct stimulation by *Dscam1* receptors themselves, or *via* a cytokine? Or is it *via* Toll, JAK, or Imd pathways? Are coreceptors involved? (5) Exploit more *Dscam1* mutants in immunological experiments. Test the alternative hypothesis mentioned above (i.e., see 4) in *Dscam1* mutants. Migration of hemocytes can be monitored beautifully in *D. melanogaster* [(79), this article has since been withdrawn]. Following mechanical disturbance hemocytes change location and return to their original position within 45 min. If *Dscam1* plays a role in controlling migration of hemocytes, *Dscam1* mutants should show differences in relocation after disturbance, the prediction being that the cells would not return properly to their location. But perhaps the pattern of hemocytes distribution in mutants would be abnormal even without disturbance! (6) Specificity of binding. One should explore the binding properties of *Dscam1* proteins to heterologous ligands to confirm its potential as a receptor or an effector. In addition, study the precise binding properties of *Dscam1*, with proteins encoded for by different exon combinations, to determine the specificity of binding heterologous ligands (if any).

Go back to testing further the epitope I-epitope II hypothesis, with the *in vitro* production of *Dscam1* molecules, similarly to what was done by Watson et al. (80). (7) Comparative functional approaches. Study the role of *Dscam* in basal arthropods. Are the functions of the *Dscam1* molecules all analogous to each other? Are they redundant? Compare again hemocytes versus other cells and investigate the presence of soluble forms.

Yuemei Dong

Many questions remain to be answered, such as how many splice variants (or groups of variants, so called “*Dscam1* clouds”) are produced uniquely or whether there is a continuous range. Moreover, the essential questions regarding the stability of the pathogen specific *Dscam1* isoform repertoires after selection, and whether the expression of *Dscam1* clouds in the renewing population are regulated, remain to be addressed.

Han-Ching Wang

There are still many missing pieces in the puzzle of *Dscam1*-mediated immunity, even in terms of *Dscam1*'s general properties. For instance, we still do not have a complete picture of *Dscam1*'s response at the mRNA level and protein level after one or multiple stimulations with various immune stimulators. Part of the difficulty in *Dscam1* research is due to the fact that it cannot easily be silenced *in vivo* in long-lived crustaceans, such as shrimp and crayfish (unpublished data). Clearly, there is a need to develop an alternative *in vivo* system to test *Dscam1* function, especially for long-term observations. As for *Dscam1*'s immune diversity, the main questions to be addressed are which factors are involved in *Dscam1* alternative splicing and which mechanisms support the generation and maintenance of the specific *Dscam1* isoforms after pathogen challenge. Meanwhile, regarding *Dscam1*'s immune specificity, instead of just focusing on one particular highly expressed exon variant, we should investigate how the entire Ig2/Ig3/Ig7 combination is involved in specific binding with the corresponding pathogen. We would also like to know which kinds of epitopes on a particular pathogen can be recognized by the corresponding pathogen-induced *Dscam1* isoforms: does *Dscam1* bind with these pathogens through the recognition of general PAMPs or by recognizing particular antigens as pathogen surface proteins? Finally, the question of immune memory is perhaps the most difficult of all. Our current approach is to document the dynamics of the *Dscam1* isoform population in long-lived pancrustaceans after multiple stimulations. From this, we hope to establish whether or not some specific *Dscam1* isoforms are consistently present after specific pathogen stimulation. Other open questions include: which cell types might act as immune memory cells? Are the kinds of pathogen-specific *Dscam1* isoforms expressed after pathogen stimulation only produced by particular cells (or cell types)? At present, we are still a long way from answering these questions. Given that penaeid shrimp culture is a global economic activity that is vulnerable to economic losses from outbreaks of viral and bacterial diseases, the study of *Dscam1*-mediated immunity is also of practical importance. For example, a clear understanding of the mechanism of *Dscam1*-mediated immunity should provide a scientific basis for optimizing

shrimp vaccination strategies. We therefore believe that further research into *Dscam1* has great potential and should very much be encouraged.

Sophie A. O. Armitage and Joachim Kurtz

In addition to the abovementioned ideas we would add: (1) test whether epitope II indeed binds to pathogen/parasites, and if so, uncover what the specific binding partner is; are there conserved aspects of e.g., viruses, bacteria, fungi, or other parasites that are involved? (2) next, generation sequencing of mRNA alternative splicing patterns of exons 4, 6, and 9 [e.g., Ref. (81)], for example, using the *A. gambiae*—*Plasmodium* interaction or crustacean—WSSV interactions, which seem to be particularly promising to understand *Dscam1* in immunity; (3) as an extension to the previous point, applying peptide sequencing to *Dscam1* after infection with a pathogen or parasite to test the variability in alternatively spliced sequences at the protein level; also determine for how long the protein persists in the haemolymph (particularly in relation to knock-down studies); (4) test whether *Dscam1* is involved in specific immune memory by varying the identity of the primary and secondary pathogen/parasite in conjunction with *Dscam* knockdown before the primary and/or before the second

pathogen/parasite exposure; and (5) it could be interesting to further characterize the influence of *Dscam1* on microbiota.

AUTHOR CONTRIBUTIONS

JK and SA conceived the questions. SA collated the answers, wrote the introduction, and produced the figure. All authors responded to the questions and revised the manuscript.

ACKNOWLEDGMENTS

We would like to thank Larry Dishaw and Gary Litman for giving us the opportunity to contribute toward the Special Research Topic on Host and Microbe Adaptations in the Evolution of Immunity. We would like to thank Dietmar Schmucker for comments on an earlier draft of this manuscript, and the two referees for their comments.

FUNDING

H-CW was financially supported by the Ministry of Science and Technology (MOST 105-2633-B-006-004).

REFERENCES

- Yamakawa K, Huot YK, Haendelt MA, Hubert R, Chen XN, Lyons GE, et al. DSCAM: a novel member of the immunoglobulin superfamily maps in a Down syndrome region and is involved in the development of the nervous system. *Hum Mol Genet* (1998) 7:227–37. doi:10.1093/hmg/7.2.227
- Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, et al. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* (2000) 101:671–84. doi:10.1016/S0092-8674(00)80878-8
- Schmucker D. Molecular diversity of Dscam: recognition of molecular identity in neuronal wiring. *Nat Rev Neurosci* (2007) 8:915–20. doi:10.1038/nrn2256
- Schmucker D, Chen B. Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev* (2009) 23:147–56. doi:10.1101/gad.1752909
- Zipursky SL, Grueber WB. The molecular basis of self-avoidance. *Annu Rev Neurosci* (2013) 36:547–68. doi:10.1146/annurev-neuro-062111-150414
- Ng TH, Chiang YA, Yeh YC, Wang HC. Review of Dscam-mediated immunity in shrimp and other arthropods. *Dev Comp Immunol* (2014) 46:129–38. doi:10.1016/j.dci.2014.04.002
- Armitage SAO, Peuß R, Kurtz J. Dscam and pancrustacean immune memory – a review of the evidence. *Dev Comp Immunol* (2015) 48:315–23. doi:10.1016/j.dci.2014.03.004
- Brites D, Du Pasquier L. Somatic and germline diversification of a putative immunoreceptor within one phylum: Dscam in arthropods. In: Hsu E, Du Pasquier L, editors. *Results and Problems in Cell Differentiation*. Berlin: Springer International Publishing (2015). p. 131–58.
- Petrovic M, Schmucker D. Axonal wiring in neural development: target-independent mechanisms help to establish precision and complexity. *Bioessays* (2015) 37:996–1004. doi:10.1002/bies.201400222
- Armitage SAO, Brites D. The immune-related roles and the evolutionary history of Dscam in arthropods. In: Malagoli D, editor. *The Evolution of the Immune System*. London, San Diego, Cambridge, Oxford: Elsevier (2016). p. 241–74.
- Neves G, Zucker J, Daly M, Chess A. Stochastic yet biased expression of multiple Dscam splice variants by individual cells. *Nat Genet* (2004) 36:240–6. doi:10.1038/ng1299
- Watson FL, Püttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* (2005) 309:1874–8. doi:10.1126/science.1116887
- Sun W, You X, Gogol-Döring A, He H, Kise Y, Sohn M, et al. Ultra-deep profiling of alternatively spliced *Drosophila* Dscam isoforms by circularization-assisted multi-segment sequencing. *EMBO J* (2013) 32:2029–38. doi:10.1038/emboj.2013.144
- Dong Y, Taylor HE, Dimopoulos G. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol* (2006) 4:e229. doi:10.1371/journal.pbio.0040229
- Yu H, Yang J, Wang J, Huang Y, Lee T. Endodomain diversity in the *Drosophila* Dscam and its roles in neuronal morphogenesis. *J Neurosci* (2009) 29:1904–14. doi:10.1523/JNEUROSCI.5743-08.2009
- Hattori D, Millard S, Wojtowicz W, Zipursky S. Dscam-mediated cell recognition regulates neural circuit formation. *Annu Rev Cell Dev Biol* (2008) 24:597–620. doi:10.1146/annurev.cellbio.24.110707.175250
- Legg D, Sutton M, Edgecombe G. Arthropod fossil data increase congruence of morphological and molecular phylogenies. *Nat Commun* (2013) 4:2485. doi:10.1038/ncomms3485
- Zipursky SL, Wojtowicz WM, Hattori D. Got diversity? Wiring the fly brain with Dscam. *Trends Biochem Sci* (2006) 31:581–8. doi:10.1016/j.tibs.2006.08.003
- Millard S, Zipursky S. Dscam-mediated repulsion controls tiling and self-avoidance. *Curr Opin Neurobiol* (2008) 18:84–9. doi:10.1016/j.conb.2008.05.005
- Wojtowicz WM, Flanagan JJ, Millard SS, Zipursky SL, Clemens JC. Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* (2004) 118:619–33. doi:10.1016/j.cell.2004.08.021
- Wojtowicz WM, Wu W, Andre I, Qian B, Baker D, Zipursky SL. A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell* (2007) 130:1134–45. doi:10.1016/j.cell.2007.08.026
- Matthews BJ, Kim ME, Flanagan JJ, Hattori D, Clemens JC, Zipursky SL, et al. Dendrite self-avoidance is controlled by Dscam. *Cell* (2007) 129:593–604. doi:10.1016/j.cell.2007.04.013
- Hattori D, Demir E, Kim HW, Viragh E, Zipursky SL, Dickson BJ. Dscam diversity is essential for neuronal wiring and self-recognition. *Nature* (2007) 449:223–7. doi:10.1038/nature06099
- Meijers R, Puettmann-Holgado R, Skiniotis G, Liu JH, Walz T, Wang JH, et al. Structural basis of Dscam isoform specificity. *Nature* (2007) 449:487–91. doi:10.1038/nature06147
- Sawaya MR, Wojtowicz WM, Andre I, Qian B, Wu W, Baker D, et al. A double S shape provides the structural basis for the extraordinary binding specificity of Dscam isoforms. *Cell* (2008) 134:1007–18. doi:10.1016/j.cell.2008.07.042

26. Brites D, Brena C, Ebert D, Du Pasquier L. More than one way to produce protein diversity: duplication and limited alternative splicing of an adhesion molecule gene in basal arthropods. *Evolution* (2013) 67:2999–3011. doi:10.1111/evo.12179
27. Palmer W, Jiggins F. Comparative genomics reveals the origins and diversity of arthropod immune systems. *Mol Biol Evol* (2015) 32:2111–29. doi:10.1093/molbev/msv093
28. Yue Y, Meng Y, Ma H, Hou S, Cao G, Hong W, et al. A large family of Dscam genes with tandemly arrayed 5' cassettes in Chelicerata. *Nat Commun* (2016) 7:11252. doi:10.1038/ncomms11252
29. Janeway C, Travers P, Walport M, Shlomchik M. *Immunobiology*. 6th ed. New York: Garland Science (2005).
30. Hoffmann J, Kafatos F, Janeway C, Ezekowitz R. Phylogenetic perspectives in innate immunity. *Science* (1999) 284:1313–8. doi:10.1126/science.284.5418.1313
31. Milutinović B, Kurtz J. Immune memory in invertebrates. *Semin Immunol* (2016) 28:328–42. doi:10.1016/j.smim.2016.05.004
32. Werner T, Borge-Renberg K, Mellroth P, Steiner H, Hultmark D. Functional diversity of the *Drosophila* PGRP-LC gene cluster in the response to lipopolysaccharide and peptidoglycan. *J Biol Chem* (2003) 278:26319–22. doi:10.1074/jbc.C300184200
33. Chou PH, Chang HS, Chen IT, Lin HY, Chen YM, Yang HL, et al. The putative invertebrate adaptive immune protein *Litopenaeus vannamei* Dscam (LvDscam) is the first reported Dscam to lack a transmembrane domain and cytoplasmic tail. *Dev Comp Immunol* (2009) 33:1258–67. doi:10.1016/j.dci.2009.07.006
34. Chou PH, Chang HS, Chen IT, Lee CW, Hung HY, Wang HC. *Penaeus monodon* Dscam (PmDscam) has a highly diverse cytoplasmic tail and is the first membrane-bound shrimp Dscam to be reported. *Fish Shellfish Immunol* (2011) 30:1109–23. doi:10.1016/j.fsi.2011.02.009
35. Wang J, Wang L, Gao Y, Jiang Q, Yi Q, Zhang H, et al. A tailless Dscam from *Eriocheir sinensis* diversified by alternative splicing. *Fish Shellfish Immunol* (2013) 35:249–61. doi:10.1016/j.fsi.2013.04.029
36. Watthanasurorot A, Jiravanichpaisal P, Liu H, Söderhäll I, Söderhäll K. Bacteria-induced Dscam isoforms of the crustacean, *Pacifastacus leniusculus*. *PLoS Pathog* (2011) 7(6):e1002062. doi:10.1371/journal.ppat.1002062
37. Hung HY, Ng TH, Lin JH, Chiang YA, Chuang YC, Wang HC. Properties of *Litopenaeus vannamei* Dscam (LvDscam) isoforms related to specific pathogen recognition. *Fish Shellfish Immunol* (2013) 35:1272–81. doi:10.1016/j.fsi.2013.07.045
38. Stuart LM, Ezekowitz RA. Phagocytosis and comparative innate immunity: learning on the fly. *Nat Rev Immunol* (2008) 8:131–41. doi:10.1038/nri2240
39. Brites D, McTaggart S, Morris K, Anderson J, Thomas K, Colson I, et al. The Dscam homologue of the crustacean *Daphnia* is diversified by alternative splicing like in insects. *Mol Biol Evol* (2008) 25:1429–39. doi:10.1093/molbev/msn087
40. Peuß R, Wensing KU, Woestmann L, Eggert H, Milutinović B, Sroka MGU, et al. *Down syndrome cell adhesion molecule 1*: testing for a role in insect immunity, behaviour and reproduction. *R Soc Open Sci* (2016) 3:160138. doi:10.1098/rsos.160138
41. Jin X-K, Li W-W, Wu M-H, Guo X-N, Li S, Yu A-Q, et al. Immunoglobulin superfamily protein Dscam exhibited molecular diversity by alternative splicing in hemocytes of crustacean, *Eriocheir sinensis*. *Fish Shellfish Immunol* (2013) 35:900–9. doi:10.1016/j.fsi.2013.06.029
42. Li D, Yu A, Li X, Zhu Y, Jin X, Li W, et al. Antimicrobial activity of a novel hypervariable immunoglobulin domain-containing receptor Dscam in *Cherax quadricarinatus*. *Fish Shellfish Immunol* (2015) 47:766–76. doi:10.1016/j.fsi.2015.10.025
43. Chiang YA, Hung HY, Lee CW, Huang YT, Wang HC. Shrimp Dscam and its cytoplasmic tail splicing activator serine/arginine (SR)-rich protein B52 were both induced after white spot syndrome virus challenge. *Fish Shellfish Immunol* (2013) 34:209–19. doi:10.1016/j.fsi.2012.10.021
44. Ng TH, Hung HY, Chiang YA, Lin JH, Chen YN, Chuang YC, et al. WSSV-induced crayfish Dscam shows durable immune behavior. *Fish Shellfish Immunol* (2014) 40:78–90. doi:10.1016/j.fsi.2014.06.023
45. Dong Y, Cirimotich CM, Pike A, Chandra R, Dimopoulos G. *Anopheles* NF- κ B-regulated splicing factors direct pathogen-specific repertoires of the hypervariable pattern recognition receptor AgDscam. *Cell Host Microbe* (2012) 12(4):521–30. doi:10.1016/j.chom.2012.09.004
46. Du Pasquier L. Insects diversify one molecule to serve two systems. *Science* (2005) 309(5742):1826–7. doi:10.1126/science.1118828
47. Boehm T. Two in one: dual function of an invertebrate antigen receptor. *Nat Immunol* (2007) 8:1031–3. doi:10.1038/nri1007-1031
48. Anderl I, Vesala L, Ihalaenen TO, Vanha-Aho LM, Andó I, Rämetsä M, et al. Transdifferentiation and proliferation in two distinct hemocyte lineages in *Drosophila melanogaster* larvae after wasp infection. *PLoS Pathog* (2016) 12:e1005746. doi:10.1371/journal.ppat.1005746
49. Smith P, Mwangi J, Afrane Y, Yan G, Obbard D, Ranford-Cartwright L, et al. Alternative splicing of the *Anopheles gambiae* Dscam gene in diverse *Plasmodium falciparum* infections. *Malar J* (2011) 10:156. doi:10.1186/1475-2875-10-156
50. Kurtz J, Armitage SAO. Alternative adaptive immunity in invertebrates. *Trends Immunol* (2006) 27:493–6. doi:10.1016/j.it.2006.09.001
51. Boehm T. Evolution of vertebrate immunity. *Curr Biol* (2012) 22:R722–32. doi:10.1016/j.cub.2012.07.003
52. Sun JC, Ugolini S, Vivier E. Immunological memory within the innate immune system. *EMBO J* (2014) 33:1295–303. doi:10.1002/emboj.201387651
53. Milutinović B, Peuss R, Ferro K, Kurtz J. Immune priming in arthropods: an update focusing on the red flour beetle. *Zoology* (2016) 119:254–61. doi:10.1016/j.zool.2016.03.006
54. Kurtz J, Franz K. Evidence for memory in invertebrate immunity. *Nature* (2003) 425:37–8. doi:10.1038/425037a
55. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science (2002).
56. Schmid-Hempel P. Natural insect host-parasite systems show immune priming and specificity: puzzles to be solved. *Bioessays* (2005) 27:1026–34. doi:10.1002/bies.20282
57. Pham LN, Dionne MS, Shirasu-Hiza M, Schneider DS. A specific primed immune response in *Drosophila* is dependent on phagocytes. *PLoS Pathog* (2007) 3:e26. doi:10.1371/journal.ppat.0030026
58. Rodrigues J, Brayner FA, Alves LC, Dixit R, Barillas-Mury C. Hemocyte differentiation mediates innate immune memory in *Anopheles gambiae* mosquitoes. *Science* (2010) 329:1353–5. doi:10.1126/science.1190689
59. Chambers MC, Schneider DS. Pioneering immunology: insect style. *Curr Opin Immunol* (2011) 24:10–14. doi:10.1016/j.coi.2011.11.003
60. Eleftherianos I, Schneider D. *Drosophila* immunity research on the move. *Fly* (2011) 5:247–54. doi:10.4161/fly.5.3.17028
61. Futo M, Armitage SAO, Kurtz J. Microbiota plays a role in oral immune priming in *Tribolium castaneum*. *Front Microbiol* (2016) 6:1383. doi:10.3389/fmicb.2015.01383
62. Vargas V, Moreno-Garcia M, Duarte-Elguea E, Lanz-Mendoza H. Limited specificity in the injury and infection priming against bacteria in *Aedes aegypti* mosquitoes. *Front Microbiol* (2016) 7:975. doi:10.3389/fmicb.2016.00975
63. Fu L, Li T, Wang Y. Potential role of LvDscam in specific immune response of *Litopenaeus vannamei* against white spot syndrome virus by oral delivery of VP28 using *Bacillus subtilis*. *Aquac Res* (2016) 47:2068–79. doi:10.1111/are.12661
64. Norouzitalab P, Baruah K, Biswas P, Vanrompay D, Bossier P. Probing the phenomenon of trained immunity in invertebrates during a transgenerational study, using brine shrimp *Artemia* as a model system. *Sci Rep* (2016) 6:21166. doi:10.1038/srep21166
65. Li J, Terry E, Fejer E, Gamba D, Hartmann N, Logsdon J, et al. Achilles is a circadian clock-controlled gene that regulates immune function in *Drosophila*. *Brain Behav Immun* (2017) 61:127–36. doi:10.1016/j.bbi.2016.11.012
66. Zhang F, Li Q, Chen X, Huo Y, Guo H, Song Z, et al. Roles of the *Laodelphax striatellus* Down syndrome cell adhesion molecule in rice stripe virus infection of its insect vector. *Insect Mol Biol* (2016) 25:413–21. doi:10.1111/imb.12226
67. Brites D, Encinas-Viso F, Ebert D, Du Pasquier L, Haag CR. Population genetics of duplicated alternatively spliced exons of the Dscam gene in *Daphnia* and *Drosophila*. *PLoS One* (2011) 6:e27947. doi:10.1371/journal.pone.0027947
68. Armitage SAO, Freiburg RY, Kurtz J, Bravo IG. The evolution of *Dscam* genes across the arthropods. *BMC Evol Biol* (2012) 12:53. doi:10.1186/1471-2148-12-53
69. Funada M, Hara H, Sasagawa H, Kitagawa Y, Kadowaki T. A honey bee Dscam family member, AbsCAM, is a brain-specific cell adhesion molecule

- with the neurite outgrowth activity which influences neuronal wiring during development. *Eur J Neurosci* (2007) 25:168–80. doi:10.1111/j.1460-9568.2006.05270.x
70. Millard SS, Flanagan JJ, Pappu KS, Wu W, Zipursky SL. Dscam2 mediates axonal tiling in the *Drosophila* visual system. *Nature* (2007) 447:720–4. doi:10.1038/nature05855
 71. Graveley BR, Kaur A, Gunning D, Zipursky SL, Rowen L, Clemens JC. The organization and evolution of the Dipteran and Hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *RNA* (2004) 10:1499–506. doi:10.1261/rna.7105504
 72. Obbard DJ, Welch JJ, Kim KW, Jiggins FM. Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet* (2009) 5:e1000698. doi:10.1371/journal.pgen.1000698
 73. Scofield V, Schlumpberger J, West L, Weissman I. Protochordate allorecognition is controlled by a MHC-like system. *Nature* (1982) 295:499–502. doi:10.1038/295499a0
 74. Ballarin L, Du Pasquier L, Rinkevich B, Kurtz J. Evolutionary aspects of allorecognition. *Invertebrate Surviv J* (2015) 12:233–6.
 75. Nyholm SV, Passegue E, Ludington WB, Voskoboinik A, Mitchel K, Weissman IL, et al. *fester*, a candidate allorecognition receptor from a primitive chordate. *Immunity* (2006) 25:163–73. doi:10.1016/j.immuni.2006.04.011
 76. Taketa D, De Tomaso A. *Botryllus schlosseri* allorecognition: tackling the enigma. *Dev Comp Immunol* (2015) 48:254–65. doi:10.1016/j.dci.2014.03.014
 77. Lah G, Li J, Millard S. Cell-specific alternative splicing of *Drosophila Dscam2* is crucial for proper neuronal wiring. *Neuron* (2014) 83:1376–88. doi:10.1016/j.neuron.2014.08.002
 78. Miura S, Martins A, Zhang K, Graveley B, Zipursky S. Probabilistic splicing of Dscam1 establishes identity at the level of single neurons. *Cell* (2013) 155:1166–77. doi:10.1016/j.cell.2013.10.018
 79. Li W, Guan KL. The Down syndrome cell adhesion molecule (DSCAM) interacts with and activates Pak. *J Biol Chem* (2004) 279:32824–32824. doi:10.1074/jbc.M401878200
 80. Petraki S, Alexander B, Bruckner K. Assaying blood cell populations of the *Drosophila melanogaster* larva. *J Vis Exp* (2015) 105:52733. doi:10.3791/52733
 81. Armitage SAO, Sun W, You X, Kurtz J, Schmucker D, Chen W. Quantitative profiling of *Drosophila melanogaster Dscam1* isoforms reveals no changes in splicing after bacterial exposure. *PLoS One* (2014) 9:e108660. doi:10.1371/journal.pone.0108660

Conflict of Interest Statement: 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.

Copyright © 2017 Armitage, Kurtz, Brites, Dong, Du Pasquier and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Leaky Gut As a Danger Signal for Autoimmune Diseases

Qinghui Mu¹, Jay Kirby¹, Christopher M. Reilly² and Xin M. Luo^{1*}

¹ Department of Biomedical Sciences and Pathobiology, Virginia-Maryland College of Veterinary Medicine, Virginia Tech, Blacksburg, VA, USA, ² Edward Via College of Osteopathic Medicine, Blacksburg, VA, USA

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, USA

Reviewed by:

Rajendra Karki,
St. Jude Children's Research
Hospital, USA
Lisa Rizzetto,
Fondazione Edmund Mach, Italy

*Correspondence:

Xin M. Luo
xinluo@vt.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 27 March 2017

Accepted: 05 May 2017

Published: 23 May 2017

Citation:

Mu Q, Kirby J, Reilly CM and Luo XM
(2017) Leaky Gut As a Danger Signal
for Autoimmune Diseases.
Front. Immunol. 8:598.
doi: 10.3389/fimmu.2017.00598

The intestinal epithelial lining, together with factors secreted from it, forms a barrier that separates the host from the environment. In pathologic conditions, the permeability of the epithelial lining may be compromised allowing the passage of toxins, antigens, and bacteria in the lumen to enter the blood stream creating a “leaky gut.” In individuals with a genetic predisposition, a leaky gut may allow environmental factors to enter the body and trigger the initiation and development of autoimmune disease. Growing evidence shows that the gut microbiota is important in supporting the epithelial barrier and therefore plays a key role in the regulation of environmental factors that enter the body. Several recent reports have shown that probiotics can reverse the leaky gut by enhancing the production of tight junction proteins; however, additional and longer term studies are still required. Conversely, pathogenic bacteria that can facilitate a leaky gut and induce autoimmune symptoms can be ameliorated with the use of antibiotic treatment. Therefore, it is hypothesized that modulating the gut microbiota can serve as a potential method for regulating intestinal permeability and may help to alter the course of autoimmune diseases in susceptible individuals.

Keywords: leaky gut, microbial translocation, gut microbiota, probiotics, autoimmunity

INTRODUCTION

For digestion and absorption purposes, mammals have developed a very complicated and highly specialized gastrointestinal system maintained by the mucosal barrier (1). However, apart from absorbable nutrients, the intestinal mucosa also faces tremendous exterior antigens, including food antigens, commensal bacteria, pathogens, and toxins. Thus, a specialized barrier function is required to block the entry of diverse exterior antigens while absorbing nutrients. Impressively, in the intestine, the front line of this barrier is maintained by only a single layer of specialized epithelial cells that are linked together by tight junction (TJ) proteins. Many other factors aid in support of this barrier including mucins, antimicrobial molecules, immunoglobulins, and cytokines. If any abnormalities occur among these factors, the intestinal permeability may increase, which is termed a “leaky gut.” A leaky gut allows the entry of exterior antigens from the gut lumen into the host, which may promote both local and systemic immune responses. Multiple diseases may arise or be exacerbated due to a leaky gut, including autoimmune diseases such as inflammatory bowel disease, celiac disease, autoimmune hepatitis, type 1 diabetes (T1D), multiple sclerosis, and systemic lupus erythematosus (SLE) (2–6). Numerous factors can affect gut permeability, such as various diet-derived compounds, alcohol consumption, and gut microbiota dysbiosis. While this review is focused on chronic inflammation and gut barrier functions in mammals, it is worth noting that leaky gut is a phenomenon that is widespread in both mammalian and non-mammalian animals (7). Thus, studies in systems

outside of mammals, such as zebrafish (7, 8), can be also helpful in our understanding of the relationship between inflammation and the intestinal barrier.

The gut microbiota has drawn intense attention in the past decade (9). Although scientists have studied gut microbiota for many years, recent advancements in molecular biology including next-generation sequencing technology has enabled researchers to gain new insight in this research field. While we are still far away from clearly understanding the exact roles and effecting modes of gut microbiota, growing evidence suggests that gut microbiota is important in modulating gut permeability and intestinal barrier functions. In this review, we summarize recent advances in the understanding of the leaky gut, bacterial translocation, and gut microbiota dysbiosis, with a particular focus on their association with extraintestinal autoimmune diseases, such as T1D and SLE.

THE INTESTINAL BARRIER

A large variety of exogenous substances colonize the gut lumen, such as microorganisms, toxins, and antigens. Without an intact and properly functioning intestinal barrier, these substances can penetrate the tissues beneath the intestinal epithelial lining, diffuse into blood and lymphatic circulations, and disrupt tissue homeostasis. However, there is an efficient multifaceted intestinal barrier system with physical, biochemical, and immunological components that prevents the entry of most pathogens (**Figure 1**). These components coordinate with each other to prevent uncontrolled translocation of luminal contents into the body. Below is a brief synopsis of the main components comprising the intestinal barrier.

Physical Barrier

In humans, the intestinal epithelium covers as large as 400 m² of surface area (1). Though only a single layer of cells, the intestinal epithelial cells (IECs) are the mainstay of the intestinal barrier and serve as a physical barrier (**Figure 1**). There are at least seven types of functional IECs—enterocytes, goblet cells, Paneth cells, microfold cells (M cells), enteroendocrine cells, cup cells, and tuft cells, although the functions of the last two cell populations are not well understood (10). Among all these cell types, enterocytes represent the absolute majority, accounting for at least 90% of crypt cells or villus cells. Enterocytes are absorptive cells and vital for nutrient uptake. However, growing evidence indicates that the functions of enterocytes are not limited to nutrient absorption. For example, enterocytes can control the abundance of Gram-positive bacteria by expressing RegIII γ , one type of antimicrobial proteins (AMPs) (11–13). All epithelial cell types originate from Lgr5⁺ intestinal epithelial stem cells, which reside within the crypts (14). The turnover rate of IECs is high and the cells are renewed every 3–5 days in the mammalian intestine (10, 15), with the exception being the Paneth cells, which have a life span of about 2 months.

The IEC lining is continuous, and the contact between IECs is sealed by TJs (16). The paracellular pathway, in contrast to transcellular pathway, allows the transport of substances across the gut epithelium through the spaces between IECs. A large variety of molecules, mainly proteins, control the plasticity of TJs. More

than 40 TJ proteins have been recognized, including occludin, claudins, junctional adhesion molecule A, and tricellulin (17). Under various pathological conditions, paracellular permeability may be increased, resulting in the entry of unwelcome, potentially harmful molecules.

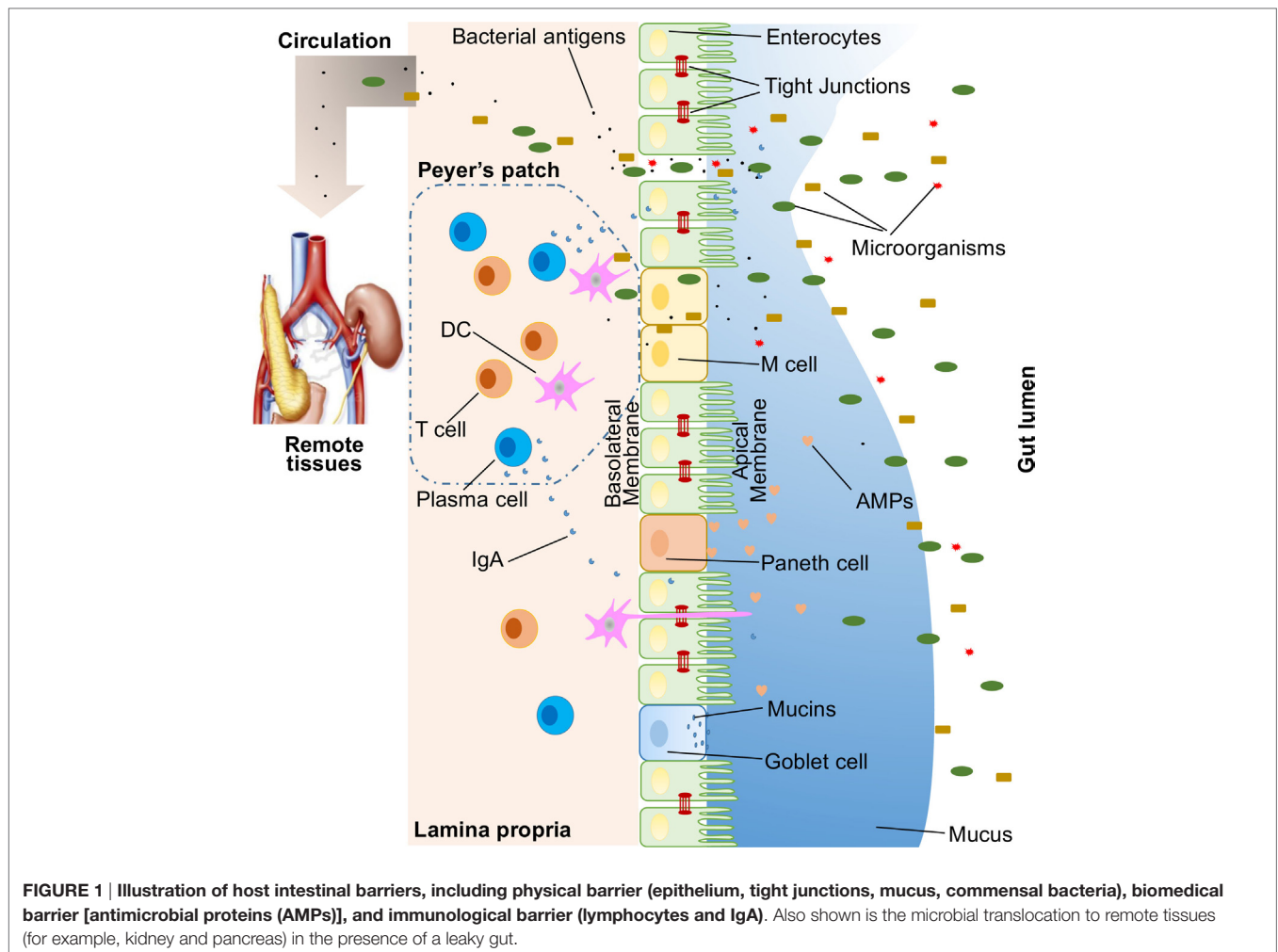
On top of the gut epithelium, there are two layers of mucus, the inner and outer layers, that cover the whole intestinal epithelial lining and provide physical protection to separate luminal microorganisms from the epithelium. Organized by its major component, a highly glycosylated gel-forming mucin MUC2, the mucus contains diverse molecules including IgA as well as enzymes and proteins, such as lactoferrin (18). Goblet cells are the central cell type for the formation of mucus. They not only produce MUC2 mucin but also secrete other mucus components such as ZG16, AGR2, FCGBP, CLCA1, and TFF3 (19, 20). Colitis would spontaneously develop in Muc2-deficient mice, indicating a critical role for MUC2 in mucosal protection (21). In addition to gel-forming mucins, there is another type of mucin that is in close proximity to epithelial cells, called transmembrane mucins. Enterocytes are the main producers of transmembrane mucins (20).

The gut commensal bacteria have been described as one component of the intestinal physical barrier primarily due to its two major functions (22). The first is to promote resistance to the colonization of harmful or pathogenic bacteria species by competing for nutrients, occupying attachment sites, and releasing antimicrobial substances (23, 24). Additionally, the gut microbiota regulates the digestion and absorption of nutrients to supply energy to epithelial cells, which are a major component of the physical barrier (25). A good example of the direct energy supply is the production of short-chain fatty acids by the gut microbiota, which are used by colonocytes for their development and metabolism (26). Taken together, IECs, the mucus layers, and gut microbial residents serve as the physical barrier to limit the entry of unfriendly luminal contents into host tissues.

Biochemical Barrier

Biochemical molecules with antimicrobial properties exist in the mucus as well as far into the lumen and include bile acids and AMPs (27, 28) (**Figure 1**). These diverse molecules form a complicated network to reduce the load of colonized bacteria and decrease the chance of contact between luminal antigens and host cells. They are a good supplement to the physical barrier and an essential component of the intestinal barrier function.

The proximal small intestine harbors very few microorganisms (29). But as the distance from the stomach increases, the pH rises and the number of colonized bacteria escalates (30). Facing a large number of microorganisms, which likely outnumber the number of host cells, multiple AMPs are generated to fight against invaders. These AMPs are divided into several types, including α - and β -defensins, C-type lectin, cathelicidin, lysozyme, and intestinal alkaline phosphatase (IAP) (27). Their detailed antimicrobial mechanisms are discussed elsewhere (31). As a major, but not exclusive, producer of AMPs, Paneth cells support and mediate the biochemical barrier function.



Immunological Barrier

Below the intestinal epithelium, there are organized lymphoid follicles, including the Peyer's patches and isolated lymphoid follicles. Inside the follicles, a variety of immune cells, including B cells, T cells, dendritic cells (DCs), and neutrophils, orchestrate the immune response by presenting antigens, secreting cytokines, and producing antigen-binding antibodies (**Figure 1**). In the intestinal epithelium where lymphoid follicles are found, M cells are present that transcytose antigens across the intestinal epithelium to the Peyer's patches underneath (14). In addition, goblet cells present acquired luminal antigens to CD103⁺ DCs in lamina propria in small intestine by forming goblet cell-associated antigen passages (GAPs) (32, 33). Interestingly, spontaneous antigen presentation was also observed in the colon, but only when the mice were raised germ-free (GF), or housed conventionally but with oral antibiotic treatment (34). This suggests that the antigen uptake process and formation of GAPs are regulated by the colonic microbiota (35). In addition, goblet cells and GAPs are capable of sensing invasive pathogens and inhibiting the translocation of pathogenic bacteria into the host immune system (36). Furthermore, intestinal mononuclear phagocytes can sense and

sample luminal contents (37, 38). CX3CR1-expressing cells are responsible for this process, and antigen sampling is dependent on structures called transepithelial dendrites (TEDs) (39, 40). The formation of TEDs is regulated by CX3CR1⁺ macrophages and the expression of CX3CL1 by certain IECs (41, 42).

Another component of the immunological barrier is secretory IgA (SIgA). As the most abundant immunoglobulin in the body, IgA resides primarily on intestinal mucosal surfaces. While some people with selective IgA deficiency appear to be healthy, SIgA is important as it presumably interacts with commensal bacteria to provide protection against pathogens. A unique feature about SIgA is that it is structurally resilient in protease-rich environments allowing it to remain functionally active compared to other antibody isotypes on mucosal surfaces (43). In adult humans, about 50 mg/kg of SIgA is produced daily by plasma cells residing in the intestinal lamina propria. Finally, SIgA can be transcytosed through the epithelium and secreted into the gut lumen.

Though not mentioned here, self-modulating factors, such as nerves and diverse cytokines, are also important for maintaining the normal functions of the intestinal barrier.

GUT MICROBIOTA AND THE INTESTINAL BARRIER

Microbiota can be sensed by the host through pattern recognition receptors (PRRs), such as toll-like receptors (TLRs) and nucleotide-binding oligomerization domain (NOD)-like receptors (NLRs). In the gut, the bacteria–host communications are largely dependent on the recognition of microbe-associated molecular patterns by PRRs expressed on immune and non-immune cells. Certain microbiota, bacterial products, and metabolites affect the intestinal barrier function and are responsible for the subsequent breakdown of tissue homeostasis. When there is a leaky gut, commensal bacteria in gut lumen, together with their products, are able to escape the lumen of the gut, which may induce inflammation and cause systemic tissue damages if translocated into peripheral circulation (**Figure 1**). This process of translocation is called microbial translocation (44).

Evidence from GF animals suggests that the development and function of the intestinal barrier are dependent on microbiota. In GF animals, due to the lack of bacterial stimulations, the thickness of the mucus layers is extremely reduced (45–48). The important role of gut microbiota in modulating mucin production from goblet cells is further evidenced in animals with lower loads of bacteria (49, 50). The thinner mucus layers would allow for bacteria penetration, which may initiate inflammation and inflammatory diseases such as colitis (46, 51). Commensal bacteria, or bacterial products such as lipopolysaccharide (LPS) and peptidoglycan, can restore the mucus layers (46, 47). A balance exists between commensal bacteria and the mucus layers, and together they contribute to the maintenance of gut homeostasis (48). Within the mucus layers, there are diverse secreted AMPs that can clear pathogens and control the colonization of commensal bacteria. Reciprocally, the production of some AMPs is regulated by microbiota and/or their products. For instance, RegIII γ is the AMP necessary for physically separating commensal bacteria from intestinal epithelium (11). RegIII γ has been shown to be suppressed in alcoholic patients and mice receiving ethanol treatment (52, 53). Prebiotics administration, or increasing probiotic *Lactobacilli* and *Bifidobacteria*, has been shown to restore the properties of RegIII γ and control bacterial overgrowth (53). Ang4, a member of angiogenin family, is another example where gut commensals are known to modulate AMP production. In one study, Gordon and coworkers found that the production and secretion of Ang4 from mouse Paneth cells were induced by a predominant gut microflora, *Bacteroides thetaiotaomicron* (54). Therefore, the antibacterial activity of Ang4 against microbes in gut lumen is, in turn, dependent on the existence of certain commensal species.

In addition, an interaction exists between gut microbes and AMPs, such as IAP. Predominately produced by IECs, IAP is active either anchored on the epithelium membrane or secreted into gut lumen (55, 56). In IAP-deficient mice, it was noted that there were fewer microbes and an altered bacteria composition compared to control wild-type animals. In particular, the researchers noted a decrease in *Lactobacillaceae* (57, 58). Upregulated IAP activity can selectively increase LPS-suppressing bacteria (e.g., *Bifidobacterium*), while reducing LPS-producing bacteria

(e.g., *Escherichia coli*) (59). Having the capacity to inactivate LPS *in vivo*, IAP is vital in preventing the translocation of LPS, the pro-inflammatory stimulus originated from bacteria (60, 61). Of note, the expression of IAP relies on the presence of microbiota. In GF zebrafish, the colonization of commensals, or even supplying LPS alone, could sufficiently induce IAP expression (62). It is worth mentioning that IAP can also regulate TJ proteins to enhance barrier function through increasing ZO-1, ZO-2, and occludin expression (63). Several others have also reported on the various types of AMPs and their function in the microbiota (64, 65).

Intestinal epithelial cells compose the single layer of intestinal epithelium, and the generation of new IECs from local intestinal stem cells is vital in maintaining the barrier function due to the high frequency of apoptosis and shedding of IECs (66). As much as 10% of all the gene transcriptions, especially genes related to immunity, cell proliferation, and metabolism, in IECs are regulated by gut microbiota (67). In GF and antibiotic-treated mice, epithelial proliferation rate is reduced, suggesting the role of microbiota on epithelium cell renewal (68, 69). LPS from *E. coli* can induce cell shedding in a dose-dependent manner (70, 71). Colonization of *Bifidobacterium breve*, or more precisely its surface component, exopolysaccharide, can positively modulate LPS-induced epithelium cell shedding through epithelial MyD88 signaling (70). The renewal of IECs relies on the activity of intestinal stem cells that are located at the base of crypts and express TLR4, the LPS receptor. TLR4 activation has been demonstrated to inhibit proliferation and promote the apoptosis of Lgr5⁺ intestinal stem cells. In mice bearing selective TLR4 deletion in intestinal stem cells, LPS is no longer able to inhibit the renewal of IECs (72). This process was found to be mediated by the p53-upregulated modulator of apoptosis (PUMA) as TLR4 activation in mice lacking PUMA was unaltered. Apart from LPS, bacterial metabolites, particularly butyrate, have also been identified as inhibitors of intestinal stem cell proliferation (73). The intestinal crypt architecture protects the intestinal stem cells from the negative effect of butyrate. As gatekeepers for the paracellular pathway, TJ complexes are also major targets of microbiota regulation (74). This is particularly true for certain probiotic species including, but not limited to, *Lactobacillus rhamnosus* (75–78), *Streptococcus thermophilus* (79), *Lactobacillus reuteri* (80), and *Bifidobacterium infantis* (81).

MECHANISMS OF LEAKY GUT

A large variety of gut barrier disruptors and/or gut microbiota disturbers may potentially result in microbial translocation and subsequent inflammation locally and systemically. These include diet, infections, alcohol consumption, and burn injury.

Diet-Induced Gut Leakiness

Nutrients and food ingredients have been reported to contribute to the maintenance or alterations of gut microbiota and the intestinal barrier function (82). A recent review by De Santis et al. detailed many dietary factors that may modulate the intestinal barrier (83). Here, we review some recent publications and emphasize the effects of diet-induced alterations of gut

microbiota on compromising the gut barrier function. Vitamin D (VD) has been recognized as an intestinal permeability protector by inducing the expression of TJ proteins ZO-1 and claudin-1. In VD receptor (VDR)-knockout mice, more severe experimental colitis has been observed, suggesting the protective effect of VD on the mucosal barrier (84). However, another group have recently found that VDR deficiency lowers, whereas VD treatment upregulates, the expression of claudin-2, a pore-forming TJ protein, which renders the intestinal epithelium leaky (85). Further analysis confirmed that VDR enhanced claudin-2 promoter activity. The exact role of VD and VDR on modulating intestinal permeability is therefore unclear and should be investigated carefully in association with gut microbiota. In a recent study by Desai et al., a low-fiber diet consumption was found to trigger the expansion of mucus-degrading bacteria, including *Akkermansia muciniphila* and *Bacteroides caccae* (45). As a result, the thickness of mucus is significantly decreased in mice fed with fiber-deficient diets, although the transcription of *Muc2* gene was surprisingly heightened, possibly as a compensatory response. The thinner mucus and compromised intestinal barrier function lead to a higher susceptibility to certain colitis-causing pathogens (45). Moreover, a diet high in saturated fat has been shown to greatly decrease *Lactobacillus* and increase *Oscillibacter*, and these changes were correlated with significantly increased permeability in the proximal colon (86). Furthermore, studies revealed that the abundance of the *Oscillospira* genus was negatively correlated with the mRNA expression of barrier-forming TJ protein ZO-1.

Stress-Induced Gut Leakiness

Under certain circumstances, stress-induced alterations of gut microbiota and the impaired intestinal barrier would allow the occurrence of microbial translocation. Burn injury and alcohol consumption are examples of such stress. Burn injury results in increased intestinal permeability, which is mediated by increased activity of myosin light-chain (MLC) kinase (87, 88). It is known that MLC phosphorylation or kinase activation can trigger epithelial TJ opening (89–91). In burn injury, TJ proteins, including ZO-1, occluding, and claudin-1, are redistributed, which can be reversed by adding an MLC phosphorylation inhibitor (87). In addition, both humans and mice experiencing burn injury undergo similar alterations of gut microbiota, in particular, with increases of the abundance of bacteria from the *Enterobacteriaceae* family (88). Importantly, microbial translocation of these Gram-negative aerobic bacteria has been observed. Another research group, using a different burn injury mouse model reported increased colonic permeability together with reduced aerobic and anaerobic bacterial populations in the gut microbiota, particularly those producing butyrate (92). As a consequence, the butyrate level in the stool was significantly decreased in mice with burn injury. Interestingly, when the experimental mice received fecal microbiota transplant, their altered bacterial counts and impaired mucosal barrier function were reversed, suggesting direct involvement of microbiota in causing gut leakiness after burn injury.

Chronic alcohol consumption is responsible for intestinal barrier dysfunction, alterations on both the quality and quantity of gut microbiota, LPS translocation, and alcoholic liver disease

(ALD). In both human and mouse, it has been well established that alcohol can disrupt intestinal barrier function, which is closely related to increased tumor necrosis factor (TNF) production from intestinal monocytes/macrophages and enterocytes bearing TNF-receptor 1, followed by downstream activation of MLC kinase (93). Notably, when mice given chronic alcohol also received oral antibiotic treatment, to remove the microbiota, the level of TNF production and intestinal permeability decreased to levels comparable to those in control mice (93). This indicates that the alcohol-induced, TNF-mediated gut leakiness is greatly dependent on gut microbiota. Indeed, though the mechanism is unknown, alcohol administration alters microbiota qualitatively and quantitatively in both human and mouse (94). Bacterial overgrowth has been observed with alcohol consumption, whereas antibiotics can decrease the bacterial load and attenuate ALD (53, 93, 95–97). Interestingly, probiotic *Lactobacillus* is significantly suppressed during alcohol consumption (53, 97). Directly supplying *Lactobacillus* strains or indirect stimulation of *Lactobacilli* with prebiotics or diets can decrease bacterial overgrowth, restore mucosal integrity of the intestine, and suppress microbial translocation (53, 94, 98, 99). Microbial translocation, especially the translocation of LPS, is involved in ALD development and progression as evidenced by the lack of ALD in mice deficient of TLR4 (100, 101). It is worth noting that some bacteria species can produce alcohol, including *E. coli* and *Weissella confusa*, and this may be the mechanism by which they compromise the intestinal barrier function (102, 103).

Infections can play a role in regulating the mucosal barrier. A good example is *Helicobacter pylori*, a Gram-negative bacterium infecting the human stomach (104). *H. pylori* is known to directly increase epithelial permeability by redistributing TJ protein ZO-1 (105, 106). In addition, bacteriophages, which are usually not considered pathogenic to mammals, can have an impact on the leaky gut. When rats were given a bacteriophage cocktail containing phages against *Salmonella enterica*, disruption of the intestinal barrier integrity was observed (107). The authors speculated that the gut microbiota might have been affected by bacteriophages, but sequencing data were not supplied to support their claims.

Taken together, perturbation of gut microbiota, which may be the consequence of diverse interventions, can lead to increased intestinal permeability and translocation of bacterial components and products. Such microbial translocation can subsequently trigger an abnormal immune response, causing inflammation and/or tissue damage in extraintestinal organs.

LEAKY GUT AND AUTOIMMUNE DISORDERS

Several disease states have been associated with gut microbiota dysbiosis, intestinal barrier dysfunction, and microbial translocation. These include Alzheimer's disease, ALD, cancer, and multiple autoimmune disorders. Autoimmune disorders are characterized by the generation of autoantibodies against self-antigens that attack the body's own tissues, resulting in damage. Genetic and environmental triggers have been long known as the major contributors to the development of autoimmunity.

Increasing evidence in recent years suggests that microbial translocation and intestinal barrier dysfunction, which may be affected by gut microbiota, are another important causative element for autoimmune disorders (2–6). T1D and SLE are examples discussed below that reveal advancements in the understanding of the mechanisms behind the interaction between the leaky gut and autoimmune disorders.

Type 1 Diabetes

Type 1 diabetes is an organ-specific autoimmune disorder characterized by an autoimmune response against the host's own pancreatic β cells, leading to insufficient insulin production from the pancreas (108). Some argue that the leaky gut is only an outcome of disease progression rather than an initiator or exacerbator of disease (109), but this should not be the case for T1D. This is supported by the following evidences. First, studies utilizing human subjects affected by T1D or T1D-prone animal models have indicated that impaired intestinal barrier function occurs before disease onset (110–112). Second, the pathogenic role that increased intestinal permeability plays in T1D is zonulin-dependent, and the production of zonulin relies on bacterial colonization (113). Reversion of intestinal barrier dysbiosis by adding a zonulin inhibitor ameliorated T1D manifestations in disease-prone rats (114). Third, a recent study has provided evidence that microbial translocation contributes to T1D development (115). In streptozotocin-induced T1D, mice treated with streptozotocin harbor a distinct microbiota compared to vehicle-treated controls. Importantly, gut bacteria were shown to be able to translocate into pancreatic lymph nodes (PLNs) and contribute to T1D development (115). When mice were treated with oral antibiotics, PLNs appeared to be sterile and the disease was attenuated. Further analysis revealed that the translocated bacteria in PLNs triggered NOD2 activation and exacerbated T1D. Altogether, these results suggest an essential role for the leaky gut in driving the progression of T1D.

Systemic Lupus Erythematosus

Systemic lupus erythematosus, or lupus, is an autoimmune disorder characterized by severe and persistent inflammation that leads to tissue damage in multiple organs (116). Although SLE affects both men and women, women of childbearing age are diagnosed about nine times more often than men. LPS, a cell wall component of Gram-negative bacteria, can promote SLE development and disease progression upon penetration of the intestinal epithelium and translocation into tissues (117). In SLE patients, the higher level of soluble CD14 suggests an increase in LPS, as soluble CD14 is released from monocytes when the cells are exposed to LPS (118). Activation of TLR4 exacerbates lupus development (119–121). Mice spontaneously develop lupus when TLR4 responsiveness is increased, whereas the exacerbated disease phenotype can be significantly ameliorated when the commensal gut flora is removed by antibiotic treatment (121). This clearly indicates that TLR4 hyperresponsiveness to gut flora (which contains LPS) contributes to the pathogenesis of SLE. Moreover, the development of lupus in wild-type mice (C57BL/6 or BALB/c) immunized with phospholipid-binding proteins can be facilitated by the administration of LPS (122–124). Conversely, inhibition of TLR4 results in reduced autoantibody production

and lowered renal glomerular IgG deposits in lupus-prone mice (125, 126). Taken together, these data suggest that LPS stimulation and TLR4 activation as disease-initiating factors for SLE. Lipoteichoic acid (LTA), a component of the Gram-positive bacterial cell wall, can also promote lupus disease. The expression of TLR2, the receptor of LTA, has been reported to be increased in SLE patients (127). In lupus-prone mice, TLR2 activation triggers lupus nephritis, whereas TLR2 knockout attenuates lupus-like symptoms (125, 128–130). Recently, another bacterial antigen that may mimic self-antigens has been recognized to induce autoantibody production (131).

Several downstream proteins in the TLR signaling cascade are highly relevant to the pathogenesis of SLE and are potential therapeutic targets, including MyD88, IRAKs, and IFN α (132). Deficiency of MyD88, in particular, has been shown to ameliorate lupus disease in MRL/lpr mice (133, 134), suggesting a potential role for TLRs to communicate with harmful bacteria in the gut microbiota. Conversely, there is a paucity of data pertaining to members of the NLR family. The most extensively characterized NLRs are associated with inflammasome formation (135, 136). Loss of NLRP3 and AIM2 inflammasome function was found to significantly contribute to lupus pathogenesis (137). Interestingly, both of these inflammasomes were found compromised in NZB mice, a lupus-prone model. Consistent with this finding, loss of ASC (apoptosis-associated speck-like protein containing CARD), a common adaptor protein required for inflammasome formation in B6-*Fas*^{lpr} mice led to exacerbation of lupus-like disease (138). These results suggest a potential role for NLRs to recognize protective bacteria in the gut microbiota. Therefore, it appears that TLRs and NLRs make distinct contributions to lupus pathogenesis by sensing harmful and protective bacteria, respectively. Both types of bacteria can come from gut microbiota through microbial translocation, especially in the presence of a leaky gut.

REVERSING THE LEAKY GUT AS A POTENTIAL THERAPY

Considering the contributions of leaky gut and bacterial translocation to inflammation and multiple diseases, reversing gut leakiness appears to be an attractive therapeutic strategy. Prebiotics and probiotics, for example, can be used to reduce intestinal permeability (139). Diverse probiotic species have been uncovered that possess the properties to protect the intestinal barrier through targeting different components of the mucosal barrier system. The human commensal *Bacteroides fragilis* may serve as such a probiotic (140). In a mouse model, autism spectrum disorder (ASD) has been shown to be accompanied by intestinal barrier dysfunction, gut microbiota dysbiosis, and leakiness of 4-ethylphenylsulfate (4EPS), which originates from the commensal bacteria. When 4EPS was given to wild-type mice, it directly caused behavioral abnormalities similar to ASD mice. Treatment with *B. fragilis* reduced the translocation of disease-causative 4EPS, and significantly ameliorated the behavior defects. The therapeutic benefit of *B. fragilis* is believed to be due to its ability to alter microbial composition and enhance intestinal barrier function (140). *B. fragilis* is also known for its capability to induce the

development of Foxp3⁺ regulatory T cells, a process regulated by another product of *B. fragilis*, polysaccharide A (PSA) (141, 142). *B. fragilis* and PSA are beneficial against inflammatory diseases, such as colitis and experimental autoimmune encephalomyelitis (141, 143). The application of *B. fragilis* to prevent the leaky gut and reverse autoimmunity warrants further investigation. In a practical point of view, probiotic candidates with different targets on reversing the leaky gut may synergistically act to attenuate disease as thus may serve as a probiotic cocktail. As probiotics are generally considered safe, it is anticipated that they will become cost-effective treatment options for people with autoimmune diseases in the foreseeable future. This is a very young but exciting field in which much still remains to be learned.

REFERENCES

- Peterson LW, Artis D. Intestinal epithelial cells: regulators of barrier function and immune homeostasis. *Nat Rev Immunol* (2014) 14(3):141–53. doi:10.1038/nri3608
- Lin R, Zhou L, Zhang J, Wang B. Abnormal intestinal permeability and microbiota in patients with autoimmune hepatitis. *Int J Clin Exp Pathol* (2015) 8(5):5153–60.
- Khaleghi S, Ju JM, Lamba A, Murray JA. The potential utility of tight junction regulation in celiac disease: focus on larazotide acetate. *Therap Adv Gastroenterol* (2016) 9(1):37–49. doi:10.1177/1756283X15616576
- Fasano A, Shea-Donohue T. Mechanisms of disease: the role of intestinal barrier function in the pathogenesis of gastrointestinal autoimmune diseases. *Nat Clin Pract Gastroenterol Hepatol* (2005) 2(9):416–22. doi:10.1038/ncpgasthep0259
- Traskalova-Hogenova H, Stepankova R, Kozakova H, Hudcovic T, Vannucci L, Tuckova L, et al. The role of gut microbiota (commensal bacteria) and the mucosal barrier in the pathogenesis of inflammatory and autoimmune diseases and cancer: contribution of germ-free and gnotobiotic animal models of human diseases. *Cell Mol Immunol* (2011) 8(2):110–20. doi:10.1038/cmi.2010.67
- Fasano A. Zonulin, regulation of tight junctions, and autoimmune diseases. *Ann N Y Acad Sci* (2012) 1258:25–33. doi:10.1111/j.1749-6632.2012.06538.x
- Jimenez JA, Uwiera TC, Douglas Inglis G, Uwiera RR. Animal models to study acute and chronic intestinal inflammation in mammals. *Gut Pathog* (2015) 7:29. doi:10.1186/s13099-015-0076-y
- Brugman S. The zebrafish as a model to study intestinal inflammation. *Dev Comp Immunol* (2016) 64:82–92. doi:10.1016/j.dci.2016.02.020
- Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* (2012) 336(6086):1268–73. doi:10.1126/science.1223490
- Gerbe F, Legraverend C, Jay P. The intestinal epithelium tuft cells: specification and function. *Cell Mol Life Sci* (2012) 69(17):2907–17. doi:10.1007/s00018-012-0984-7
- Vaishnava S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, et al. The antibacterial lectin RegIIIgamma promotes the spatial segregation of microbiota and host in the intestine. *Science* (2011) 334(6053):255–8. doi:10.1126/science.1209791
- Cash HL, Whitham CV, Behrendt CL, Hooper LV. Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* (2006) 313(5790):1126–30. doi:10.1126/science.1127119
- Vaishnava S, Behrendt CL, Ismail AS, Eckmann L, Hooper LV. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc Natl Acad Sci U S A* (2008) 105(52):20858–63. doi:10.1073/pnas.0808723105
- Mabbott NA, Donaldson DS, Ohno H, Williams IR, Mahajan A. Microfold (M) cells: important immunosurveillance posts in the intestinal epithelium. *Mucosal Immunol* (2013) 6(4):666–77. doi:10.1038/mi.2013.30
- van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu Rev Physiol* (2009) 71:241–60. doi:10.1146/annurev.physiol.010908.163145

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

Preparation of this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number R03AI117597. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

- Anderson JM, Van Itallie CM. Physiology and function of the tight junction. *Cold Spring Harb Perspect Biol* (2009) 1(2):a002584. doi:10.1101/cshperspect.a002584
- Yamazaki Y, Okawa K, Yano T, Tsukita S, Tsukita S. Optimized proteomic analysis on gels of cell-cell adhering junctional membrane proteins. *Biochemistry* (2008) 47(19):5378–86. doi:10.1021/bi8002567
- Singh PK, Parsek MR, Greenberg EP, Welsh MJ. A component of innate immunity prevents bacterial biofilm development. *Nature* (2002) 417(6888):552–5. doi:10.1038/417552a
- Hansson GC. Role of mucus layers in gut infection and inflammation. *Curr Opin Microbiol* (2012) 15(1):57–62. doi:10.1016/j.mib.2011.11.002
- Pelaseyed T, Bergstrom JH, Gustafsson JK, Ermund A, Birchenough GM, Schutte A, et al. The mucus and mucins of the goblet cells and enterocytes provide the first defense line of the gastrointestinal tract and interact with the immune system. *Immunol Rev* (2014) 260(1):8–20. doi:10.1111/imr.12182
- Van der Sluis M, De Koning BA, De Bruijn AC, Velcich A, Meijerink JP, Van Goudoever JB, et al. Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* (2006) 131(1):117–29. doi:10.1053/j.gastro.2006.04.020
- Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. *Physiol Rev* (2010) 90(3):859–904. doi:10.1152/physrev.00045.2009
- Baumler AJ, Sperandio V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* (2016) 535(7610):85–93. doi:10.1038/nature18849
- Ming L, Zhang Q, Yang L, Huang JA. Comparison of antibacterial effects between antimicrobial peptide and bacteriocins isolated from *Lactobacillus plantarum* on three common pathogenic bacteria. *Int J Clin Exp Med* (2015) 8(4):5806–11.
- Ramakrishna BS. Role of the gut microbiota in human nutrition and metabolism. *J Gastroenterol Hepatol* (2013) 28(Suppl 4):9–17. doi:10.1111/jgh.12294
- Krajmalnik-Brown R, Ilhan ZE, Kang DW, DiBaise JK. Effects of gut microbes on nutrient absorption and energy regulation. *Nutr Clin Pract* (2012) 27(2):201–14. doi:10.1177/0884533611436116
- Dupont A, Heinbockel L, Brandenburg K, Hornef MW. Antimicrobial peptides and the enteric mucus layer act in concert to protect the intestinal mucosa. *Gut Microbes* (2014) 5(6):761–5. doi:10.4161/19490976.2014.972238
- Mukherjee S, Hooper LV. Antimicrobial defense of the intestine. *Immunity* (2015) 42(1):28–39. doi:10.1016/j.immuni.2014.12.028
- Hofmann AF, Eckmann L. How bile acids confer gut mucosal protection against bacteria. *Proc Natl Acad Sci U S A* (2006) 103(12):4333–4. doi:10.1073/pnas.0600780103
- Ostaf MJ, Stange EF, Wehkamp J. Antimicrobial peptides and gut microbiota in homeostasis and pathology. *EMBO Mol Med* (2013) 5(10):1465–83. doi:10.1002/emmm.201201773
- Kopp ZA, Jain U, Van Limbergen J, Stadnyk AW. Do antimicrobial peptides and complement collaborate in the intestinal mucosa? *Front Immunol* (2015) 6:17. doi:10.3389/fimmu.2015.00017

32. Howe SE, Lickteig DJ, Plunkett KN, Ryerse JS, Konjufca V. The uptake of soluble and particulate antigens by epithelial cells in the mouse small intestine. *PLoS One* (2014) 9(1):e86656. doi:10.1371/journal.pone.0086656
33. McDole JR, Wheeler LW, McDonald KG, Wang B, Konjufca V, Knoop KA, et al. Goblet cells deliver luminal antigen to CD103+ dendritic cells in the small intestine. *Nature* (2012) 483(7389):345–9. doi:10.1038/nature10863
34. Knoop KA, McDonald KG, McCrater S, McDole JR, Newberry RD. Microbial sensing by goblet cells controls immune surveillance of luminal antigens in the colon. *Mucosal Immunol* (2015) 8(1):198–210. doi:10.1038/mi.2014.58
35. Birchenough GM, Johansson ME, Gustafsson JK, Bergstrom JH, Hansson GC. New developments in goblet cell mucus secretion and function. *Mucosal Immunol* (2015) 8(4):712–9. doi:10.1038/mi.2015.32
36. Knoop KA, McDonald KG, Kulkarni DH, Newberry RD. Antibiotics promote inflammation through the translocation of native commensal colonic bacteria. *Gut* (2016) 65(7):1100–9. doi:10.1136/gutjnl-2014-309059
37. Rescigno M, Urbano M, Valzasina B, Francolini M, Rotta G, Bonasio R, et al. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat Immunol* (2001) 2(4):361–7. doi:10.1038/86373
38. Vallon-Eberhard A, Landsman L, Yogev N, Verrier B, Jung S. Transepithelial pathogen uptake into the small intestinal lamina propria. *J Immunol* (2006) 176(4):2465–9. doi:10.4049/jimmunol.176.4.2465
39. Diehl GE, Longman RS, Zhang JX, Breart B, Galan C, Cuesta A, et al. Microbiota restricts trafficking of bacteria to mesenteric lymph nodes by CX3CR1(hi) cells. *Nature* (2013) 494(7435):116–20. doi:10.1038/nature11809
40. Niess JH, Brand S, Gu X, Landsman L, Jung S, McCormick BA, et al. CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* (2005) 307(5707):254–8. doi:10.1126/science.1102901
41. Gross M, Salame TM, Jung S. Guardians of the gut – murine intestinal macrophages and dendritic cells. *Front Immunol* (2015) 6:254. doi:10.3389/fimmu.2015.00254
42. Kim KW, Vallon-Eberhard A, Zigmund E, Farache J, Shezen E, Shakhar G, et al. In vivo structure/function and expression analysis of the CX3C chemokine fractalkine. *Blood* (2011) 118(22):e156–67. doi:10.1182/blood-2011-04-348946
43. Woof JM, Russell MW. Structure and function relationships in IgA. *Mucosal Immunol* (2011) 4(6):590–7. doi:10.1038/mi.2011.39
44. Brenchley JM, Douek DC. Microbial translocation across the GI tract. *Annu Rev Immunol* (2012) 30:149–73. doi:10.1146/annurev-immunol-020711-075001
45. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* (2016) 167(5):1339–53. doi:10.1016/j.cell.2016.10.043
46. Johansson ME, Jakobsson HE, Holmen-Larsson J, Schutte A, Ermund A, Rodriguez-Pineiro AM, et al. Normalization of host intestinal mucus layers requires long-term microbial colonization. *Cell Host Microbe* (2015) 18(5):582–92. doi:10.1016/j.chom.2015.10.007
47. Petersson J, Schreiber O, Hansson GC, Gendler SJ, Velcich A, Lundberg JO, et al. Importance and regulation of the colonic mucus barrier in a mouse model of colitis. *Am J Physiol Gastrointest Liver Physiol* (2011) 300(2):G327–33. doi:10.1152/ajpgi.00422.2010
48. Wrzosek L, Miquel S, Noordine ML, Bouet S, Joncquel Chevalier-Curt M, Robert V, et al. *Bacteroides thetaiotaomicron* and *Faecalibacterium prausnitzii* influence the production of mucus glycans and the development of goblet cells in the colonic epithelium of a gnotobiotic model rodent. *BMC Biol* (2013) 11:61. doi:10.1186/1741-7007-11-61
49. Forder RE, Howarth GS, Tivey DR, Hughes RJ. Bacterial modulation of small intestinal goblet cells and mucin composition during early posthatch development of poultry. *Poult Sci* (2007) 86(11):2396–403. doi:10.3382/ps.2007-00222
50. Wlodarska M, Willing B, Keeney KM, Menendez A, Bergstrom KS, Gill N, et al. Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated *Citrobacter rodentium*-induced colitis. *Infect Immun* (2011) 79(4):1536–45. doi:10.1128/IAI.01104-10
51. Johansson ME, Gustafsson JK, Sjöberg KE, Petersson J, Holm L, Sjövall H, et al. Bacteria penetrate the inner mucus layer before inflammation in the dextran sulfate colitis model. *PLoS One* (2010) 5(8):e12238. doi:10.1371/journal.pone.0012238
52. Wang L, Fouts DE, Starkel P, Hartmann P, Chen P, Llorente C, et al. Intestinal REG3 lectins protect against alcoholic steatohepatitis by reducing mucosa-associated microbiota and preventing bacterial translocation. *Cell Host Microbe* (2016) 19(2):227–39. doi:10.1016/j.chom.2016.01.003
53. Yan AW, Fouts DE, Brandl J, Starkel P, Torralba M, Schott E, et al. Enteric dysbiosis associated with a mouse model of alcoholic liver disease. *Hepatology* (2011) 53(1):96–105. doi:10.1002/hep.24018
54. Hooper LV, Stappenbeck TS, Hong CV, Gordon JI. Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol* (2003) 4(3):269–73. doi:10.1038/ni888
55. Eliakim R, Mahmood A, Alpers DH. Rat intestinal alkaline phosphatase secretion into lumen and serum is coordinately regulated. *Biochim Biophys Acta* (1991) 1091(1):1–8. doi:10.1016/0167-4889(91)90213-H
56. Nakano T, Inoue I, Alpers DH, Akiba Y, Katayama S, Shinozaki R, et al. Role of lysophosphatidylcholine in brush-border intestinal alkaline phosphatase release and restoration. *Am J Physiol Gastrointest Liver Physiol* (2009) 297(1):G207–14. doi:10.1152/ajpgi.90590.2008
57. Malo MS, Alam SN, Mostafa G, Zeller SJ, Johnson PV, Mohammad N, et al. Intestinal alkaline phosphatase preserves the normal homeostasis of gut microbiota. *Gut* (2010) 59(11):1476–84. doi:10.1136/gut.2010.211706
58. Manichanh C, Rigottier-Gois L, Bonnaud E, Gloux K, Pelletier E, Frangeul L, et al. Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* (2006) 55(2):205–11. doi:10.1136/gut.2005.073817
59. Kaliannan K, Wang B, Li XY, Kim KJ, Kang JX. A host-microbiome interaction mediates the opposing effects of omega-6 and omega-3 fatty acids on metabolic endotoxemia. *Sci Rep* (2015) 5:11276. doi:10.1038/srep11276
60. Bates JM, Akerlund J, Mitge E, Guillemin K. Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota. *Cell Host Microbe* (2007) 2(6):371–82. doi:10.1016/j.chom.2007.10.010
61. Yang Y, Millan JL, Mecsas J, Guillemin K. Intestinal alkaline phosphatase deficiency leads to lipopolysaccharide desensitization and faster weight gain. *Infect Immun* (2015) 83(1):247–58. doi:10.1128/IAI.02520-14
62. Bates JM, Mitge E, Kuhlman J, Baden KN, Cheesman SE, Guillemin K. Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Dev Biol* (2006) 297(2):374–86. doi:10.1016/j.ydbio.2006.05.006
63. Liu W, Hu D, Huo H, Zhang W, Adiliaghdam F, Morrison S, et al. Intestinal alkaline phosphatase regulates tight junction protein levels. *J Am Coll Surg* (2016) 222(6):1009–17. doi:10.1016/j.jamcollsurg.2015.12.006
64. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nat Rev Microbiol* (2011) 9(5):356–68. doi:10.1038/nrmicro2546
65. Gallo RL, Hooper LV. Epithelial antimicrobial defence of the skin and intestine. *Nat Rev Immunol* (2012) 12(7):503–16. doi:10.1038/nri3228
66. Williams JM, Duckworth CA, Burkitt MD, Watson AJ, Campbell BJ, Pritchard DM. Epithelial cell shedding and barrier function: a matter of life and death at the small intestinal villus tip. *Vet Pathol* (2015) 52(3):445–55. doi:10.1177/0300985814559404
67. Sommer F, Nookaew I, Sommer N, Fogelstrand P, Backhed F. Site-specific programming of the host epithelial transcriptome by the gut microbiota. *Genome Biol* (2015) 16:62. doi:10.1186/s13059-015-0614-4
68. Reikvam DH, Erofeev A, Sandvik A, Gric V, Jahnsen FL, Gaustad P, et al. Depletion of murine intestinal microbiota: effects on gut mucosa and epithelial gene expression. *PLoS One* (2011) 6(3):e17996. doi:10.1371/journal.pone.0017996
69. Stedman A, Nigro G, Sansonetti PJ. [Microbiota-intestinal stem cells dialog: a key element for intestinal regeneration]. *Med Sci (Paris)* (2016) 32(11):983–90. doi:10.1051/medsci/20163211014
70. Hughes KR, Harnisch LC, Alcon-Giner C, Mitra S, Wright CJ, Ketskemety J, et al. *Bifidobacterium breve* reduces apoptotic epithelial cell shedding in an exopolysaccharide and MyD88-dependent manner. *Open Biol* (2017) 7(1):160155. doi:10.1098/rsob.160155
71. Williams JM, Duckworth CA, Watson AJ, Frey MR, Miguel JC, Burkitt MD, et al. A mouse model of pathological small intestinal

- epithelial cell apoptosis and shedding induced by systemic administration of lipopolysaccharide. *Dis Model Mech* (2013) 6(6):1388–99. doi:10.1242/dmm.013284
72. Neal MD, Sodhi CP, Jia H, Dyer M, Egan CE, Yazji I, et al. Toll-like receptor 4 is expressed on intestinal stem cells and regulates their proliferation and apoptosis via the p53 up-regulated modulator of apoptosis. *J Biol Chem* (2012) 287(44):37296–308. doi:10.1074/jbc.M112.375881
 73. Kaiko GE, Ryu SH, Koues OI, Collins PL, Solnica-Krezel L, Pearce EJ, et al. The colonic crypt protects stem cells from microbiota-derived metabolites. *Cell* (2016) 167(4):1137. doi:10.1016/j.cell.2016.10.034
 74. Ulluwishewa D, Anderson RC, McNabb WC, Moughan PJ, Wells JM, Roy NC. Regulation of tight junction permeability by intestinal bacteria and dietary components. *J Nutr* (2011) 141(5):769–76. doi:10.3945/jn.110.135657
 75. Donato KA, Gareau MG, Wang YJ, Sherman PM. *Lactobacillus rhamnosus* GG attenuates interferon- γ and tumour necrosis factor- α -induced barrier dysfunction and pro-inflammatory signalling. *Microbiology* (2010) 156(Pt 11):3288–97. doi:10.1099/mic.0.040139-0
 76. Johnson-Henry KC, Donato KA, Shen-Tu G, Gordanpour M, Sherman PM. *Lactobacillus rhamnosus* strain GG prevents enterohemorrhagic *Escherichia coli* O157: H7-induced changes in epithelial barrier function. *Infect Immun* (2008) 76(4):1340–8. doi:10.1128/IAI.00778-07
 77. Miyauchi E, Morita H, Tanabe S. *Lactobacillus rhamnosus* alleviates intestinal barrier dysfunction in part by increasing expression of zonula occludens-1 and myosin light-chain kinase in vivo. *J Dairy Sci* (2009) 92(6):2400–8. doi:10.3168/jds.2008-1698
 78. Seth A, Yan F, Polk DB, Rao RK. Probiotics ameliorate the hydrogen peroxide-induced epithelial barrier disruption by a PKC- and MAP kinase-dependent mechanism. *Am J Physiol Gastrointest Liver Physiol* (2008) 294(4):G1060–9. doi:10.1152/ajpgi.00202.2007
 79. Resta-Lenert S, Barrett KE. Live probiotics protect intestinal epithelial cells from the effects of infection with enteroinvasive *Escherichia coli* (EIEC). *Gut* (2003) 52(7):988–97. doi:10.1136/gut.52.7.988
 80. Ewaschuk JB, Diaz H, Meddings L, Diederichs B, Dmytrash A, Backer J, et al. Secreted bioactive factors from *Bifidobacterium infantis* enhance epithelial cell barrier function. *Am J Physiol Gastrointest Liver Physiol* (2008) 295(5):G1025–34. doi:10.1152/ajpgi.90227.2008
 81. Ahrne S, Hagslåt ML. Effect of lactobacilli on paracellular permeability in the gut. *Nutrients* (2011) 3(1):104–17. doi:10.3390/nu3010104
 82. Suzuki T. Regulation of intestinal epithelial permeability by tight junctions. *Cell Mol Life Sci* (2013) 70(4):631–59. doi:10.1007/s00018-012-1070-x
 83. De Santis S, Cavalcanti E, Mastronardi M, Jirillo E, Chieppa M. Nutritional keys for intestinal barrier modulation. *Front Immunol* (2015) 6:612. doi:10.3389/fimmu.2015.00612
 84. Kong J, Zhang Z, Musch MW, Ning G, Sun J, Hart J, et al. Novel role of the vitamin D receptor in maintaining the integrity of the intestinal mucosal barrier. *Am J Physiol Gastrointest Liver Physiol* (2008) 294(1):G208–16. doi:10.1152/ajpgi.00398.2007
 85. Zhang YG, Wu S, Lu R, Zhou D, Zhou J, Carmeliet G, et al. Tight junction CLDN2 gene is a direct target of the vitamin D receptor. *Sci Rep* (2015) 5:10642. doi:10.1038/srep10642
 86. Lam YY, Ha CW, Campbell CR, Mitchell AJ, Dinudom A, Oscarsson J, et al. Increased gut permeability and microbiota change associate with mesenteric fat inflammation and metabolic dysfunction in diet-induced obese mice. *PLoS One* (2012) 7(3):e34233. doi:10.1371/journal.pone.0034233
 87. Chen C, Wang P, Su Q, Wang S, Wang F. Myosin light chain kinase mediates intestinal barrier disruption following burn injury. *PLoS One* (2012) 7(4):e34946. doi:10.1371/journal.pone.0034946
 88. Earley ZM, Akhtar S, Green SJ, Naqib A, Khan O, Cannon AR, et al. Burn injury alters the intestinal microbiome and increases gut permeability and bacterial translocation. *PLoS One* (2015) 10(7):e0129996. doi:10.1371/journal.pone.0129996
 89. Turner JR. Molecular basis of epithelial barrier regulation: from basic mechanisms to clinical application. *Am J Pathol* (2006) 169(6):1901–9. doi:10.2353/ajpath.2006.060681
 90. Turner JR. Intestinal mucosal barrier function in health and disease. *Nat Rev Immunol* (2009) 9(11):799–809. doi:10.1038/nri2653
 91. Cunningham KE, Turner JR. Myosin light chain kinase: pulling the strings of epithelial tight junction function. *Ann N Y Acad Sci* (2012) 1258:34–42. doi:10.1111/j.1749-6632.2012.06526.x
 92. Kuethe JW, Armocida SM, Midura EF, Rice TC, Hildeman DA, Healy DP, et al. Fecal microbiota transplant restores mucosal integrity in a murine model of burn injury. *Shock* (2016) 45(6):647–52. doi:10.1097/SHK.0000000000000551
 93. Chen P, Starkel P, Turner JR, Ho SB, Schnabl B. Dysbiosis-induced intestinal inflammation activates tumor necrosis factor receptor 1 and mediates alcoholic liver disease in mice. *Hepatology* (2015) 61(3):883–94. doi:10.1002/hep.27489
 94. Chen P, Torralba M, Tan J, Embree M, Zengler K, Starkel P, et al. Supplementation of saturated long-chain fatty acids maintains intestinal eubiosis and reduces ethanol-induced liver injury in mice. *Gastroenterology* (2015) 148(1):203–14 e16. doi:10.1053/j.gastro.2014.09.014
 95. Adachi Y, Moore LE, Bradford BU, Gao W, Thurman RG. Antibiotics prevent liver injury in rats following long-term exposure to ethanol. *Gastroenterology* (1995) 108(1):218–24. doi:10.1016/0016-5085(95)90027-6
 96. Casafont Morencos F, de las Heras Castano G, Martin Ramos L, Lopez Arias MJ, Ledesma F, Pons Romero F. Small bowel bacterial overgrowth in patients with alcoholic cirrhosis. *Dig Dis Sci* (1996) 41(3):552–6. doi:10.1007/BF02282340
 97. Mutlu EA, Gillevet PM, Rangwala H, Sikaroodi M, Naqvi A, Engen PA, et al. Colonic microbiome is altered in alcoholism. *Am J Physiol Gastrointest Liver Physiol* (2012) 302(9):G966–78. doi:10.1152/ajpgi.00380.2011
 98. Forsyth CB, Farhadi A, Jakate SM, Tang Y, Shaikh M, Keshavarzian A. *Lactobacillus* GG treatment ameliorates alcohol-induced intestinal oxidative stress, gut leakiness, and liver injury in a rat model of alcoholic steatohepatitis. *Alcohol* (2009) 43(2):163–72. doi:10.1016/j.alcohol.2008.12.009
 99. Wang Y, Kirpich I, Liu Y, Ma Z, Barve S, McClain CJ, et al. *Lactobacillus rhamnosus* GG treatment potentiates intestinal hypoxia-inducible factor, promotes intestinal integrity and ameliorates alcohol-induced liver injury. *Am J Pathol* (2011) 179(6):2866–75. doi:10.1016/j.ajpath.2011.08.039
 100. Parlesak A, Schafer C, Schutz T, Bode JC, Bode C. Increased intestinal permeability to macromolecules and endotoxemia in patients with chronic alcohol abuse in different stages of alcohol-induced liver disease. *J Hepatol* (2000) 32(5):742–7. doi:10.1016/S0168-8278(00)80242-1
 101. Uesugi T, Froh M, Arteel GE, Bradford BU, Thurman RG. Toll-like receptor 4 is involved in the mechanism of early alcohol-induced liver injury in mice. *Hepatology* (2001) 34(1):101–8. doi:10.1053/jhep.2001.25350
 102. Elshaghabee FM, Bockelmann W, Meske D, de Vrese M, Walte HG, Schrezenmeier J, et al. Ethanol production by selected intestinal microorganisms and lactic acid bacteria growing under different nutritional conditions. *Front Microbiol* (2016) 7:47. doi:10.3389/fmicb.2016.00047
 103. Tyler AD, Knox N, Kabakchiev B, Milgrom R, Kirsch R, Cohen Z, et al. Characterization of the gut-associated microbiome in inflammatory pouch complications following ileal pouch-anal anastomosis. *PLoS One* (2013) 8(9):e66934. doi:10.1371/journal.pone.0066934
 104. Yu QH, Yang Q. Diversity of tight junctions (TJs) between gastrointestinal epithelial cells and their function in maintaining the mucosal barrier. *Cell Biol Int* (2009) 33(1):78–82. doi:10.1016/j.cellbi.2008.09.007
 105. Amieva MR, Vogelmann R, Covacci A, Tompkins LS, Nelson WJ, Falkow S. Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. *Science* (2003) 300(5624):1430–4. doi:10.1126/science.1081919
 106. Fedwick JP, Lapointe TK, Meddings JB, Sherman PM, Buret AG. *Helicobacter pylori* activates myosin light-chain kinase to disrupt claudin-4 and claudin-5 and increase epithelial permeability. *Infect Immun* (2005) 73(12):7844–52. doi:10.1128/IAI.73.12.7844-7852.2005
 107. Tetz G, Tetz V. Bacteriophage infections of microbiota can lead to leaky gut in an experimental rodent model. *Gut Pathog* (2016) 8:33. doi:10.1186/s13099-016-0109-1
 108. Kawasaki E. Type 1 diabetes and autoimmunity. *Clin Pediatr Endocrinol* (2014) 23(4):99–105. doi:10.1297/cpe.23.99
 109. Odenwald MA, Turner JR. Intestinal permeability defects: is it time to treat? *Clin Gastroenterol Hepatol* (2013) 11(9):1075–83. doi:10.1016/j.cgh.2013.07.001
 110. Damci T, Nuhoglu I, Devranoglu G, Osar Z, Demir M, Ilkova H. Increased intestinal permeability as a cause of fluctuating postprandial blood glucose

- levels in type 1 diabetic patients. *Eur J Clin Invest* (2003) 33(5):397–401. doi:10.1046/j.1365-2362.2003.01161.x
111. Meddings JB, Jarand J, Urbanski SJ, Hardin J, Gall DG. Increased gastrointestinal permeability is an early lesion in the spontaneously diabetic BB rat. *Am J Physiol* (1999) 276(4 Pt 1):G951–7.
 112. Secondulfo M, Iafusco D, Carratu R, deMagistris L, Sapone A, Generoso M, et al. Ultrastructural mucosal alterations and increased intestinal permeability in non-celiac, type I diabetic patients. *Dig Liver Dis* (2004) 36(1):35–45. doi:10.1016/j.dld.2003.09.016
 113. El Asmar R, Panigrahi P, Bamford P, Berti I, Not T, Coppa GV, et al. Host-dependent zonulin secretion causes the impairment of the small intestine barrier function after bacterial exposure. *Gastroenterology* (2002) 123(5):1607–15. doi:10.1053/gast.2002.36578
 114. Watts T, Berti I, Sapone A, Gerarduzzi T, Not T, Zielke R, et al. Role of the intestinal tight junction modulator zonulin in the pathogenesis of type I diabetes in BB diabetic-prone rats. *Proc Natl Acad Sci U S A* (2005) 102(8):2916–21. doi:10.1073/pnas.0500178102
 115. Costa FR, Francozo MC, de Oliveira GG, Ignacio A, Castoldi A, Zamboni DS, et al. Gut microbiota translocation to the pancreatic lymph nodes triggers NOD2 activation and contributes to T1D onset. *J Exp Med* (2016) 213(7):1223–39. doi:10.1084/jem.20150744
 116. Tsokos GC. Systemic lupus erythematosus. *N Engl J Med* (2011) 365(22):2110–21. doi:10.1056/NEJMr1100359
 117. Mu Q, Zhang H, Luo XM. SLE: another autoimmune disorder influenced by microbes and diet? *Front Immunol* (2015) 6:608. doi:10.3389/fimmu.2015.00608
 118. Nockher WA, Wigand R, Schoeppe W, Scherberich JE. Elevated levels of soluble CD14 in serum of patients with systemic lupus erythematosus. *Clin Exp Immunol* (1994) 96(1):15–9. doi:10.1111/j.1365-2249.1994.tb06222.x
 119. Lee TP, Huang JC, Liu CJ, Chen HJ, Chen YH, Tsai YT, et al. Interactions of surface-expressed TLR-4 and endosomal TLR-9 accelerate lupus progression in anti-dsDNA antibody transgenic mice. *Exp Biol Med* (Maywood) (2014) 239(6):715–23. doi:10.1177/1535370214525299
 120. Lee TP, Tang SJ, Wu MF, Song YC, Yu CL, Sun KH. Transgenic overexpression of anti-double-stranded DNA autoantibody and activation of Toll-like receptor 4 in mice induce severe systemic lupus erythematosus syndromes. *J Autoimmun* (2010) 35(4):358–67. doi:10.1016/j.jaut.2010.07.007
 121. Liu B, Yang Y, Dai J, Medzhitov R, Freudenberg MA, Zhang PL, et al. TLR4 up-regulation at protein or gene level is pathogenic for lupus-like autoimmune disease. *J Immunol* (2006) 177(10):6880–8. doi:10.4049/jimmunol.177.10.6880
 122. Levine JS, Subang R, Nasr SH, Fournier S, Lajoie G, Wither J, et al. Immunization with an apoptotic cell-binding protein recapitulates the nephritis and sequential autoantibody emergence of systemic lupus erythematosus. *J Immunol* (2006) 177(9):6504–16. doi:10.4049/jimmunol.177.9.6504
 123. Levine JS, Subang R, Setty S, Cabrera J, Laplante P, Fritzler MJ, et al. Phospholipid-binding proteins differ in their capacity to induce autoantibodies and murine systemic lupus erythematosus. *Lupus* (2014) 23(8):752–68. doi:10.1177/0961203314525676
 124. Tolomeo T, Rico De Souza A, Roter E, Dieude M, Amireault P, Subang R, et al. T cells demonstrate a Th1-biased response to native beta2-glycoprotein I in a murine model of anti-phospholipid antibody induction. *Autoimmunity* (2009) 42(4):292–5. doi:10.1080/08916930902828254
 125. Lartigue A, Colliou N, Calbo S, Francois A, Jacquot S, Arnoult C, et al. Critical role of TLR2 and TLR4 in autoantibody production and glomerulonephritis in lpr mutation-induced mouse lupus. *J Immunol* (2009) 183(10):6207–16. doi:10.4049/jimmunol.0803219
 126. Summers SA, Hoi A, Steinmetz OM, O'Sullivan KM, Ooi JD, Odobasic D, et al. TLR9 and TLR4 are required for the development of autoimmunity and lupus nephritis in pristane nephropathy. *J Autoimmun* (2010) 35(4):291–8. doi:10.1016/j.jaut.2010.05.004
 127. Liu Y, Liao J, Zhao M, Wu H, Yung S, Chan TM, et al. Increased expression of TLR2 in CD4(+) T cells from SLE patients enhances immune reactivity and promotes IL-17 expression through histone modifications. *Eur J Immunol* (2015) 45(9):2683–93. doi:10.1002/eji.201445219
 128. Leiss H, Niederreiter B, Bandur T, Schwarzecker B, Bluml S, Steiner G, et al. Pristane-induced lupus as a model of human lupus arthritis: involvement of autoantibodies, internal organ and joint inflammation. *Lupus* (2013) 22(8):778–92. doi:10.1177/0961203313492869
 129. Pawar RD, Castrezana-Lopez L, Allam R, Kulkarni OP, Segerer S, Radomska E, et al. Bacterial lipopeptide triggers massive albuminuria in murine lupus nephritis by activating toll-like receptor 2 at the glomerular filtration barrier. *Immunology* (2009) 128(1 Suppl):e206–21. doi:10.1111/j.1365-2567.2008.02948.x
 130. Urbonaviciute V, Starke C, Pirschel W, Pohle S, Frey S, Daniel C, et al. Toll-like receptor 2 is required for autoantibody production and development of renal disease in pristane-induced lupus. *Arthritis Rheum* (2013) 65(6):1612–23. doi:10.1002/art.37914
 131. Gallo PM, Rapsinski GJ, Wilson RP, Oppong GO, Sriram U, Goulian M, et al. Amyloid-DNA composites of bacterial biofilms stimulate autoimmunity. *Immunity* (2015) 42(6):1171–84. doi:10.1016/j.immuni.2015.06.002
 132. Wu YW, Tang W, Zuo JP. Toll-like receptors: potential targets for lupus treatment. *Acta Pharmacol Sin* (2015) 36(12):1395–407. doi:10.1038/aps.2015.91
 133. Teichmann LL, Schenten D, Medzhitov R, Kashgarian M, Shlomchik MJ. Signals via the adaptor MyD88 in B cells and DCs make distinct and synergistic contributions to immune activation and tissue damage in lupus. *Immunity* (2013) 38(3):528–40. doi:10.1016/j.immuni.2012.11.017
 134. Nickerson KM, Christensen SR, Shupe J, Kashgarian M, Kim D, Elkon K, et al. TLR9 regulates TLR7- and MyD88-dependent autoantibody production and disease in a murine model of lupus. *J Immunol* (2010) 184(4):1840–8. doi:10.4049/jimmunol.0902592
 135. Karki R, Man SM, Kanneganti TD. Inflammasomes and cancer. *Cancer Immunol Res* (2017) 5(2):94–9. doi:10.1158/2326-6066.CIR-16-0269
 136. Karki R, Man SM, Malireddi RK, Kesavardhana S, Zhu Q, Burton AR, et al. NLR3 is an inhibitory sensor of PI3K-mTOR pathways in cancer. *Nature* (2016) 540:583–87. doi:10.1038/nature20597
 137. Sester DP, Sagulenko V, Thygesen SJ, Cridland JA, Loi YS, Cridland SO, et al. Deficient NLRP3 and AIM2 inflammasome function in autoimmune NZB mice. *J Immunol* (2015) 195(3):1233–41. doi:10.4049/jimmunol.1402859
 138. Lech M, Lorenz G, Kulkarni OP, Grosser MO, Stigrot N, Darisipudi MN, et al. NLRP3 and ASC suppress lupus-like autoimmunity by driving the immunosuppressive effects of TGF-beta receptor signalling. *Ann Rheum Dis* (2015) 74(12):2224–35. doi:10.1136/annrheumdis-2014-205496
 139. Lamprecht M, Bogner S, Schippinger G, Steinbauer K, Fankhauser F, Hallstroem S, et al. Probiotic supplementation affects markers of intestinal barrier, oxidation, and inflammation in trained men; a randomized, double-blinded, placebo-controlled trial. *J Int Soc Sports Nutr* (2012) 9(1):45. doi:10.1186/1550-2783-9-45
 140. Hsiao EY, McBride SW, Hsien S, Sharon G, Hyde ER, McCue T, et al. Microbiota modulate behavioral and physiological abnormalities associated with neurodevelopmental disorders. *Cell* (2013) 155(7):1451–63. doi:10.1016/j.cell.2013.11.024
 141. Round JL, Mazmanian SK. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci U S A* (2010) 107(27):12204–9. doi:10.1073/pnas.0909122107
 142. Telesford KM, Yan W, Ochoa-Reparaz J, Pant A, Kircher C, Christy MA, et al. A commensal symbiotic factor derived from *Bacteroides fragilis* promotes human CD39(+)Foxp3(+) T cells and Treg function. *Gut Microbes* (2015) 6(4):234–42. doi:10.1080/19490976.2015.1056973
 143. Ochoa-Reparaz J, Mielcarz DW, Wang Y, Begum-Haque S, Dasgupta S, Kasper DL, et al. A polysaccharide from the human commensal *Bacteroides fragilis* protects against CNS demyelinating disease. *Mucosal Immunol* (2010) 3(5):487–95. doi:10.1038/mi.2010.29

Conflict of Interest Statement: 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.

Copyright © 2017 Mu, Kirby, Reilly and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Under Pressure: Interactions between Commensal Microbiota and the Teleost Immune System

Cecelia Kelly and Irene Salinas*

Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM, USA

Commensal microorganisms inhabit every mucosal surface of teleost fish. At these surfaces, microorganisms directly and indirectly shape the teleost immune system. This review provides a comprehensive overview of how the microbiota and microbiota-derived products influence both the mucosal and systemic immune system of fish. The cross talk between the microbiota and the teleost immune system shifts significantly under stress or disease scenarios rendering commensals into opportunists or pathogens. Lessons learnt from germ-free fish models as well as from oral administration of live probiotics to fish highlight the vast impact that microbiota have on immune development, antibody production, mucosal homeostasis, and resistance to stress. Future studies should dissect the specific mechanisms by which different members of the fish microbiota and the metabolites they produce interact with pathogens, with other commensals, and with the teleost immune system.

Keywords: microbiota, commensals, teleost, fish, immunity, mucosal immunity, evolution

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, USA

Reviewed by:

Miki Nakao,
Kyushu University, Japan
Jeffrey A. Yoder,
North Carolina State University, USA

*Correspondence:

Irene Salinas
isalinas@unm.edu

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 02 April 2017

Accepted: 26 April 2017

Published: 15 May 2017

Citation:

Kelly C and Salinas I (2017) Under
Pressure: Interactions between
Commensal Microbiota
and the Teleost Immune System.
Front. Immunol. 8:559.
doi: 10.3389/fimmu.2017.00559

INTRODUCTION

Teleost fish are colonized soon after hatching by a diverse set of microbes, which interact with, and shape the development of the host immune system. Like mammals, teleosts have mucosal surfaces, which serve as the first line of defense against invading pathogens, but also harbor non-pathogenic microbes, which makeup the host microbiome. Microbiome studies in a variety of plant and animal species have been accomplished in recent years using rapidly advancing sequencing technologies. These studies continue to expand the knowledge base needed to implement microbiome manipulations with the goal of improving host health. Understanding microbiota-immune system interactions in teleosts is important both for developing solutions to aquacultural problems and for further refinement of fish models, such as the zebrafish (*Danio rerio*), as useful models for biomedical research.

Studies on a number of metazoan hosts have shown that the composition of the microbiota does not merely reflect that of the environment but rather a specific selection of microbial assemblages by hosts has occurred over time (1). Several studies have already determined the bacterial community composition at different teleost mucosal sites (2–5), as well as the presence of a core microbiome in the gut of zebrafish (6). Unfortunately, studies pertaining to mycobionomes and viriomes of fish are lacking. As a consequence, this review only discusses interactions between bacteria and fish immune systems.

As discussed throughout this review, microbiota exert direct effects on the teleost immune system through their display of microbe-associated molecular patterns (MAMPs) and secretion of

factors. Microbiota and their secreted molecules can act locally on the mucosal epithelium or systemically if they enter host circulation or activate immune cells that then travel from mucosal sites to systemic lymphoid tissues. Additionally, these microbes can exert immunostimulatory or immunosuppressive effects on both innate and adaptive immune cells, specific examples of which will be discussed further in this review.

One of the most intimate relationships between microbiota and vertebrate mucosal immune systems is the coevolution between microorganisms and mucosal antibodies (7). Thus, in this review, we will describe in detail, current findings regarding how microbiota shape teleost B cell and antibody responses and how mucosal antibodies and secretory component (SC) allow the host to sculpt its microbial communities. Despite immune exclusion mechanisms present in teleosts, it is clear that certain microbes are capable of reaching and occupying the epithelium of teleosts (2, 8). Similar to mammals, microbial populations vary greatly over the various mucosal body sites of a single fish, with the biggest differences seen between GI tract and external mucosal surfaces (i.e., skin, nose, and gill) bacterial communities (2), suggesting unique and specialized symbiotic relationships at each mucosal site. Conversely, microbial species specific to different mucosal sites may have coevolved alongside the host to perform essential physiological or metabolic duties critical for the optimal functioning of each mucosal immune compartment.

Following the identification of whole microbiome compositions in various fish species, several groups have gained ground in identifying specific microbial species, which are capable of modulating the immune system by colonizing germ-free fish with a single microbe (monocolonization) or a defined group of microbes. These studies have primarily been accomplished using zebrafish (*Danio rerio*) a model for which good germ-free rearing techniques were developed in 2004. Zebrafish are a small, genetically manipulable, and provide the advantage of being transparent during the larval life stage, which makes them a useful model for studying immune system dynamics in response to microbial colonization. Only recently, germ-free seabass (*Dicentrarchus labrax*) have been produced, allowing the study of the interactions between microbiota and a fish host in the seawater environment (9). Future work focused on the identification of candidate microbial species, which can be introduced *via* probiotics or eliminated using antibiotics will be essential to produce treatment plans applicable to improving fish health in aquaculture conditions.

In this review, we will focus on the role of the microbiota in the development and function of the teleost immune system. We will discuss mucosal immune responses at the various tissues that harbor these microbial communities, as well as systemic immune responses, which are regulated by microbiota and their products. We will also review recent studies, which have shed more light on the abilities of individual microbial species to influence the teleost immune system or provide protection from pathogens. Last, we also aim to synthesize known information and create a big picture model showing the different ways microbes and microbial products influence teleost immunity. This model takes into consideration the influence of the environment as

well as other factors that can break the equilibrium between the microbiota and the fish host.

THE IMMUNE SYSTEM OF TELEOST FISH

The immune system of teleost has been studied for decades. Teleost fish have an immune system that resembles that of other jawed vertebrates. The teleost innate immune system provides a first line of defense by detecting and eliminating invading pathogens in an immediate and non-specific manner. Teleost fish also have an adaptive immune system, which relies on somatic recombination of germline-encoded V-D-J fragments to generate a vast repertoire of antigen receptors expressed on the membrane of T and B lymphocytes.

Due to the large number and diversity of teleost species (>30,000), we find unique evolutionary innovations in certain clades. At times, these innovations challenge the current dogma of mammalian immune systems. For instance, the Gadoid family lacks MHC-II expression and CD4 T cell-related molecules. Thus, this teleost group does not rely on traditional antigen presentation *via* the MHC-II and activation of T helper cells to mount adaptive immune responses and instead displays an expansion in the number of MHC-I genes (10, 11).

With regards to the anatomical organization of the teleost immune system, teleosts possess both primary and secondary lymphoid tissues. Primary lymphoid tissues include the thymus, where T cell development occurs, and the head-kidney, which performs hematopoietic functions similar to the mammalian bone marrow. Secondary lymphoid tissues include the spleen and the mucosa-associated lymphoid tissues (MALTs).

Teleost fish have four MALT, the gut-associated lymphoid tissue (GALT), the gill-associated lymphoid tissue (GIALT), the skin-associated lymphoid tissue (SALT), and the nasopharynx-associated lymphoid tissue (NALT) (12). These four MALT share important canonical features that underscore the conserved mechanisms of mucosal immunity in teleost fish (13, 14). Due to the important and direct interactions between commensals and teleost mucosal surfaces, we will describe in further details the organization and functioning of teleost MALT and their components in this review.

A continuously produced mucus layer covers the intestinal, gill, skin, and nasal mucosal surfaces of fish. The teleost mucus layer contains molecules with immunologically important properties, which interact directly with commensal microbial populations at mucosal surfaces. Thus, the composition of the teleost commensal bacteria, fungal, and viral communities is likely shaped by the physicochemical properties of the mucosal secretions. Currently, how the microbiota modulates the amount of mucus secretion as well as the specific composition of the secretions in teleosts is not well understood. While we know that mucosal infections in teleosts can alter the amount of mucus produced as well as the glycosylation levels of mucins (15, 16), how these changes alter the microbiome requires careful investigation.

Generally speaking, teleost MALTs do not contain organized lymphoid structures such as those found in endotherms. Thus, teleost MALTs are composed of a diffuse network of myeloid and lymphoid cells. However, within the GALT, there are some

accumulations of T lymphocytes known as the interbranchial lymphoid tissue (ILT) (17). Although this structure does not present fully organized B and T cell regions and lacks germinal centers, it represents an ancient example of lymphocytic groupings at mucosal surfaces.

In mammals, the microbiota plays a pivotal role in the education of local antigen-presenting cells. The mechanisms of antigen uptake and antigen presentation in teleost MALT are not as well defined as those present in mammalian MALT, but it is clear that teleost MALT have significant numbers of antigen-presenting cells at mucosal sites. Dendritic cells (DCs), macrophages, IgT/Z⁺ B cells, and granulocytes have all been described to uptake antigen in teleost MALT (12, 18, 19). Additionally, enterocytes can uptake antigens by endocytosis (18). Finally, putative M-like cells have been described in the gut of rainbow trout (20). In mammals, luminal sampling DCs can directly sample symbiotic bacteria and transport them to draining lymph nodes (21). Importantly, the presence of the microbiota is required for the establishment of a tolerogenic phenotype in mucosal APCs. To date, the interactions between the microbiota and mucosal APCs of fish have not been investigated.

T cells are the most abundant of all the immune cells present in the MALT of teleost fish. Mucosal T cells include both CD8⁺ and CD4⁺ T cells. Recent reports in zebrafish and trout have shown that CD4⁺ T cells account for 10 and 20% of all T cells in gills and gut (22). However, phenotypic and functional studies on teleost mucosal CD4⁺ T cells are still lacking. CD8⁺ T cells are also present in GALT, GIALT, SALT, and NALT (23–25). Mucosal CD8⁺ T cells appear to have a cytotoxic (CTL) phenotype (12, 23, 25). Compared to systemic CD8 α T cells, mucosal CD8 α T cells also display markers characteristic of mammalian tissue resident memory T cells. Importantly, each teleost MALT harbors unique CD8 α T cell subpopulations, as evidenced by the unique expression of adhesion molecules and receptors in NALT- and GALT-sorted CD8 α T cells. Additionally, trout NALT contains two different populations of CD8 α T cells located in the apical mucosal epithelium and the lateral neuroepithelium, respectively (25). Whether other teleost MALT harbor unique tissue microenvironments containing unique T cell subsets is unknown.

B cells are also part of all teleost MALT and have been fairly well characterized in all four MALT of rainbow trout (13, 14, 26, 27). In sharp contrast to the distribution of B cells in systemic lymphoid tissues, teleost MALT consistently contains a 50/50% distribution of IgM⁺ and IgT⁺ B cells (13, 14, 26, 27). The discovery of IgT as the chief mucosal Ig in teleosts opened up a number of questions regarding the role of this molecule in the maintenance of symbiotic communities in teleost fish. As discussed later, mucosal IgT responses take place in a compartmentalized manner in response to mucosal pathogens. Importantly, commensal bacteria modulate B cells and mucosal Igs.

THE TELEOST FISH MICROBIOME

Although the presence of microbial communities on the mucosal surfaces of teleost fish has been acknowledged for decades, the composition, topography, and environmental factors that shape

teleost bacterial microbiomes have only recently been unveiled thanks to deep sequencing of the 16S rDNA variable region.

Currently, most of the research efforts, which aim to understand the fish microbiota have focused on sequencing bacterial communities from aquacultured species. Microbiome studies from wild fish are also available (28) but less well studied (3, 29). Since phylogeny is a determining factor of the microbial composition of the host (1), and given the large number and taxonomic diversity of extant teleost species, it is likely that new efforts to sequence microbiomes from distantly related teleost species will reveal different assemblages to the ones so far reported. The bacterial communities present at different body sites (2), or under different conditions such as varying host developmental stages (3, 30), different diet regimes (30, 31) or following antibiotic treatment (32), have been sequenced. Importantly, fish also influence the bacterial composition of the tank water as evidenced by two different zebrafish studies (30, 33). Interindividual variation in microbial community composition has been reported in many different fish microbiome studies (2, 33, 34). Stephens et al. showed that in a group of zebrafish siblings raised in the same conditions, the gut microbiota still displays considerable inter-individual variation. This variation can be explained at least in part by neutral processes of drift and dispersal (34). Additionally, ontogenic studies in zebrafish have shown that as the fish age, their gut microbial communities become increasingly different from that of the surrounding environment (30, 33). Whether these changes are also partially controlled by the host immune system is currently unknown. However, as discussed later, inter-host variability in the mucosal Ig repertoire may partially explain bacterial colonization in certain individuals but not others.

Based on sequencing studies from the gut and skin of turbot (35) and trout (2), respectively, it appears that fish are quite permissive in terms of mucosal tissue colonization. In other words, bacteria are not completely excluded from invading epidermal cells and goblet cells (2). This observation may have important consequences when investigating the interactions between the microbiota and the mucosal immune system of fish and further studies are required to understand the nature of this observed “permissiveness.”

Only a few comprehensive functional studies have provided a mechanistic view of the specific interactions that occur between bacterial symbionts and the fish immune system. Based on human microbiome studies, it is clear that microbiota regulates almost every aspect of the host physiology, including the immune response. Based on the seminal study on zebrafish gut responses to microbiota (36), it is tempting to speculate that most of the mechanisms underlying the control of immune systems by the microbiota in mammals may be conserved in teleosts. Undoubtedly, the great taxonomic diversity of fishes as well as their diverse physiological strategies and habitats likely results in very unique adaptations and coevolutionary processes not found in other vertebrate groups.

Whereas 16S rDNA next generation sequencing (NGS) has increased our understanding on bacterial communities of fish, future studies should investigate the archeal, fungal, and viral microbiota of fishes. Moreover, the inter-kingdom interactions between fish viromes, mycobiomes, and bacteriomes remain

unexplored. Similarly, functional studies of fish microbial community composition at different mucosal sites of the same individual require investigation.

GERM-FREE TELEOST MODELS: WHAT HAVE WE LEARNED?

The development of germ-free zebrafish rearing techniques allowed researchers to compare the phenotype of zebrafish larvae, which develop in the absence of the microbiome with that of conventionally reared fish. Due to the laboratory research tools currently available, the majority of zebrafish studies have focused on the interactions between microbiota and the innate immune system. Germ-free zebrafish larvae have impaired neutrophil migration to injury sites (37), decreased larval resistance to viral infection (38), lack expression of innate immune genes, and altered gut epithelial cell turnover (39). Upon colonization with the natural microbiota, zebrafish larvae regain these immune functions. Thus, similar to mammals, teleost immune systems depend on the microbiota for stimulation to maintain a natural state of activity, which benefits the host.

Germ-free larvae can be used to conduct reassociation studies using the natural microbiota, single microbial species, or defined groups of microbes to determine direct effects of microbial presence on the immune system. These types of studies allow the identification of specific bacterial species and their interactions with the host immune system. Pioneer works on zebrafish revealed three main types of responses to specific bacterial colonization at the transcriptional level: innate immune responses, nutrient metabolism, and epithelial cell regeneration (36). Not all species are able to induce all three classes and bacterial products such as lipopolysaccharide (LPS) failed to elicit nutrient metabolism responses (36). Interestingly, germ-free zebrafish mono-associated with *Aeromonas hydrophila* achieve higher induction of *serum amyloid a* expression and similar levels of C3 expression as conventionalized larvae (39). The former result suggests that interactions between different members of the microbiota can serve to balance immunostimulatory effects of a single microbial member, while the latter result shows that single microbial species are sufficient to induce immunostimulatory effects. An elegant study by Rolig demonstrated that while in fish diassociated with *Vibrio* and *Shewanella*, *Vibrio* was the numerically dominant taxa, *Shewanella* presence significantly reduced neutrophil numbers compared to fish mono-associated with *Vibrio* (40). The latter challenges the assumption that the most abundant taxa exert the largest effects on host physiological and immune processes and suggests rarer species in the microbiota can exert potent effects on the immune system. Future studies on the extent of the immunomodulatory power of specific species within the microbiome, and whether these populations are sensitive to manipulation using antibiotics and probiotics will be highly impactful.

Some limitations of the germ-free zebrafish model are lack of known cell markers for immune cells, especially adaptive immune cells, which are not prominent during the early larval stages. Additionally, it is difficult to maintain the germ-free status of larvae past 7 dpf, as the larvae transition from relying on yolk

sac nutrients to eating food. While it is possible, though labor intensive, to maintain a zebrafish under germ-free conditions past this early life stage, no studies have been published using adult germ-free zebrafish. Conversely, a germ-free seabass model that incorporates germ-free feeding of live prey has recently been developed allowing for larvae to survive for at least 16 days post hatching, if not longer (9). Future refinement of the germ-free rearing technique in zebrafish and other teleost species, as well as identification of cell markers, production of reagents, and production of transgenic lines with reporters for or knockouts of important immune genes, will allow for a deeper understanding of the types of systemic immune responses that microbes are capable of inducing during development and adulthood in teleosts.

INTERACTIONS BETWEEN MICROBIOTA AND THE TELEOST MUCOSAL IMMUNE SYSTEM

All fish mucosal sites are colonized by microbes, which interact with both the adaptive and innate immune system. Successful maintenance of immune homeostasis at these sites allow the microbiota to live as an extension of the teleost's own physiology, providing essential functions in nutrient metabolism, maintenance of mucosal barriers, and protection from pathogens. In order to maintain this balance, microbes must either suppress or evade the host immune system, and the host immune system must be calibrated to prevent infection by opportunists, but remain tolerant to a natural number and diversity of microbes, which inhabit various niches in the mucosal microenvironment.

Both innate and adaptive immune pathways regulate bacterial colonization of mucosal surfaces (38, 41). With regards to innate immune pathways, MyD88 signaling appears to be critical (38). Activation of this pathway occurs due to the presence of MAMPs in the microbiota that exert innate immunomodulatory effects. For example, Bates and colleagues demonstrated, in 2007, in zebrafish that detection of LPS can induce intestinal alkaline phosphatase (IAP) expression *via* TLR4 detection and MyD88 signaling. In turn, IAP serves to detoxify LPS and maintain intestinal homeostasis. As mentioned earlier, germ-free teleost models have provided a detailed view of how microbial colonization triggers the transcription of different innate immune genes.

A sizeable fraction of microbes present at trout mucosal surfaces are coated by secreted IgT, IgM, and IgD as well as free SC (13, 14, 26, 27, 42). In mammals, it is generally thought that this coating is a form of immune exclusion, which allows the host to neutralize bacterial adhesion molecules to limit access to the host epithelium. Binding may be mediated by both antigen specific interactions between the Fab region of the antibody and non-specific interactions between glycosylated regions of the SC and antibodies and microbial surface receptors (43). Recently, *Flectobacillus major*-specific IgT titers were recorded in healthy hatchery rainbow trout gill and skin mucus. Interestingly, some fish also had *F. major*-specific IgM titers in plasma. Since both mucosal IgT and systemic IgM titers against this trout commensal strain were low, it was speculated that these antibodies are either

natural antibodies or low-affinity cross-reactive antibodies that recognize common epitopes present in different commensal bacteria (44). Further studies should address whether exposure to commensals elicits compartmentalized Ig responses in mucosal and systemic sites similar to those elicited by pathogens.

Sepahi and Cordero also showed that *F. major*, an abundant microbe at trout mucosal surfaces, produces sphingolipids that induce IgT production in trout gill explants (44). *F. major*-derived sphingolipids injected intravenously into rainbow trout were capable of increasing the systemic IgT to IgM producing B cell ratio. Assuming other members of the microbial community are also producing an array of products, which can interact with immune system receptors, and acknowledging the co-evolution of the teleost immune system alongside the microbiota, it seems likely that the interplay between microbes, their products, and the immune system is highly complex and requires the balance between microbial and host molecules to have the tenacity to rebound to steady state conditions after stresses such as disease and environmental changes are placed on the fish. Future studies regarding the dynamics of how this balance is maintained depends on both the continued exploration of specific host–microbe interactions, as well as building a more accurate big-picture view of host–microbe interactions at mucosal surfaces.

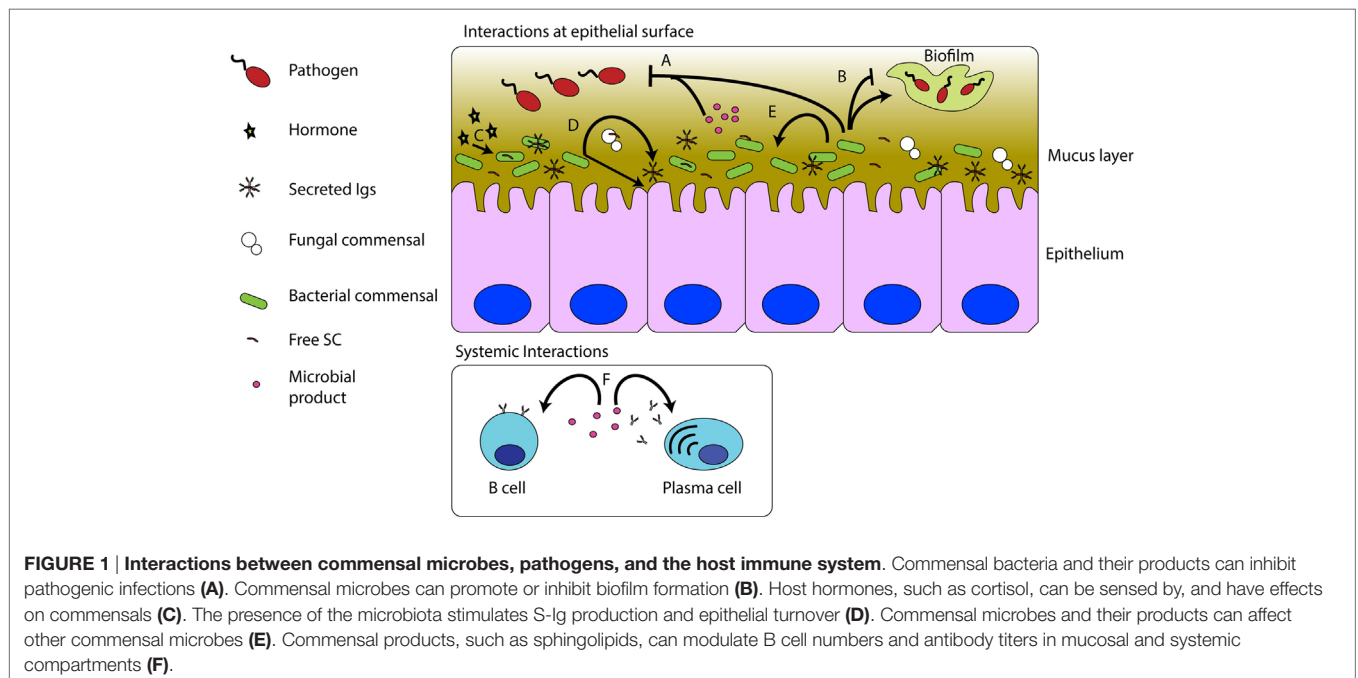
Apart from interactions between B cell/Ig and microbiota, teleost T cells also shape the intestinal microbial composition (45). Adoptive transfer of T cells into Rag1-deficient zebrafish reduces the outgrowth of *Vibrio* sp. The *in vivo* mechanisms behind this inhibitory effect remain unexplored, but T lymphocytes exposed to the microbiota of Rag1-deficient zebrafish *in vitro* produced more IFN γ and TNF α compared to T lymphocytes exposed to the microbiota of wild-type zebrafish, suggesting T cell-mediated inflammatory responses may play a role in shaping the microbiome.

Finally, it is worth highlighting the notion that microbiota contribute to the host's array of immune defenses. Microbial products such as the aforementioned sphingolipids can affect the growth of other symbionts (44) or secrete molecules such as entericidin produced by *Enterobacter* sp., a trout commensal, which directly inhibits pathogen growth (46) in the same manner as host antimicrobial peptides would (Figure 1). On the other hand, when microbiota grows out of control, resident opportunists may favor colonization of pathogens, as demonstrated in the case of the commensal *Staphylococcus warneri* and the pathogen *Vibrio anguillarum* (8).

INTERACTIONS BETWEEN MICROBIOTA AND THE TELEOST SYSTEMIC IMMUNE SYSTEM

Despite the fact that multiple studies have shown that delivery of probiotic bacteria in fish diets can modulate teleost systemic immune responses and disease resistance (47–51), the mechanisms of this interaction remain unknown. As shown in Figure 1, fish commensal bacteria present in the gut mucosa can regulate certain systemic immune parameters. However, there is a clear knowledge gap concerning how these effects are achieved.

Possible indirect interactions between the microbiota and the teleost systemic immune system include production of metabolites such as carbohydrates, aminoacids, or lipids that can be uptaken by gut enterocytes and travel *via* the blood stream to systemic lymphoid tissues such as the HK or the spleen. For instance, PHB produced by *Bacteroides thuringensis* and delivered orally to Nile tilapia increases serum antibodies as well as innate immune parameters (52). However, how the PHB send this message to the systemic immune system is not understood.



Systemic delivery of commensal-derived metabolites has provided some useful insights into the possible mechanisms by which these bacterial products can regulate the fish immune system. For instance, intravenous (i.v.) delivery of *F. major* sphingolipids is able to change IgM and IgT percentages in the HK (44). An overall increase in the proportion of lymphocytes in the HK 72 h after i.v. delivery suggests that this microbial product is able to stimulate B cell proliferation when it reaches systemic circulation. Thus, if the gill and skin of trout is able to extract sphingolipids from *F. major* or *F. major* itself is able to secrete these products and they can enter the bloodstream across the epithelial barriers, then systemic (HK) B cells could directly be controlled by symbiont products.

The contribution of commensal-derived aminoacids, CH, and lipids to the teleost host metabolic composition is unknown. Additionally, we do not know what metabolites commensal communities of fish are capable of producing and how they get secreted and absorbed. This lack of knowledge highlights the fact that implementation of microbiome intervention in aquaculture is still at its infancy.

INTERACTIONS BETWEEN MICROBIOTA AND THE TELEOST IMMUNE SYSTEM DURING STRESS OR DISEASE

Microorganisms interact with each other to form resilient associations in humans (53). The application of microbial ecology concepts to the study of human microbiomes suggests that competitive rather than cooperative interactions between microbes foster the stability of the microbial communities (54). Spatiotemporal changes in the microbial composition of any given community take place during disturbances. In response to perturbations, functionally redundant members may become more abundant aiding in the preservation of community functionality. Environmental disturbances may differentially affect certain mucosal microenvironments. Thus, protected microenvironments could then act as reservoirs for recolonization of the disturbed regions (53). This theoretical framework and modeling has largely been applied to human gut microbiome studies as well as the assembly of the zebrafish microbiome during development. However, how fish microbial assemblages respond to disturbance is less well understood. It is worth noting that adapting this conceptual framework to fish likely needs to consider the greater influence of the environment on aquatic microbial communities compared to their terrestrial counterparts since water is a medium that highly supports microbial growth (Figure 2).

Overall, microbe–microbe interactions, host–microbiota interactions and host–pathogen interactions are complex and poorly understood (55). The dynamics of this triangle under homeostatic conditions require further investigation and may vary between teleost species. Additionally, although it is clear that any changes (i.e., altered microbiota or dysbiosis; altered host status such as stress of ongoing immune responses or altered pathogen loads) will result in loss of homeostasis and an unfavorable outcome for the host (Figure 2), the mechanisms that operate resilience and preservation of fish microbial communities

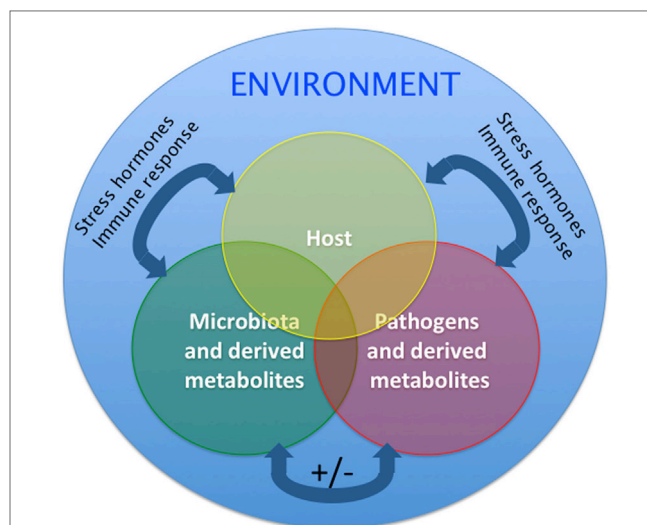


FIGURE 2 | Proposed model of the host–pathogen–microbiota interactions in aquatic animals such as teleost fish. Two-way and three-way interactions among host, pathogen, and microbiota are possible and are overall affected by the environment. Interactions can be positive (synergistic) or negative (inhibitory). Additionally, the host microbial communities and host physiology modify the environment where the fish live. These interactions are, therefore, likely different in laboratory settings, aquaculture settings, and the wild. Homeostatic interactions form a delicate equilibrium. Under stress conditions, host, commensals, and pathogens can produce stress hormones that are molecularly conserved, will be released and alter the interactions of the triangle, likely resulting in decreased immune responses and outgrowth of opportunists and pathogens. Similarly, during the course of immune responses, release of immune molecules from host, commensals, and pathogens will shift the equilibrium of this triangle. The mechanisms by which teleost regain homeostasis following perturbations are largely unknown.

remain poorly understood. Finally, it is very important to bear in mind that these interactions are likely different in a laboratory setting compared to the wild or a fish farm operation (12) as evidenced by the differences in the composition of zebrafish gut microbiomes from different laboratories (6).

A number of studies have shed some light onto the interactions between the microbiota and the teleost immune system during stress responses. For instance, transportation stress results in increased numbers of culturable skin mucus bacteria in rainbow trout. These changes in the skin microbiome were paralleled by change in gene expression of skin mucins, tight junction genes, and anti-inflammatory cytokines (56). Changes in bacterial numbers result in sharp differences in the host mucosal immune response. As suggested by a number of authors, the line between a symbiont and a pathogen is often a blurry one. Symbionts are generally defined as microorganisms that induce anti-inflammatory cytokine expression in the host, whereas pathogens induce pro-inflammatory responses (57, 58). However, even commensals will eventually trigger pro-inflammatory responses in the host if present at high enough numbers. It appears that this paradigm holds true in teleosts, since the commensal bacterium *S. warneri* induces anti-inflammatory cytokines in the skin of rainbow trout when present at low concentrations but pro-inflammatory cytokine expression is upregulated if high concentrations of the

bacterium are achieved. Thus, stress-induced immunosuppression likely allows local bacteria to overgrow.

In a separate study, hypoxic stress was shown to increase the relative abundance of putative pathogenic taxa such as *Psychrobacter*, *Steroidobacter*, *Pseudomonas*, *Acinetobacter*, and *Aeromonas* on trout skin (55). Stress has long been recognized as a key modulator of fish immunity with general immunosuppressive effects (59, 60). Thus, not surprisingly, stress alters teleost microbiomes and results in dysbiosis. However, the mechanisms underlying stress-induced dysbiosis are unknown. Both direct (effects of hormones on the ability of certain bacterial taxa to grow) and indirect (inhibition of host immune responses by glucocorticoids) likely play a role.

We currently know very little about the impact of pathogens on the fish microbiota. One study evaluated the microbiota of wild tropical fish as well as their parasitic loads and found increased diversity of symbionts and lower presence of opportunists in fish that had greater parasitic burdens (61). This study, therefore, reveals a correlation between the presence of parasites and decreased presence of opportunistic pathogens. Whether the immune response of the host against the parasites is playing a role in decreasing opportunistic bacteria requires further investigation. Recently, the commercially important and devastating parasitic copepod *Lepeophtheirus salmonis* was shown to cause major changes in the Atlantic salmon skin microbiome by reducing the alpha diversity and causing destabilization of the microbial community composition (62).

Two reports have given some insights into bacterial diseases and the microbiome of fish (63). The first was conducted in farmed turbot and compared three different farms. This study, although it did not use deep-sequencing of the 16S rDNA, revealed that even healthy fish have a high abundance of bacteria present in internal organs such as the liver and kidney (63). However, the mucosal microbiomes of these fish were not studied and; therefore, it is unknown whether the internal organ microbial communities came from the healthy microbiota reservoir. More recently, the skin mucus microbiome of Atlantic salmon and smallmouth bass (*Micropterus dolomieu*) was studied using plate counts. Bacterial diversity was evaluated over time following natural *Aeromonas salmonicida* outbreaks in the fish farm (64). Despite the obvious limitation of the plate count method, authors concluded that microbial diversity decreased over time due to an over representation of *A. salmonicida* in the community. However,

infection does not always result in losses in overall diversity of the microbiota. For instance, a recent study in laboratory seawater Atlantic salmon found no significant changes in the skin microbiome diversity (alpha diversity) of control and salmon alphavirus-infected fish due to high interindividual variability. However, experimentally infected salmon lost the majority of the proteobacteria and had increased abundances of opportunistic taxa (65). Thus, this study highlights a negative interaction between viral infection and the host–microbiota relationship. In both cases, the contribution of the host immune response to this outcome was not investigated.

CONCLUDING REMARKS

Metazoans draw many benefits from the symbioses with prokaryotes. Unique partnerships have been selected through evolution in order to optimally exploit the metabolic capabilities of microorganisms. Teleost fish include >33,000 different extant species and; therefore, this diversity must be matched by a great diversity of selected microbial assemblages, which inhabit every fish mucosal barrier. Due to the conduciveness of the aquatic environment for microbial growth (66), it appears that minute changes in the host immune status can trigger states of dysbiosis. How teleost fish cope with these perturbations and how the microbial communities regain homeostasis is not fully understood. The complexity of the interactions between the environment, the teleost immune system, and the microbiota can now be dissected; thanks to NGS techniques, germ-free models, mono-association studies, and infection models. In the future, bacterial metagenomics and transcriptomic studies would be beneficial to advance our understanding of the functionality of fish microbiomes and their partnership with the fish immune system.

AUTHOR CONTRIBUTIONS

IS conceptually designed the paper and wrote the paper. CK wrote the paper and made figures.

FUNDING

This work was funded by NIH grants 2R01GM085207-05. CK was funded by the Stephanie Ruby fellowship.

REFERENCES

- Brooks AW, Kohl KD, Brucker RM, van Opstal EJ, Bordenstein SR. Phyllosymbiosis: relationships and functional effects of microbial communities across host evolutionary history. *PLoS Biol* (2016) 15(1):e1002587. doi:10.1371/journal.pbio.2000225
- Lowrey L, Woodhams DC, Tacchi L, Salinas I. Topographical mapping of the rainbow trout (*Oncorhynchus mykiss*) microbiome reveals a diverse bacterial community with antifungal properties in the skin. *Appl Environ Microbiol* (2015) 81(19):6915–25. doi:10.1128/AEM.01826-15
- Llewellyn MS, McGinnity P, Dionne M, Letourneau F, Thonier F, Carvalho GR, et al. The biogeography of the Atlantic salmon (*Salmo salar*) gut microbiome. *ISME J* (2015) 10:1280–4. doi:10.1038/ismej.2015.189
- Gajardo K, Rodiles A, Kortner TM, Kroghdahl A, Bakke AM, Merrifield DL, et al. A high-resolution map of the gut microbiota in Atlantic salmon (*Salmo salar*): a basis for comparative gut microbial research. *Sci Rep* (2016) 6:30894. doi:10.1038/srep30893
- Larsen AM, Bullard SA, Womble M, Arias CR. Community structure of skin microbiome of gulf killifish, *Fundulus grandis*, is driven by seasonality and not exposure to oiled sediments in a Louisiana salt marsh. *Microb Ecol* (2015) 70(2):534–44. doi:10.1007/s00248-015-0578-7
- Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, et al. Evidence for a core gut microbiota in the zebrafish. *ISME J* (2011) 5:1595–608. doi:10.1038/ismej.2011.38
- Kaetzel CS. Coevolution of mucosal immunoglobulins and the polymeric immunoglobulin receptor: evidence that the commensal microbiota provided the driving force. *ISRN Immunol* (2014) 2014:1–20. doi:10.1155/2014/541537
- Musharrafieh R, Tacchi L, Trujeque J, LaPatra S, Salinas I. *Staphylococcus warneri*, a resident skin commensal of rainbow trout (*Oncorhynchus mykiss*) with pathobiont characteristics. *Vet Microbiol* (2014) 169(1–2):80–8. doi:10.1016/j.vetmic.2013.12.012

9. Schaeck M, De Swaef E, Van den Broeck W, Van Nevel S, Boon N, De Geyter N, et al. Germ-free sea bass *Dicentrarchus labrax* larval model: a valuable tool in the study of host-microbe interactions. *Dis Aquat Organ* (2016) 117(3):177–85. doi:10.3354/dao02950
10. Malmstrom M, Jentoft S, Gregers TF, Jakobsen KS. Unraveling the evolution of the Atlantic cod's (*Gadus morhua* L.) alternative immune strategy. *PLoS One* (2013) 8(10):e74004. doi:10.1371/journal.pone.0074004
11. Malmstrom M, Matschiner M, Torresen OK, Star B, Snipen LG, Hansen TF, et al. Evolution of the immune system influences speciation rates in teleost fishes. *Nat Genet* (2016) 48:1204–10. doi:10.1038/ng.3645
12. Salinas I. The mucosal immune system of teleost fish. *Biology (Basel)* (2015) 4(3):525–39. doi:10.3390/biology4030525
13. Xu Z, Takizawa F, Parra D, Gomez D, von Gersdorff Jorgensen L, LaPatra SE, et al. Mucosal immunoglobulins at respiratory surfaces mark an ancient association that predates the emergence of tetrapods. *Nat Commun* (2016) 7:10728. doi:10.1038/ncomms10728
14. Tacchi L, Musharrafieh R, Larragoite ET, Crossey K, Erhardt EB, Martin SAM, et al. Nasal immunity is an ancient arm of the mucosal immune system. *Nat Commun* (2014) 5:5205. doi:10.1038/ncomms6205
15. Van der Marel M, Caspari N, Neuhaus H, Meyer W, Enss M, Steinhagen D. Changes in skin mucus of common carp, *Cyprinus carpio* L., after exposure to water with a high bacterial load. *J Fish Dis* (2010) 33:431–9. doi:10.1111/j.1365-2761.2010.01140.x
16. Estensoro I, Jung-Schroers V, Álvarez-Pellitero P, Steinhagen D, Sitjà-Bobadilla A. Effects of *Enteromyxum leei* (Myxozoa) infection on gilthead sea bream (*Sparus aurata*) (Teleostei) intestinal mucus: glycoprotein profile and bacterial adhesion. *Parasitol Res* (2013) 112:567–76. doi:10.1007/s00436-012-3168-3
17. Haugarvoll E, Bjerkaas I, Nowak BF, Hordvik I, Koppang EO. Identification and characterization of a novel intraepithelial lymphoid tissue in the gills of Atlantic salmon. *J Anat* (2008) 213(2):202–9. doi:10.1111/j.1469-7580.2008.00943.x
18. Rombout JH, Abelli L, Picchietti S, Scapigliati G, Kiron V. Teleost intestinal immunology. *Fish Shellfish Immunol* (2011) 31(5):616–26. doi:10.1016/j.fsi.2010.09.001
19. Resseguier J, Delaune E, Coolen AL, Levraud JB, Boudinot P, Le Guellec D, et al. Specific and efficient uptake of surfactant-free poly(lactic acid) nanovaccine vehicles by mucosal dendritic cells in adult zebrafish after bath immersion. *Front Immunol* (2017) 8:190. doi:10.3389/fimmu.2017.00190
20. Fuglem B, Jirillo E, Bjerkaas I, Kiyono H, Nochi T, Yuki Y, et al. Antigen-sampling cells in the salmonid intestinal epithelium. *Dev Comp Immunol* (2010) 34(7):768–74. doi:10.1016/j.dci.2010.02.007
21. Rescigno M, Urbano M, Valzasina B, Francolini M, Rotta G, Bonasio R, et al. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nat Immunol* (2001) 2:361–7. doi:10.1038/86373
22. Dee CT, Nagaraju RT, Athanasiadis EI, Gray C, del Alma LF, Johnston SA, et al. CD4-transgenic zebrafish reveal tissue-resident Th2 and regulatory T cell-like populations and diverse mononuclear phagocytes. *J Immunol* (2016) 197(9):3520–30. doi:10.4049/jimmunol.1600959
23. Takizawa F, Araki K, Kobayashi I, Moritomo T, Ototake M, Nakanishi T. Molecular cloning and expression analysis of T-bet in ginbuna crucian carp (*Carassius auratus langsdorffii*). *Mol Immunol* (2008) 45(1):127–36. doi:10.1016/j.molimm.2007.05.005
24. Leal E, Granja AG, Zarza C, Tafalla C. Distribution of T cells in rainbow trout (*Oncorhynchus mykiss*) skin and responsiveness to viral infection. *PLoS One* (2016) 11(1):e0147477. doi:10.1371/journal.pone.0147477
25. Sepahi A, Casadei E, Tacchi L, Munoz P, LaPatra SE, Salinas I. Tissue micro-environments in the nasal epithelium of rainbow trout (*Oncorhynchus mykiss*) define two distinct CD8 α cell populations and establish regional immunity. *J Immunol* (2016) 197(11):4453–63. doi:10.4049/jimmunol.1600678
26. Zhang YA, Salinas I, Li J, Parra D, Bjork S, Xu Z, et al. IgT, a primitive immunoglobulin class specialized in mucosal immunity. *Nat Immunol* (2010) 11(9):827–35. doi:10.1038/ni.1913
27. Xu Z, Parra D, Gomez D, Salinas I, Zhang YA, von Gersdorff Jorgensen L, et al. Teleost skin, an ancient mucosal surface that elicits gut-like immune responses. *Proc Natl Acad Sci U S A* (2013) 110(32):13097–102. doi:10.1073/pnas.1304319110
28. Tarnecki AM, Patterson WF III, Arias CR. Microbiota of wild-caught red snapper *Lutjanus campechanus*. *BMC Microbiol* (2016) 16(1):245. doi:10.1186/s12866-016-0864-7
29. Sime-Ngando T, Lafferty KD, Biron DG. Roles and mechanisms of parasitism in aquatic microbial communities. *Front Microbiol* (2015) 6:446. doi:10.3389/fmicb.2015.00446
30. Wong S, Stephens WZ, Burns AR, Stagaman K, David LA, Bohannan BJM, et al. Ontogenic differences in dietary fat influence microbiota assembly in the zebrafish gut. *mBio* (2015) 6(5):e687–615. doi:10.1128/mBio.00687-15
31. Xia JH, Lin G, Fu GH, Wan ZY, Lee M, Wang L, et al. The intestinal microbiome of fish under starvation. *BMC Genomics* (2014) 15:266. doi:10.1186/1471-2164-15-266
32. Gaulke CA, Barton CL, Proffitt S, Tanguay RL, Sharpton TJ. Triclosan exposure is associated with rapid restructuring of the microbiome in adult zebrafish. *PLoS One* (2016) 11(5):e0154632. doi:10.1371/journal.pone.0154632
33. Stephens ZW, Burns AR, Stagaman K, Wong S, Rawls JF, Guillemin K, et al. The composition of the zebrafish intestinal microbial community varies across development. *ISME J* (2016) 10:644–54. doi:10.1038/ismej.2015.140
34. Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, et al. Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* (2016) 10(3):655–64. doi:10.1038/ismej.2015.142
35. Xing M, Hou Z, Yuan J, Liu Y, Qu Y, Liu B. Taxonomic and functional metagenomics profiling of gastrointestinal tract microbiome of the farmed adult turbot (*Scophthalmus maximus*). *FEMS Microbiol Ecol* (2013) 86(3):432–43. doi:10.1111/1574-6941.12174
36. Rawls JF, Mahowald MA, Ley RE, Gordon JL. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* (2006) 127(2):423–33. doi:10.1016/j.cell.2006.08.043
37. Kanther M, Tomkovich S, Xiaolun S, Grosser MR, Koo J, Flynn EJ III, et al. Commensal microbiota stimulate systemic neutrophil migration through induction of serum amyloid A. *Cell Microbiol* (2014) 16(7):1053–67. doi:10.1111/cmi.12257
38. Gallindo-Villegas J, Garcia-Moreno D, Oliveira S, Meseguer J, Mulero V. Regulation of immunity and disease resistance by commensal microbes and chromatin modifications during zebrafish development. *Proc Natl Acad Sci U S A* (2012) 109(39):E2605–14. doi:10.1073/pnas.1209920109
39. Rawls JF, Samuel BS, Gordon JL. Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proc Natl Acad Sci U S A* (2004) 101(13):4596–601. doi:10.1073/pnas.0400706101
40. Rolig AS, Parthasarathy R, Burns AR, Bohannan BJM, Guillemin K. Individual members of the microbiota disproportionately modulate host innate immune responses. *Cell Host Microbe* (2015) 18(5):613–20. doi:10.1016/j.chom.2015.10.009
41. Bates JM, Akerlund J, Mittge E, Guillemin K. Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota. *Cell Host Microbe* (2007) 2(6):371–82. doi:10.1016/j.chom.2007.10.010
42. Kelly C, Takizawa F, Sunyer JO, Salinas I. Rainbow trout (*Oncorhynchus mykiss*) secretory component binds to commensal bacteria and pathogens. *Sci Rep* (2017) 7:41753. doi:10.1038/srep41753
43. Mantis NJ, Rol N, Corthesy B. Secretory IgA's complex roles in immunity and mucosal homeostasis in the gut. *Mucosal Immunol* (2013) 4(6):603–11. doi:10.1038/mi.2011.41
44. Sepahi A, Cordero H, Goldfine H, Esteban MA, Salinas I. Symbiont-derived sphingolipids modulate mucosal homeostasis and B cells in teleost fish. *Sci Rep* (2016) 6:39054. doi:10.1038/srep39054
45. Brugman S, Schneeberger K, Wittle M, Klein MR, van den Bogert B, Boekhorst J, et al. T lymphocytes control microbial composition by regulating the abundance of *Vibrio* in the zebrafish gut. *Gut Microbes* (2014) 5(6):737–47. doi:10.4161/19490976.2014.972228
46. Schubiger CB, Orfe LH, Sudheesh PS, Cain KD, Shah DH, Call DR. Entericidin is required for a probiotic treatment (*Enterobacter* sp. strain C6-6) to protect trout from cold-water disease challenge. *Appl Environ Microbiol* (2015) 81(2):658–65. doi:10.1128/AEM.02965-14
47. Salinas I, Cuesta A, Esteban MA, Meseguer J. Dietary administration of *Lactobacillus delbrueckii* and *Bacillus subtilis*, single or combine, on gilthead seabream cellular innate immune responses. *Fish Shellfish Immunol* (2005) 19(1):67–77. doi:10.1016/j.fsi.2004.11.007
48. Salinas I, Abelli L, Bertoni F, Picchietti S, Roque A, Furones D, et al. Monospecies and multispecies probiotic formulations produce different systemic and local immunostimulatory effects in the gilthead seabream (*Sparus*

- aurata* L.). *Fish Shellfish Immunol* (2008) 25(1–2):114–23. doi:10.1016/j.fsi.2008.03.011
49. Kumar R, Mukherjee SC, Ranjan R, Nayak SK. Enhanced innate immune parameters in *Labeo rohita* (Ham.) following oral administration of *Bacillus subtilis*. *Fish Shellfish Immunol* (2008) 24(2):168–72. doi:10.1016/j.fsi.2007.10.008
 50. Son VM, Chang CC, Wu MC, Guu YK, Chiu CH, Cheng W. Dietary administration of the probiotic, *Lactobacillus plantarum*, enhanced growth, innate immune responses, and disease resistance of the grouper *Epinephelus coioides*. *Fish Shellfish Immunol* (2009) 26(5):691–8. doi:10.1016/j.fsi.2009.02.018
 51. Liu CH, Chiu CH, Wang SW, Cheng W. Dietary administration of the probiotic, *Bacillus subtilis* E20, enhances the growth, innate immune responses, and disease resistance of the grouper, *Epinephelus coioides*. *Fish Shellfish Immunol* (2012) 33(4):699–706. doi:10.1016/j.fsi.2012.06.012
 52. Suguna P, Binuramesh C, Abirami P, Saranya V, Poornima K, Rajeswari V, et al. Immunostimulation by poly- β hydroxybutyrate-hydroxyvalerate (PHB-HV) from *Bacillus thuringiensis* in *Oreochromis mossambicus*. *Fish Shellfish Immunol* (2014) 36(1):90–7. doi:10.1016/j.fsi.2013.10.012
 53. Van den Abbeele P, Van de Wiele T, Verstraete W, Possemiers S. The host selects mucosal and luminal associations of coevolved gut microorganisms: a novel concept. *FEMS Microbiol Rev* (2011) 35(4):861–704. doi:10.1111/j.1574-6976.2011.00270.x
 54. Coyte KZ, Schluter J, Foster KR. The ecology of the microbiome: networks, competition, and stability. *Science* (2015) 350(6261):663–6. doi:10.1126/science.aad2606
 55. Boutin S, Bernatchez L, Audet C, Derome N. Network analysis highlights complex interactions between pathogen, host and commensal microbiota. *PLoS One* (2013). doi:10.1371/journal.pone.0084772
 56. Tacchi L, Lowrey L, Musharrafieh R, Crossey K, Larragoite ET, Salinas I. Effects of transportation stress and addition of salt to transport water on the skin mucosal homeostasis of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* (2015) 435:120–7. doi:10.1016/j.aquaculture.2014.09.027
 57. Nussbaum JC, Locksley RM. Infectious (non)tolerance-frustrated commensalism gone awry? *Cold Spring Harb Perspect Biol* (2012) 4(5). doi:10.1101/cshperspect.a007328
 58. Kaci G, Goudercourt D, Dennin V, Pot B, Dore J, Ehrlich SD, et al. Anti-inflammatory properties of *Streptococcus salivarius*, a commensal bacterium of the oral cavity and digestive tract. *Appl Environ Microbiol* (2014) 80(3):928–34. doi:10.1128/AEM.03133-13
 59. Pickering AD, Pottinger TG. Stress responses and disease resistance in salmonid fish: effects of chronic elevation of plasma cortisol. *Fish Physiol Biochem* (1989) 7(1–6):253–8. doi:10.1007/BF00004714
 60. Parra D, Reyes-Lopez FE, Tort L. Mucosal immunity and B cells in teleosts: effect of vaccination and stress. *Front Immunol* (2015) 6:354. doi:10.3389/fimmu.2015.00354
 61. Hennersdorf P, Kleinertz S, Theisen S, Abdul-Aziz MA, Mrotzek G, Palm HW, et al. Microbial diversity and parasitic load in tropical fish of different environmental conditions. *PLoS One* (2016) 11(3):e0151594. doi:10.1371/journal.pone.0151594
 62. Llewellyn MS, Leadbeater S, Garcia C, Sylvian FE, Custodio M, Ang KP, et al. Parasitism perturbs the mucosal microbiome of Atlantic salmon. *Sci Rep* (2017) 7:43465. doi:10.1038/srep43465
 63. Toranzo AE, Novoa B, Romalde JL, Nunez S, Devesa S, Marino E, et al. Microflora associated with healthy and diseased turbot (*Scophthalmus maximus*) from three farms in Northwest Spain. *Aquaculture* (1993) 114:189–202. doi:10.1016/0044-8486(93)90295-A
 64. Cipriano RC, Dove A. Far from superficial: microbial diversity associated with the dermal mucus of fish. In: Cipriano RC, Schelkunov I, editors. *Health and Diseases of Aquatic Organisms: Bilateral Perspectives*. East Lansing: MSU Press (2011). p. 156–67.
 65. Reid KM, Patel S, Robinson AJ, Bu L, Jarungsriapisit J, Moore LJ, et al. Salmonid alphavirus infection causes skin dysbiosis in Atlantic salmon (*Salmo salar* L.) post-smolts. *PLoS ONE* (2017) 12(3):e0172856. doi:10.1371/journal.pone.0172856
 66. Salinas I, Parra D. Fish mucosal immunity: intestine. In: Beck BH, Peatmen E, editors. *Mucosal Health in Aquaculture*. Massachusetts, USA: Academic Press (2015). p. 135–58.

Conflict of Interest Statement: 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.

Copyright © 2017 Kelly and Salinas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida
St. Petersburg, USA

Reviewed by:

Yuko Ota,
University of Maryland Baltimore, USA

Tony De Tomaso,
University of California
Santa Barbara, USA

*Correspondence:

L. Courtney Smith
csmith@gwu.edu

[†]Present address:

Cheng Man Lun,
HIV Dynamics and Replication
Program, Virus-Cell Interaction
Section, Center for Cancer Research,
National Cancer Institute,
Frederick, MD, USA;
Robin L. Samuel,
MedImmune, Frederick, MD, USA;
Susan D. Gillmor,
National Institutes of Health,
Center for Scientific Review,
Bethesda, MD, USA;
Anthony Boyd,
College of Optometry, State
University of New York,
New York, NY, USA

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 03 February 2017

Accepted: 06 April 2017

Published: 12 May 2017

Citation:

Lun CM, Samuel RL, Gillmor SD,
Boyd A and Smith LC (2017) The
Recombinant Sea Urchin Immune
Effector Protein, rSpTransformer-E1,
Binds to Phosphatidic Acid and
Deforms Membranes.
Front. Immunol. 8:481.
doi: 10.3389/fimmu.2017.00481

The Recombinant Sea Urchin Immune Effector Protein, rSpTransformer-E1, Binds to Phosphatidic Acid and Deforms Membranes

Cheng Man Lun^{1†}, Robin L. Samuel^{2†}, Susan D. Gillmor^{2†}, Anthony Boyd^{1†} and L. Courtney Smith^{1*}

¹ Department of Biological Sciences, George Washington University, Science and Engineering Hall, Washington, DC, USA,

² Department of Chemistry, George Washington University, Science and Engineering Hall, Washington, DC, USA

The purple sea urchin, *Strongylocentrotus purpuratus*, possesses a sophisticated innate immune system that functions without adaptive capabilities and responds to pathogens effectively by expressing the highly diverse *SpTransformer* gene family (formerly the *Sp185/333* gene family). The swift gene expression response and the sequence diversity of *SpTransformer* cDNAs suggest that the encoded proteins have immune functions. Individual sea urchins can express up to 260 distinct SpTransformer proteins, and their diversity suggests that different versions may have different functions. Although the deduced proteins are diverse, they share an overall structure of a hydrophobic leader, a glycine-rich N-terminal region, a histidine-rich region, and a C-terminal region. Circular dichroism analysis of a recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1) demonstrates that it is intrinsically disordered and transforms to α helical in the presence of buffer additives and binding targets. Although native SpTrf proteins are associated with the membranes of perinuclear vesicles in the phagocyte class of coelomocytes and are present on the surface of small phagocytes, they have no predicted transmembrane region or conserved site for glycoposphatidylinositol linkage. To determine whether native SpTrf proteins associate with phagocyte membranes through interactions with lipids, when rSpTrf-E1 is incubated with lipid-embedded nylon strips, it binds to phosphatidic acid (PA) through both the glycine-rich region and the histidine-rich

Abbreviations: SpTrf, transformer genes from the sea urchin *Strongylocentrotus purpuratus*; SpTrf, transformer proteins from the sea urchin *Strongylocentrotus purpuratus*; HeTrf, Trf proteins from the sea urchin species *Heliocidaris erythrogramma*; natSpTrf, native SpTrf proteins; Ni-natSpTrf, nickel-isolated natSpTrf proteins; rSpTrf-E1, recombinant transformer protein with an E1 element pattern; rSpTrf-E1-FITC, biotinylated rSpTrf-E1 labeled with NeutrAvidin fluorescein; rSpTrf-E1-2PA, rSpTrf-E1 bound to two PA lipids; Gly-rich region, the glycine-rich region that is located at the N-terminus of the mature natSpTrf proteins; rGly-rich fragment, the recombinant glycine-rich fragment of rSpTrf-E1; His-rich region, the histidine-rich region that is located near the C-terminus of the mature natSpTrf proteins; rHis-rich fragment, the recombinant Histidine-rich fragment of rSpTrf-E1; rC-Gly, the recombinant fragment that is located at the C-terminal end of the glycine-rich region of rSpTrf-E1; BSA-FITC, biotinylated bovine serum albumin labeled with NeutrAvidin fluorescein; SUV, small unilamellar vesicle; GUV, giant unilamellar vesicle; LUV, large unilamellar vesicle; dextran-488, dextran labeled with Alexa Fluor® 488; PA, phosphatidic acid; NBD-PA, 1-oleoyl-2-(6-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]hexanoyl)-sn-glycero-3-phosphate; PC, phosphatidylcholine; TFE, trifluoroethanol; IDP, intrinsically disordered protein; IDR, intrinsically disordered region.

region. Synthetic liposomes composed of PA and phosphatidylcholine show binding between rSpTrf-E1 and PA by fluorescence resonance energy transfer, which is associated with leakage of luminal contents suggesting changes in lipid organization and perhaps liposome lysis. Interactions with liposomes also change membrane curvature leading to liposome budding, fusion, and invagination, which is associated with PA clustering induced by rSpTrf-E1 binding. Longer incubations result in the extraction of PA from the liposomes, which form disorganized clusters. CD shows that when rSpTrf-E1 binds to PA, it changes its secondary structure from disordered to α helical. These results provide evidence for how SpTransformer proteins may associate with molecules that have exposed phosphates including PA on cell membranes and how the characteristic of protein multimerization may drive changes in the organization of membrane lipids.

Keywords: Sp185/333, echinoderm, innate immunity, conformational plasticity, liposomes, lipid clusters

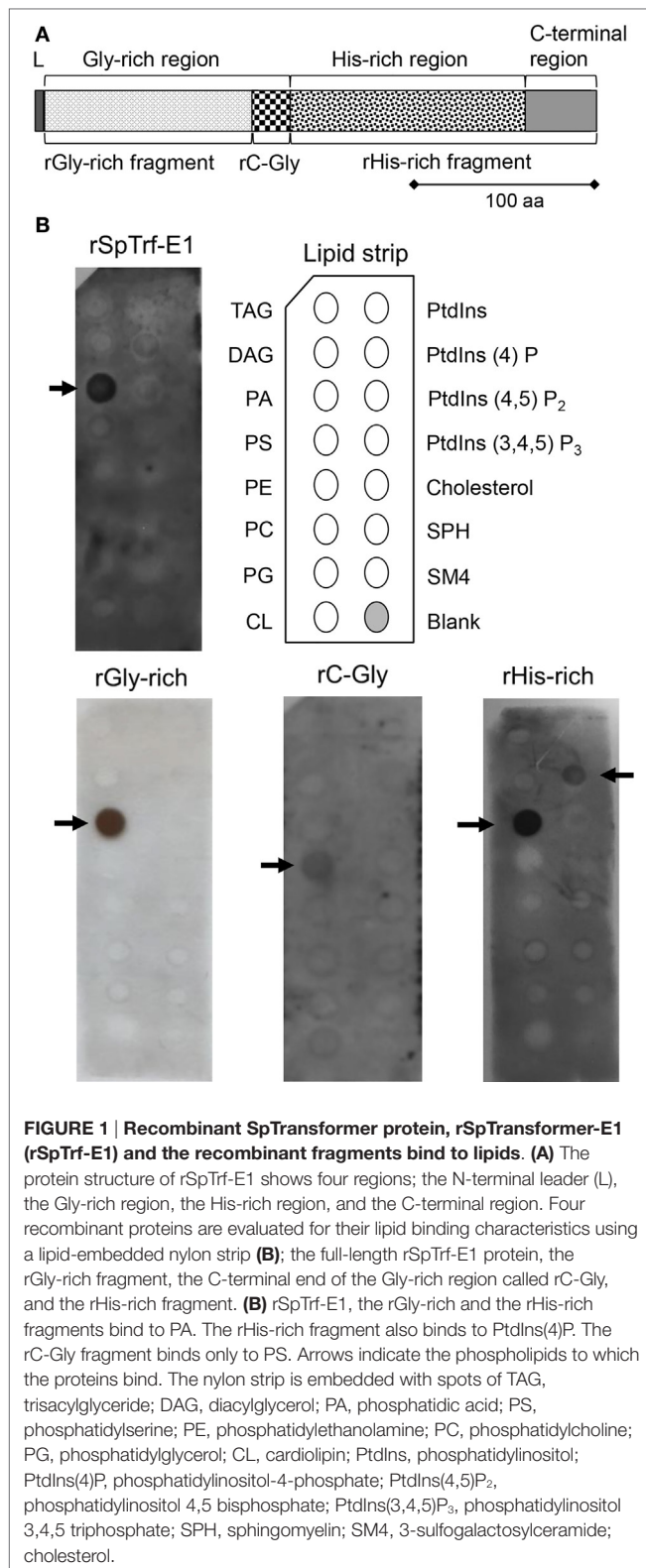
INTRODUCTION

The genome of the California purple sea urchin (*Strongylocentrotus purpuratus*) has a number of large immune gene families that are quite complex (1–3). One of the families, *SpTransformer* (*SpTrf*, formerly *Sp185/333*), is unique to sea urchin species that are members of the euechinoidea subclass and show no homology to gene families in other organisms including the cidaroidea subclass of echinoids. The *SpTrf* gene family has been estimated to have ~50 members, and the genes have two exons that encode the leader and the mature protein (4–9) that respond with swift increases in expression upon challenges from microbes and pathogen-associated molecular patterns (PAMPs) (4, 10, 11). Alignments of genes and transcripts require the insertion of artificial gaps in the second exon that defines 25–27 blocks of sequences known as *elements* (5). The presence and absence of elements create different mosaics of elements that are repeatedly identified and called *element patterns* [(4, 5, 11), reviewed in Ref. (8)]. Despite the sequence diversity of the genes and transcripts, the encoded proteins have a generic structure that is composed of an N-terminal leader, a glycine-rich (Gly-rich) region with an arginine–glycine–aspartic acid motif, a histidine-rich (His-rich) region, and a C-terminal region (**Figure 1A**) (4, 12). The diversity of element patterns, plus putative editing of the *SpTrf*mRNAs (13) that introduces missense sequence and early stop codons produces a wide range of deduced SpTrf proteins of 4–55 kDa (4, 11). An evaluation of the native (nat)SpTrf proteins in one sea urchin suggests that it can express up to 260 different variants and that the native proteins appear unexpectedly large on Western blots relative to the deduced protein size predictions (14–16). Increases in size are likely the result of multimerization of natSpTrf proteins that is induced by isolation and processing, which can also be induced for a recombinant (r)SpTrf-E1 protein (originally called rSp0032) after isolation from *E. coli* and in the absence of other sea urchin proteins (12, 14). Once multimerized, SpTrf proteins, whether native or recombinant, cannot be separated to monomers [see supplemental materials in Ref. (16)] and once bound cannot be dissociated from the marine bacteria, *Vibrio diazotrophicus*, or Baker's yeast, *Saccharomyces cerevisiae* (12). Furthermore, recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), binds tightly to lipopolysaccharide

(LPS), β -1,3-glucan, and flagellin and once bound cannot be dissociated from one PAMP for rebinding to another.

The deduced amino acid composition of the SpTrf proteins indicates two major regions within the proteins, the Gly-rich and His-rich regions, that have been predicted to be functionally different (4, 11). The first functional evaluation of rSpTrf-E1 plus three discrete recombinant fragments, the rGly-rich fragment, the C-terminal end of the Gly-rich region called the rC-Gly, and the rHis-rich fragment (**Figure 1A**) demonstrated that rSpTrf-E1 has restricted binding to a subset of bacterial species (12). However, the recombinant fragments all show expanded bacterial binding suggesting differences in the activities of the separated fragments. The rC-Gly fragment is consistently multimerized upon isolation and likely mediates multimerization of full-length natSpTrf proteins. In the absence of the rC-Gly fragment, neither the rGly-rich nor the rHis-rich fragments show multimerization either upon isolation, after storage, or upon binding targets. When using yeast as a binding target, rSpTrf-E1 partially competes with the rGly-rich fragment and fully competes with the rHis-rich fragment, indicating that both ends of the protein bind to yeast, but that the rGly-rich fragment has expanded binding activities. This result is noteworthy because mRNA editing that tends to occur prior to immune challenge (16) produces truncated proteins that consist only of the Gly-rich region in which the expanded binding activity may act in immune surveillance in the sea urchin (12).

Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), is an intrinsically disordered protein (IDP) in which monomers undergo secondary structural transformation from disordered to 78–95% α helical in sodium dodecyl sulfate (SDS), trifluoroethanol (TFE), or LPS (17). This secondary structural transformation is the basis for the new name of SpTrf proteins and rSpTrf-E1 has an E1 element pattern, hence the name extension. For details on element patterns and naming convention, see Ref. (4, 5, 9, 11, 18). Based on the overall structural similarities among the SpTrf proteins and bioinformatic predictions that all may be IDPs as suggested by amino acid sequences, we have speculated that many may have similar transforming capabilities and show structural changes that are induced by binding targets. The rHis-rich and rGly-rich fragments also show changes in secondary structural conformation; however, the changes are unexpected relative to results for rSpTrf-E1. The rGly-rich and



rHis-rich fragments are 15–30% α helical in phosphate buffer rather than disordered like rSpTrf-E1. In the presence of SDS, TFE, or LPS, the fragments either enhance their α helical structure or switch to β strand structure (17). These results suggest

that the Gly-rich and His-rich regions within rSpTrf-E1 likely interact and influence both the specificity of the target to which they bind and the subsequent folding upon binding to a target.

Native SpTrf proteins are present within the perinuclear vesicles of all types of phagocytes and are present on the surface of small phagocytes, although the percentage of cells that express the proteins is variable among animals (14, 19). HeTransformer proteins (HeTrf, formerly He185/333) have also been noted in association with vesicle membranes and with plasma membranes of gut-associated amoebocytes (an alternative term for phagocytes) from another sea urchin species, *Heliocidaris erythrogramma* (20, 21). Yet, the association of Trf proteins from both sea urchin species with cell membranes is not predicted from their deduced amino acid sequences; there are no recognizable transmembrane regions and no predicted conserved motifs for glycosylphosphatidylinositol linkages (4, 14, 21). To understand this association, we investigated possible interactions of rSpTrf-E1 and the recombinant fragments with phospholipids and identified specific binding to phosphatidic acid (PA) by the full-length protein and the rGly-rich and rHis-rich fragments. In addition, the rHis-rich fragment also binds to phosphatidyl inositol 4 phosphate [PtnIns(4)P], although with lower affinity. rSpTrf-E1 binds to liposomes that are composed of 10% PA and 90% phosphatidylcholine (10% PA:PC) and transforms from disordered to ~70% α helical in the presence of PA. rSpTrf-E1 induces changes in membrane curvature of 10% PA:PC liposomes, which show budding, invagination, and fusion that is associated with PA clustering. rSpTrf-E1 induces leakage of materials captured within liposome lumens, and longer incubations with 10% PA:PC liposomes result in PA extraction from the membranes. We speculate that accessible phosphate groups may be a binding target and that this may be a mechanism by which a subset of natSpTrf proteins may interact with coelomocyte membranes in sea urchins and perhaps may aid in initiating membrane curvature through PA clustering leading to phagocytosis of bacteria.

MATERIALS AND METHODS

Expression, Isolation, and Purification of natSpTrf proteins, rSpTrf-E1, and the Recombinant Fragments

The expression, isolation, and purification of rSpTrf-E1, and the three recombinant fragments from *E. coli* were performed as described (12). Nickel affinity was used to isolate natSpTrf proteins (Ni-natSpTrf) according to Sherman et al. (16) with an additional step using anti-SpTrf (formerly anti-Sp185/333) antibodies linked in an affinity column according to Lun et al. (12). Following isolation by Ni-affinity, Ni-natSpTrf, rSpTrf-E1, and the recombinant fragments were verified by analysis of flow-through and elution fractions on Any KD™ Mini-PROTEAN® TGX precast gels (Bio-Rad Laboratories, Inc.) that were electrophoresed for 20 min at 300 V and constant voltage. Two precast gels were run simultaneously; one was processed for Western blot evaluation with anti-SpTrf antibodies, and the other was stained with Biosafe Coomassie stain (Bio-Rad Laboratories) as described (12).

Phospholipid Nylon Strip Binding

Nylon strips with embedded phosphatidylinositol (PtdIns) lipids and other phospholipids (100 pmol per spot; Echelon Biosciences) were pre-incubated in blocking buffer [3% bovine serum albumin (BSA; w/v; fatty acid free) in standard phosphate-buffered saline (PBS) pH 7.4 with 0.1% Tween-20 (PBST)] for 2 h at room temperature (rt) on a rocking platform. rSpTrf-E1 or recombinant fragments (~20 nM) were incubated with a lipid-embedded strip in fresh blocking buffer for 2 h at rt with rocking. Unbound proteins were removed with three washes of PBST. Strips were incubated for 2 h at rt with primary antibodies composed of three polyclonal rabbit anti-SpTrf antibodies [anti-SpTrf-66, -68, and -71; 1:3,500 dilution (14, 15)] in blocking buffer, washed, and post-incubated with goat anti-rabbit IgG conjugated to horseradish peroxidase (G&RIg-HRP; 1:7,000 dilution in blocking buffer; Thermo Scientific Pierce) for 1 h at rt with rocking. Antibody–protein complexes on washed strips were visualized by incubation with enhanced chemiluminescence Western blotting substrate (Thermo Scientific Pierce) and exposed to autoradiography film (MidSci). Experiments were performed at least three times to confirm binding between rSpTrf-E1 or the recombinant fragments and the phospholipids. Negative controls omitted either rSpTrf-E1, the recombinant fragments, or the primary antibodies.

Liposome Preparation

Small unilamellar vesicles (SUVs; <100 nm) were prepared according to Kessler et al. (22) with modifications using varying mixtures of lipid concentrations including 100, 95, 90, and 80% of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (dioleoyl PC) and a corresponding 0, 5, 10, and 20% of 1,2-dioleoyl-*sn*-glycero-3-phosphate (PA; Avanti Polar Lipids, Inc.). For fluorescence resonance energy transfer (FRET) assays (see below), 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiI) was added to the lipid mixture at a mass ratio of 1:800 (DiI:PC). The lipid mixture was dried under nitrogen in test tubes that were cleaned with base bath (ethanol with potassium hydroxide solution) and rinsed several times with distilled water. Lipids were dissolved in chloroform and rotor-evaporated under nitrogen for 30 min followed by vacuum desiccation (Bel-Art Products) to remove all organic solvents. Sucrose (2%) in PBS at 80°C was added to the desiccated lipids and vortexed for 30 s at rt until the solution became opaque. The sucrose–lipid mixtures were incubated at 80°C in an Isotemp oven (Thermo Fisher Scientific) for 15 min followed by vortexing for 30 s, which was repeated twice. Lipids were resuspended in distilled water to multilamellar vesicles and converted to SUVs by bath sonication for 1–2 h with an UltraSonic Cleaner FS30H (Thermo Fisher Scientific). The SUV size range was determined using dynamic light scattering on a Beckman Coulter N5 submicron particle size analyzer with a 1-cm path length cuvette with a 30-min equilibrium time at rt with light scattering angle of 90°. The average size from three repetitions was evaluated and reported as the size of the SUVs.

Large unilamellar vesicles (LUVs; ~100–1,000 nm) were generated by mixing lipids in a ratio of 10% PA to 90% PC in chloroform followed by initial drying under nitrogen gas followed by a secondary drying step under vacuum for 2 h. Vesicles were rehydrated in

standard PBS with 10 mM 8-aminonaphthalene-1,3,6-trisulfonic acid disodium salt (ANTS; dye) and 15 mM *p*-xylene-Bis-pyridinium bromide (DPX; quencher). Liposomes were allowed to swell for 5 min before vortexing for 30 s, followed by heating to 45°C and undergoing five cycles of freeze/thaw using a dry ice–ethanol bath for 3 min per cycle, and ending with an incubation at 45°C for 5 min. Excess dye and quencher surrounding the loaded liposomes were removed by gel filtration through a Sephadex® G-25 Medium (Sigma-Aldrich) column. This procedure generated liposomes of 90% LUVs (50–1,000 nm) and 10% SUVs as measured by dynamic light scattering (Wyatt Technologies). The concentration of LUVs was determined according to Antimisariis (23).

Giant unilamellar vesicles (GUVs, >1 µm) were synthesized using electroformation according to Angelova and Dimitrov (24) with modifications from Kessler et al. (22). The lipids were combined in 1:9 M ratio (v:v) of PA:PC in a chloroform and methanol solvent. The fluorescent dye 1,1'-dioctadecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate (DiD) was incorporated into the lipid bilayers for an overall 0.08 mol% of DiD to lipid. For PA clustering detection, 6% 1-oleoyl-2-(6-[(7-nitro-2-(1,3-benzoxadiazol-4-yl)amino)hexanoyl]-*sn*-glycero-3-phosphate (NBD-PA) and 4% unlabeled PA were mixed with 90% PC. Lipids and dyes were mixed thoroughly to homogeneity and 10 µl of the mixed sample was coated onto two platinum wire electrodes (1.2 mm diameter). The electrodes with the lipid mixture were placed inside a vacuum desiccator to complete the solvent evaporation followed by emersion in a non-electrolyte buffer solution of 2% (w/v) sucrose. For microscopy imaging, 0.167 µM dextran labeled with Alexa Fluor® 488 (dextran-488, 3,000 MW, Anionic; Thermo Scientific Invitrogen) was added to the sucrose solution. Electrodes were connected to a waveform generator (Hewlett Packard) and incubated at 80°C during the electroformation procedure that started at 0.7 V with a frequency of 10 Hz, followed by stepwise voltage increases of 0.05 V every 5 min to 1.4 V, which was maintained for 3 h. Vesicles were separated from the electrodes by a final step of 0.6 V and 4 Hz, and sample cells were allowed to cool slowly to rt overnight. To separate the vesicles from the dextran-488 in solution that was not incorporated into the vesicles lumens, 300 µl of vesicles in solution were mixed with 100 µl of sugar solution (1.8% sucrose, 0.2% glucose) and spun at $15.8 \times 10^3 \times g$ for 10 min in a microfuge (Eppendorf). The top 200 µl of the solution was removed, and 200 µl of sucrose/glucose solution was mixed with the remaining vesicles, spun, and repeated three times. The density difference between sucrose and glucose allowed for a gentle separation and removal of the excess dye. Evaluation by confocal microscopy was used to verify that the vesicle lumens exhibited a stronger signal and greater concentration of the dextran-488 compared to the exterior solution. Vesicle sizes that are observable by conventional microscopy are limited to 1 µm to 1 mm. Because the images displayed a high contrast between the bilayer labeled with DiD and the background, the number of micelles, SUVs, and LUVs that might interfere with imaging was minimal.

Fluorescence Resonance Energy Transfer

Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), or BSA was biotinylated with 50 µM of EZ-Link® Sulfo-NHS-LC-LC-Biotin (Thermo Fisher Scientific) following the manufacturer's

instructions and mixed with NeutrAvidin-fluorescein isothiocyanate (NA-FITC) (1:100 dilution; Pierce) according to Lun et al. (12). Biotinylated rSpTrf-E1 labeled with NA-FITC (rSpTrf-E1-FITC) was added to each well of a black 96-well round bottom plate (Corning Costar) containing 100 μ l of SUVs. Samples were mixed and immediately excited at 450 nm to initiate FRET and emission was recorded at 560 nm. Excitation was repeated three times for each sample, recorded with a SpectraMax M5 Microplate Reader (Molecular Devices), and analyzed with the microplate data software SoftMax Pro (ver. 5) in the Read Mode setting for Spectrum and Fluorescence. After each reading, additional rSpTrf-E1-FITC was added to the SUVs and FRET was re-evaluated. The concentration of rSpTrf-E1-FITC added to the SUVs ranged from 0 to 10 μ g. Background was determined from samples that omitted rSpTrf-E1-FITC and were evaluated for FRET with increasing concentrations of NA-FITC. Negative controls employed 0–10 μ g of biotinylated BSA labeled with NA-FITC (BSA-FITC). The background levels were subtracted from the experimental results to generate the net FRET for the SUVs with rSpTrf-E1-FITC and BSA-FITC. A two-tailed, paired *t*-test was used to determine statistical significance among the net FRET results, which was recognized at $p \leq 0.05$. Means and SDs of the FRET data were calculated for each assay.

Microscopy

Giant unilamellar vesicles (200–300 μ l) in solution were placed in a Granier CELLSTAR® 96-well flat-bottom plate (Sigma-Aldrich) and allowed to settle for 30 min to 1 h and verified by confocal microscopy. Either 10 μ M rSpTrf-E1 or 1 μ l PBS (background control) was added to a region of the well in which there were many GUVs, which was imaged in multiple fields every 30 s for 30 min. Images were collected using an inverted Zeiss LSM 510 confocal microscope with a 63×1.2 NA water objective lens. DiD was excited with HeNe 633 nm laser, and images were collected with an emission range of 650–750 nm. Images with dextran-488 and NBD-PA were collected using an argon 488 nm laser with an emission range of 515–750 nm. Image J (National Institutes of Health¹) was used to view and assemble the images.

Circular Dichroism (CD)

Circular dichroism spectra of rSpTrf-E1 were obtained using a measurement range of 190–260 nm with 50 nm/min scanning speed, 1 nm bandwidth, 8 s response time with 1.0 nm data pitch for five scans as described (17). rSpTrf-E1 (0.25 μ M) was evaluated alone or in the presence of 1 mM SUVs that were composed of 10% PA:PC or 100% PA after equilibration for at least 10 min and not more than 30 min at rt. Background baseline CD spectra of 10 mM sodium phosphate buffer (pH 7.4) were subtracted from samples including rSpTrf-E1. Boxcar smoothing was used to remove noise from the signal. CD spectra were used to calculate the mean residue ellipticity, or θ , with standard units of degrees (deg) \times cm² \times dmol⁻¹. The fractional helicity was calculated using the ellipticity ratio ($R = \theta_{222}/\theta_{207}$) with the spectral data at 222 and 207 nm (25). CD spectra results were deconvoluted to calculate

the percentage of protein secondary structure using the CDNN program² (26, 27), and DichroWeb server³ (17, 28).

Vesicle Leakage Assay

Large unilamellar vesicles loaded with ANTS and DPX (see above) were mixed with rSpTrf-E1, and fluorescence was detected with a SpectraMax M5 (Molecular Devices, LLC) in which ANTS was excited at 360 nm and detected at 520 nm. rSpTrf-E1 (10 μ M) was added at $t = 0$, and data collection was terminated when the fluorescence signal ceased to increase and appeared to reach a steady state. All analyses were performed with 10 μ M lipid concentration and corrected for background fluorescence obtained for the lipids alone. LUVs loaded with ANTS and DPX were lysed with 0.1% Tween-20 and used as the positive control to determine the maximum fluorescence in the absence of quenching. Fractional fluorescence (f_i) was calculated by

$$f_i = (F_t - F_0) / (F_{\max} - F_0),$$

where F_0 is the initial fluorescence measured prior to rSpTrf-E1 addition, F_{\max} is the maximum fluorescence obtained when loaded LUVs were lysed in detergent, and F_t is the fluorescence measured at time t . The kinetics of ANTS leakage was modeled and fitted with a simple three variable equation.

$$f_i(t) = A_0 + A_1[1 - e^{-k_1 t}].$$

In this model, A_0 is the fraction that is released initially, and A_1 is the fraction that is released with a rate of k_1 per time (t) in seconds. All kinetic curves were fit using Matlab (The Mathworks, Inc.) in which the three variables were varied until the sum of the square error was minimized (29–32).

RESULTS

rSpTrf-E1 and Recombinant Fragments Bind to Specific Lipids

Native SpTrf and HeTrf proteins in cells are associated with vesicle membranes and are present on the exterior surface of the plasma membrane (8, 14, 20), which does not agree with predictions from amino acid sequences that these proteins have no obvious means for membrane association. Consequently, to determine whether the membrane association observed by microscopy could be replicated using other approaches, rSpTrf-E1 and the three recombinant fragments (**Figure 1A**) of the full-length protein were incubated with a lipid-embedded nylon strip to screen for binding to phospholipids, a few phosphatidylinositol (PtdIns) lipids, and seven other biologically important lipids. rSpTrf-E1 and the rGly-rich fragment bound only to PA, whereas the rHis-rich fragment bound to PA and to phosphatidylinositol-4-phosphate [PtdIns(4)P] although the spot intensity for PtdIns(4)P suggested weaker binding (**Figure 1B**). Alternatively, the rC-Gly fragment bound weakly only to phosphatidylserine (PS). The structures of PA and PtdIns(4)P to which rSpTrf-E1 and the rGly-rich and

¹<http://imagej.nih.gov/ij/>.

²<http://gerald-boehm.de/download/cdnn>.

³<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>.

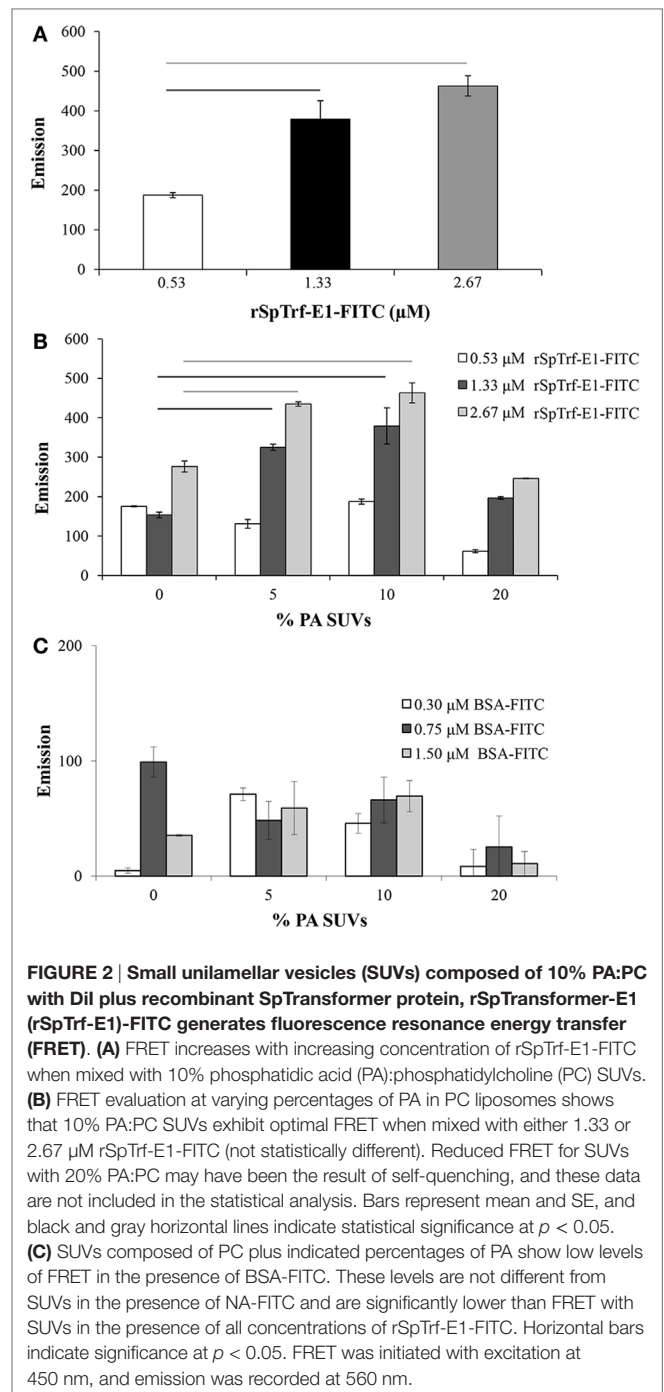
rHis-rich fragments bound suggested that exposed phosphates with extended or terminal chemical orientations may be the basis for interactions with the proteins. This result was in agreement with speculations that rSpTrf-E1 bound to charged groups including phosphates on LPS (12) and the sulfate group on SDS (17).

rSpTrf-E1 Interacts Closely with PA

To verify a close interaction between rSpTrf-E1 and PA, FRET was used to evaluate the emission from FITC linked to rSpTrf-E1 to excite DiI in liposome membranes containing PA. The recombinant fragments were not evaluated in the experiments using FRET because the rGly-rich and rHis-rich fragments gave the same results as rSpTrf-E1 on the lipid strips, and the rC-Gly fragment, which multimerizes upon isolation (12), resulted in a different lipid-binding signature that may not reflect the activities of the intact protein. PC was used as the neutral lipid background to stabilize the negatively charged PA because it is commonly found in most cell membranes (33) and was not bound by rSpTrf-E1 (Figure 1B). In an initial experiment, SUVs composed of 10% PA:PC and labeled with DiI were mixed with increasing concentrations of rSpTrf-E1-FITC and energy transfer was measured at 560 nm. Background was determined by the 560 nm emission of 10% PA:PC SUVs and DiI plus increasing concentrations of unlabeled rSpTrf-E1, which was subtracted from the experimental signal to determine the net FRET. FRET results, which are generally accepted to indicate that molecules are within 10 nm of each other, suggested that FITC and DiI, and therefore rSpTrf-E1-FITC and PA, were in very close association (Figure 2A). To determine the optimal percentage of PA in liposomes to optimize FRET with rSpTrf-E1-FITC, SUVs with increasing concentrations of PA were compared to SUVs composed only of PC (background control). In general, net FRET increased with increasing concentrations of rSpTrf-E1-FITC plus SUVs with a given percentage of PA, and SUVs composed of 5–10% PA:PC produced significantly increased FRET with increasing concentrations of rSpTrf-E1 (Figure 2B). FRET resulting from 0.53 μ M rSpTrf-E1-FITC did not change with respect to the percentage of PA in the SUVs suggesting that this concentration of rSpTrf-E1-FITC was too low to initiate FRET. SUVs composed of 20% PA:PC showed signs of self-quenching and produced lower net FRET (Figure 2B) likely because the higher concentration of NA-FITC in the controls interfered with emission detection (34). Therefore, results with 20% PA:PC SUVs were not included in the statistical analyses and were not evaluated further. rSpTrf-E1-FITC binding to SUVs containing PA appeared to be specific, because increasing concentrations of BSA-FITC did not show significant changes in net FRET when evaluated with SUVs with various percentages of PA (Figure 2C). FRET emission results suggested that rSpTrf-E1 associated closely with PA (Figure 2B) and confirmed results for rSpTrf-E1 binding to the lipid-embedded nylon strip (Figure 1B). The combination of 10% PA:PC liposomes and 2.67 μ M rSpTrf-E1-FITC was used for further analyses.

rSpTrf-E1 Causes Budding, Invagination, Fusion, and Leakage of GUVs

To visualize the close physical association between rSpTrf-E1 and liposomes containing PA suggested by FRET, rSpTrf-E1-FITC



was added slowly to one edge of a well in a flat-bottom plate containing GUVs labeled with DiD. Images captured by confocal microscopy over 20 min (four scans per min) did not show a colocalization of FITC and DiD likely because of limited sensitivity by the imaging system, which did not detect the low concentration of rSpTrf-E1-FITC relative to DiD. However, unexpected morphological changes to the GUVs were observed in the presence of rSpTrf-E1-FITC, which were not observed in the absence of the protein. After about 9 min, GUVs showed evidence of

budding, fusion, and invagination (Figure S1 and Movie S1 in Supplementary Material; white arrows indicate fusion and budding). When the GUVs were imaged again after a several hours, they were completely lysed.

Based on the initial results suggesting that rSpTrf-E1 induced morphological changes in GUVs, improved visualization of GUVs employed dextran-488 in the lumen, DiD in the membrane, and unlabeled rSpTrf-E1, with images captured every 30 s for 20–40 min by confocal microscopy. Images confirmed the initial results and showed changes in membrane curvature for some GUVs after the addition of rSpTrf-E1 that appeared as budding, invagination, and perhaps lysis (Figure 3). Similar morphological changes were not observed for GUVs in the absence of rSpTrf-E1, which remained as spheres for the duration of the observations (Figure S2 in Supplementary Material).

A progression of budding for two GUVs over 2.5 min resulted in the appearance of two or three smaller sized GUVs (Figure 3A, a–d, white and yellow arrows). GUV fusion was also observed in which two different sized GUVs came together and fused forming a kidney bean-shaped GUV (Figure 3B, a–e; orange arrows). This kidney bean-shaped GUV proceeded to invaginate into a multilamellar vesicle (Figure 3B, f–h) with an internal vesicle labeled with DiD but without dextran-488 in the lumen (Figure 3B, g,h; orange arrows). GUV invagination was also observed in which an elongated vesicle changed its morphology to a multilamellar GUV over 3 min in which the resulting internal vesicle was also devoid of dextran-488 in the lumen (Figure 3C, a–h; red arrows). What appeared to be GUV lysis in the presence of rSpTrf-E1 was observed when a bright green fluorescent GUV disappeared within 30 s (Figure 3C, d,e; blue arrows) suggesting

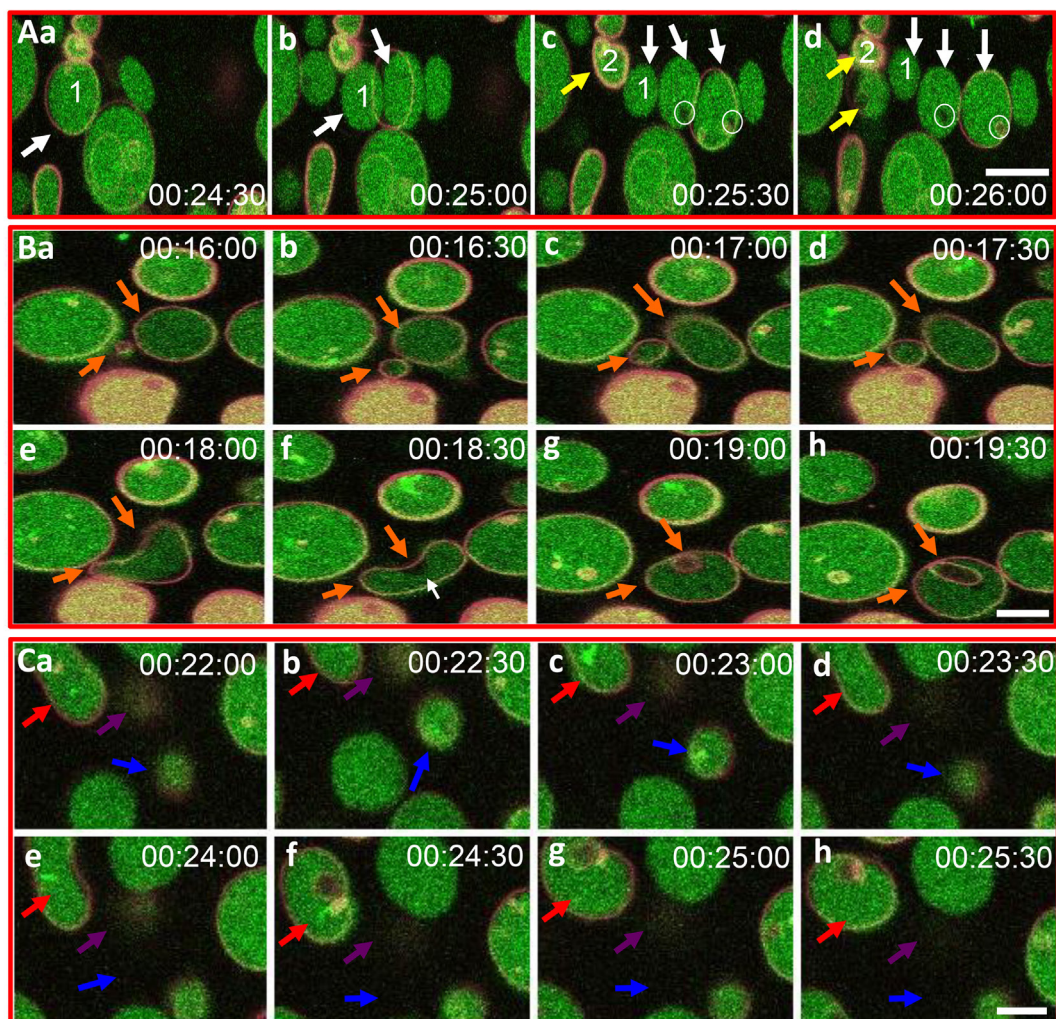


FIGURE 3 | Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), induces giant unilamellar vesicles (GUVs) to bud, fuse, invaginate, leak, and disappear. (A) Confocal microscopy images show budding of two independent GUVs into two or three smaller vesicles (a–d, white and yellow arrows). Leakage of dextran-488 appears as black spaces in the lumen of two GUVs (c,d, white circles). **(B)** Images show GUV fusion between two GUVs (a–e, orange arrows), leakage at the convex curve of the membrane (white arrow), which is the site of invagination of the fused GUV (f–h, orange arrows). **(C)** Images show invagination (a–h, red arrows), lysis (a–h, blue arrows), and a slow decrease in dextran-488 fluorescence in a GUV (a–h, purple arrows) suggestive of slow leakage leading to lysis. Image acquisition is every 30 s as indicated after the addition of rSpTrf-E1. All scale bars indicate 10 μm.

that the protein may induce membrane destabilization leading to the complete release of vesicle contents.

rSpTrf-E1 Causes Leakage of Luminal Contents from LUVs

In addition to the apparent GUV invagination, fusion, budding, and lysis events, an uneven distribution of the green dextran appeared as dark regions within the lumen of some GUVs (Figure 3A, c,d; white circles) suggesting that the dextran-488 may have leaked from the GUVs. For example, a dark region in the lumen was noted near the convex curve in the fused GUV (Figure 3B, f; white arrow) just prior to invagination that occurred at this same location. This change in the distribution of luminal dextran-488 was not observed in the control GUVs in the absence of rSpTrf-E1 (Figure S2 in Supplementary Material). Although, lysis of one particular GUV was suggested above (Figure 3C, a–e; purple arrows) an alternative possibility was that rSpTrf-E1 may alter the membrane to allow dextran solution to escape from the liposome and diffuse into the surrounding buffer to concentrations below detection by microscopy. To verify that these changes were due to lysis and/or leakage and to quantify the leakage rate, LUVs were loaded with ANTS (fluorescent dye) and DPX (quencher) and incubated with rSpTrf-E1 (both monomers and dimers) and with Ni-natSpTrf proteins isolated from two different sea urchins (Figure 4A). Based on the identification of PA binding (Figure 1), both the rGly-rich and rHis-rich fragments (Figure 4B) were also evaluated for GUV leakage. The rC-Gly fragment was not employed in this assay because it multimerizes upon isolation, bound poorly to PS and did not bind to PA (Figure 1B), and shows non-specific binding to a range of foreign targets (12). Negative control proteins included BSA and unknown proteins isolated by nickel affinity from non-induced *E. coli* that served as the negative control for the isolation protocol for the recombinant proteins (Figure 4A). After 2 h, increased ANTS fluorescence was only detected from LUVs incubated with monomeric rSpTrf-E1 or the rHis-rich fragment, which could be measured based on the separation of ANTS from DPX upon release from the liposome and diffusion into the buffer (Figure 4C). Although the rGly-rich fragment bound to PA (Figure 1B), it did not induce luminal content leakage from the LUVs. Leakage was not induced by either dimerized rSpTrf-E1 or the Ni-natSpTrf protein isolates, which were entirely multimerized upon collection from two sea urchins. This was the first evidence that dimerized rSpTrf-E1 was not active compared to the monomer and inferred that multimerization of the natSpTrf proteins may have been an attribute of the lack of leakage activity. These results also suggested differences in the activities of the His-rich and Gly-rich regions in rSpTrf-E1.

Recorded fractional fluorescence for ANTS release in the presence of rSpTrf-E1 or the rHis-rich fragment did not plateau by 2 h (Figure 4) indicating that neither protein had induced maximum leakage within that time frame. Because rSpTrf-E1 and the rHis-rich fragment had very similar fractional fluorescence results, only rSpTrf-E1 was used in the subsequent leakage assay of 5 h to identify the maximum leakage by reaching the fluorescence plateau. Three independent assays demonstrated that rSpTrf-E1 induced reproducible leakage, and these results were well described by a two-step process of an instantaneous first step

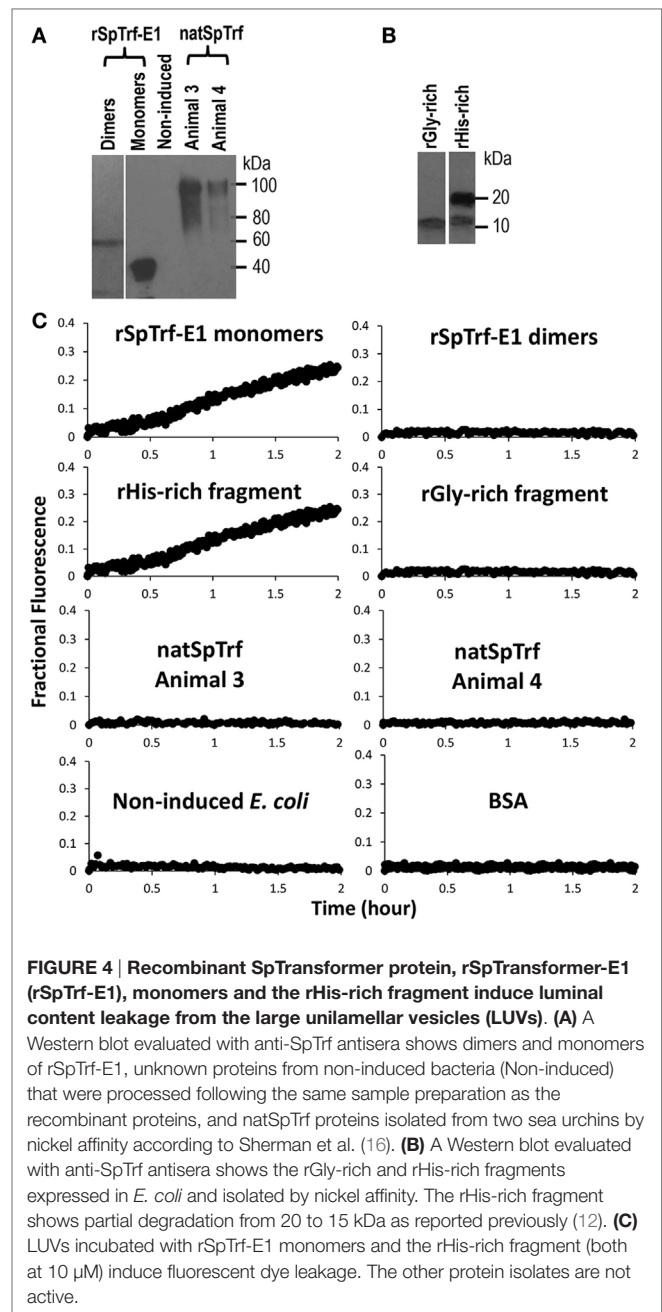


FIGURE 4 | Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), monomers and the rHis-rich fragment induce luminal content leakage from the large unilamellar vesicles (LUVs). (A) A Western blot evaluated with anti-SpTrf antisera shows dimers and monomers of rSpTrf-E1, unknown proteins from non-induced bacteria (Non-induced) that were processed following the same sample preparation as the recombinant proteins, and natSpTrf proteins isolated from two sea urchins by nickel affinity according to Sherman et al. (16). (B) A Western blot evaluated with anti-SpTrf antisera shows the rGly-rich and rHis-rich fragments expressed in *E. coli* and isolated by nickel affinity. The rHis-rich fragment shows partial degradation from 20 to 15 kDa as reported previously (12). (C) LUVs incubated with rSpTrf-E1 monomers and the rHis-rich fragment (both at 10 μ M) induce fluorescent dye leakage. The other protein isolates are not active.

and a slower rate-determining second step with a measurable rate (Figure 5). Calculations yielded an average leakage fraction of ~ 0.58 (A_1) that was released with an average kinetic rate (k_1) of $\sim 1.17 \times 10^{-4} \text{ s}^{-1}$. The fractional fluorescence showed a slow leakage process for 10 μ M rSpTrf-E1 that required 4–5 h before the fluorescence reached a stable plateau. The mode of action for rSpTrf-E1 appeared to require more time and may be more subtle and less drastic than interactions between known antimicrobial peptides and membranes (35). Although the actual sequence of events of natSpTrf proteins binding to targets *in vivo* is unknown and may have several steps and involve multiple natSpTrf isoforms, the kinetic findings for rSpTrf-E1 suggested a general interpretation of a first step as the protein binding to PA, and a

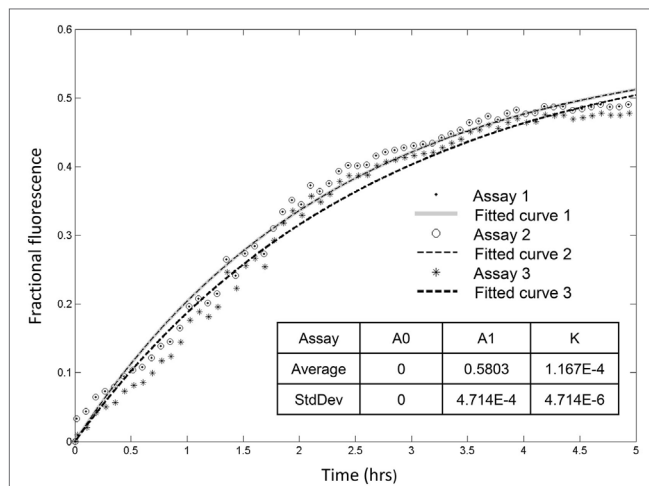


FIGURE 5 | Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), induces leakage that plateaus at about 5 h. Three independent leakage assays with 10 μ M rSpTrf-E1 show that reaching the fluorescence leakage plateau requires about 5 h. The table insert shows that the results are reproducible at 0 initial leakage rate (A_0) when rSpTrf-E1 is added to the sample with average fraction of ~ 0.58 (A_1) that is released with an average kinetic rate (k) of $\sim 1.17 \times 10^{-4} \text{ s}^{-1}$.

second step as a specific interaction or re-arrangement of proteins and lipids that led to membrane destabilization and leakage.

rSpTrf-E1 Causes PA to Cluster

Giant unilamellar vesicles in the presence of rSpTrf-E1 showed changes in membrane curvature leading to invagination or budding, which was not observed when PA was not incorporated into the GUVs. To determine whether changes in membrane curvature was a result of PA clustering, which might occur because PA has a conical shape resulting from the very small phosphate head group (36), GUVs of 6% NBD-PA, 4% PA, 90% PC plus DiD were imaged in the presence or absence of rSpTrf-E1. The 20-min time point was chosen to begin imaging because this was the point at which most changes in morphology were observed for GUVs loaded with dextran-488 after the addition of rSpTrf-E1 (**Figure 3**). Images of selected GUVs after the addition of rSpTrf-E1 showed clustered NBD-PA that formed bright blue fluorescent patches in the lipid bilayer (**Figures 6A–D**, white arrows; **Figures S3A–C** in Supplementary Material). NBD-PA clusters were sometimes present at the intersection of two GUVs (**Figure 6A**), in regions of membrane curvature (**Figure 6B**), and positioned at points of contact between membranes within multilamellar GUVs (**Figure S3B** in Supplementary Material). Confocal Z-stack images of the

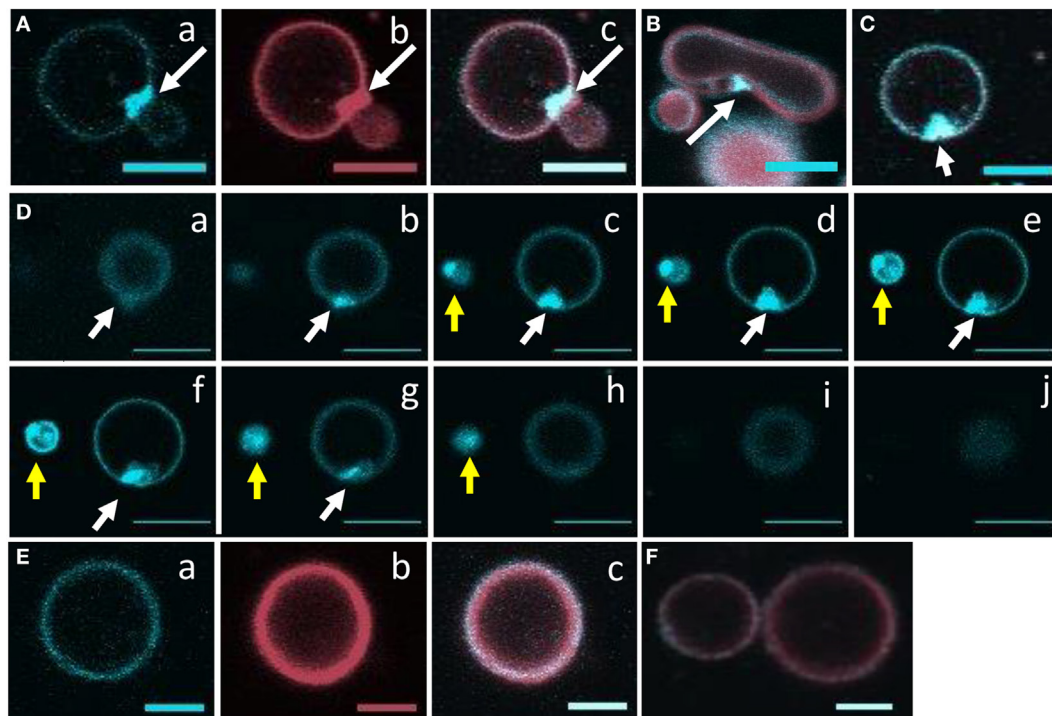


FIGURE 6 | Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), causes NBD-PA to cluster in the lipid bilayer. Confocal microscopy images were captured 20 min after the addition of rSpTrf-E1 to giant unilamellar vesicles (GUVs) that are composed of 6% NBD-PA, 4% PA, and 90% PC (100% g/ml). **(A)** An NBD-PA cluster (arrow) is present at the intersection of two GUVs. Images show NBD-PA (a), DiD in the GUV membrane (b), and the merge (c). **(B)** The merged image shows an NBD-PA cluster (arrow) at a region of concave curvature of a GUV membrane. **(C)** A single cluster of NBD-PA is present in a GUV membrane. **(D)** A Z-stack of images (a–j) from the bottom to the top of two GUVs (white and yellow arrows) shows that each GUV has a single NBD-PA cluster. **(E)** A GUV without added rSpTrf-E1 shows no change in NBD-PA distribution at 20 min. Images NBD-PA (a), DiD in the GUV membrane (b), and the merge (c). **(F)** Two GUVs without added rSpTrf-E1 show an even distribution of NBD-PA at 20 min. All scale bars indicate 10 μ m.

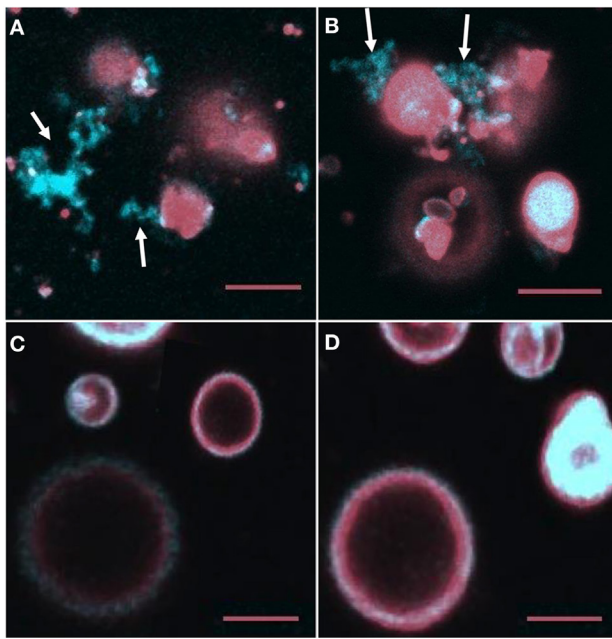


FIGURE 7 | NBD-PA becomes separated from giant unilamellar vesicles (GUVs) after 2 h of incubation with recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1). (A,B) NBD-PA (arrows) forms clusters that are separated from the GUVs after 2 h of incubation with rSpTrf-E1. (C,D) GUVs in the absence of rSpTrf-E1 show an even distribution of NBD-PA and DiD at 2 h. Differences in the GUV sizes and content of NBD-PA are an outcome of GUV preparation. All images are merged for NBD-PA (blue) and DiD (red) as captured by confocal microscopy. All scale bars indicate 10 μ m.

NBD-PA clusters in GUVs in the presence of rSpTrf-E1 showed that a single PA cluster was typically present per liposome rather than multiple clusters (Figure 6D). These morphological attributes were consistent with PA clusters being the basis for membrane curvature, budding, and invagination. NBD-PA clustering in GUVs was not observed in the absence of rSpTrf-E1 and showed an even distribution in the spherical GUVs (Figures 6E,F; Figures S3D–G in Supplementary Material). After 2 h of incubation of GUVs with rSpTrf-E1, the NBD-PA appeared in disordered clusters associated with but outside of the GUV membranes (Figures 7A,B; Figures S3H,I in Supplementary Material). In the absence of rSpTrf-E1 at 2 h, there was an even distribution of NBD-PA in the GUVs and no clusters of NBD-PA appeared within or outside of the GUV membranes (Figures 7C,D). These results suggested that the clustering of PA induced by rSpTrf-E1 proceeded to PA extraction from the membranes.

Many of the vesicles in the experiments reported here displayed no changes in membrane morphology in the presence of rSpTrf-E1 (Figure 3), and several factors may have been the basis for this observation. First, confocal imaging only has a small window for observation and recording of the events, which limited the number of vesicles that could be evaluated. Second, the addition of rSpTrf-E1 was added to an edge of the wells to minimize disturbing the settled vesicles, and likely induced a gradient of the protein across the well as it diffused into the solution. Third,

variations in the PA concentration among vesicles are known to occur (see Figures S3D–G in Supplementary Material). It was likely that a combination of all resulted in variations in the numbers of PA–rSpTrf-E1 interactions among individual vesicles that led to morphological changes in some vesicles and not in others.

rSpTrf-E1 Transforms from Disordered to α Helical in the Presence of PA and PA/PC Liposomes

Previous bioinformatic predictions and CD analysis of rSpTrf-E1 indicated that it is an IDP that transforms from disordered to α helical upon interactions with SDS, TFE, or LPS (12, 17). Based on the structural similarity between SDS (a single acyl chain linked to a sulfate group) and PA (two acyl chains linked to a phosphate head group), we hypothesized that rSpTrf-E1 binding to PA might drive similar secondary structural changes in the protein. Results from CD spectra of rSpTrf-E1 in the presence of PA, either as 100% PA SUVs or as 10% PA:PC SUVs, demonstrated that rSpTrf-E1 transformed from disordered to ~70% α helical structure (Figure 8). In the presence of fully neutral SUVs composed of 100% PC or in the absence of lipids, rSpTrf-E1 remained intrinsically disordered (~2% α helical) in agreement with a disordered structure in the absence of binding targets (17). These results suggested that the interaction between PA and rSpTrf-E1 was similar to observations with SDS and transformed the protein to α helical secondary structure.

DISCUSSION

Native SpTrf and HeTrf proteins are found in all morphotypes of sea urchin phagocytes, on the surface of small phagocytes in association with the plasma membrane and with the membranes of cytoplasmic vesicles (14, 19–21). But rather than integrated into membranes *via* transmembrane regions or associated through GPI linkages, rSpTrf-E1 and its rGly-rich and rHis-rich fragments may associate with membranes, at least in part, by binding directly to PA or other lipids with exposed phosphate groups. Interactions between rSpTrf-E1 and liposomes that include PA alter membrane curvature, which has been noted as a characteristic of cone-shaped PA in other systems (36, 37), and correlates with PA clustering that is likely the basis for budding, fusion, and invagination. Both monomeric rSpTrf-E1 and the rHis-rich fragment cause slow leakage of luminal contents demonstrating that the proteins do not induce sudden membrane disruption unlike activities of some antimicrobial peptides (35, 38). Both the rGly-rich and rHis-rich fragments bind PA suggesting that the full-length protein is at least bivalent, which is similar to results from other proteins with PA-binding domains that are likely multivalent (39). Although the rGly-rich fragment binds to PA, it does not induce leakage indicating that the His-rich region of rSpTrf-E1 is likely responsible for this activity. When rSpTrf-E1 and Ni-natSpTrf proteins are dimerized or multimerized prior to mixing with liposomes they do not induce leakage suggesting that only monomers are active. Irreversible multimerization among natSpTrf proteins has been noted repeatedly (12, 14–16), and we speculate that this may be an intrinsic

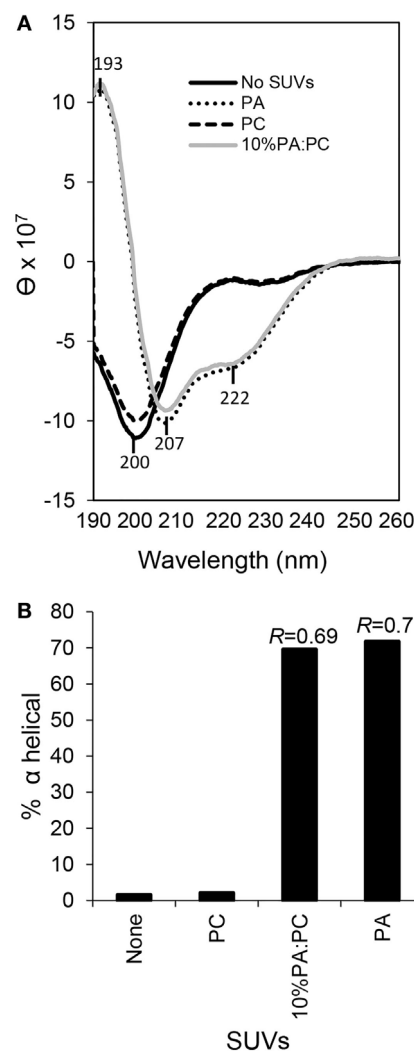


FIGURE 8 | Secondary structure of recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), transforms from intrinsically disordered to α helical in the presence of phosphatidic acid (PA) small unilamellar vesicles (SUVs). (A) CD spectra show intrinsic disorder or random coils for 0.25 mM rSpTrf-E1 in 10 mM sodium phosphate buffer in the absence of PA or in the presence of 100% phosphatidylcholine SUVs (PC). rSpTrf-E1 transforms to α helical secondary structure in the presence of 10% PA:PC SUVs or 100% PA SUVs (PA). θ is the mean residue ellipticity with standard units of degrees \times cm² \times dmol⁻¹ as described (17). (B) The percentage of α helical structure for rSpTrf-E1 is 1.57% in the absence of lipids and 2.1% in the presence of PC. However, in the presence of 100% PA SUVs or 10% PA/PC SUVs, the α helical structure of rSpTrf-E1 is 69.6 and 71.8%, respectively. The percentage of secondary structure for rSpTrf-E1 in the presence of PA are based the deconvolution of the spectra data using the DichroWeb server. The R values [ellipticity ratios: $R = \theta_{222}/\theta_{207}$ shown in panel (A)] are indicated for the CD analysis of rSpTrf-E1 with 10% PA:PC and for 100% PA (17).

control mechanism for natSpTrf proteins that do not bind quickly to pathogens or to other non-self targets and may limit the potential for destructive activities toward self. In the presence of PA and SDS, which have similar anionic and amphipathic structures, rSpTrf-E1 transforms from disordered to α helical structure (17). Similarly, speculations on the PA-binding domains from yeast

SNARE proteins also suggest protein disorder in the cytosol that alters to amphipathic α helical structure after binding to PA (39). Our findings provide the first evidence of a possible means by which natSpTrf proteins may associate with exposed phosphate groups on PAMPs including PA in membranes and that the His-rich region within rSpTrf-E1 has destabilizing activities for simple membranes.

Binding between rSpTrf-E1 and PA

There is no commonly recognized site or domain for any protein that binds PA; however, clusters of positively charged amino acids are speculated to be responsible for this interaction (40). The amino acid composition of rSpTrf-E1 is 24.8% positively charged amino acids [76 of 307 amino acids (aa); 27 His, 4 Lys, 45 Arg; see Table S3 in Ref. (12)]. Similarly, positively charged amino acids compose 30.4% of the rHis-rich fragment (56 of 184 aa; 27 His, 3 Lys, 26 Arg) and 17.5% of the rGly-rich fragment (15 of 86 aa; 1 Lys, 14 Arg), and each of these recombinant proteins binds to PA. In comparison, none of the recombinant proteins tested here bind to diacylglycerol, which is identical to PA but without the phosphate head group, suggesting that the interaction is focused on the phosphate. In addition to PA, the rHis-rich fragment binds to PtdIns(4)P, which also has an exposed phosphate on the inositol head group, although binding appears to have lower affinity compared to PA (Figure 1B). Binding to PtdIns(4)P may require a higher percentage of positively charged amino acids that are present in the rHis-rich fragment compared to the other recombinant proteins tested in this study and may offset the possibility that the phosphate on PtdIns(4)P may be less accessible than on PA. It is noteworthy that rSpTrf-E1 that includes the His-rich region does not bind to PtdIns(4)P suggesting an interaction between the Gly-rich and His-rich regions within the full-length protein to enhance or restrict binding to PA, which is relaxed when the rHis-rich fragment is expressed alone. Although the rC-Gly fragment has 15.4% positively charged amino acids (6 Arg of 39 aa), it does not bind to PA, which may be due to the spacing of the 6 arginines that are spread out as 2 singles and 2 doubles in this short fragment. The lipid binding by the rC-Gly fragment to PS, albeit weak based on spot intensity on the lipid-embedded strip (Figure 1B), may be an example of its characteristic of multimerization upon expression and its expanded range of microbial species to which it binds compared to rSpTrf-E1 (12).

The relatively high content of positively charged amino acids in rSpTrf-E1 and the rHis-rich and rGly-rich fragments are congruent with a proposed molecular model of an electrostatic/hydrogen bond switch (40) that may explain the interactions between the monomeric rSpTrf-E1 and PA in a lipid bilayer. This model proposes that upon the initial attraction, the positively charged amino acid side groups in the protein may interact electrostatically with PA in the bilayer and form hydrogen bonds with the negatively charged and exposed phosphate. When in close proximity, the hydrogen bonds between the negative charges on the phosphates and positively charged side groups increase due to deprotonation that strengthens the electrostatic attraction (40). The enhanced negative charges plus hydrogen bonds may result in a tight bond between PA and rSpTrf-E1 or the recombinant fragments resulting in docking of the protein to the lipid

(Figures 9A,B). Speculations on the electrostatic interactions between rSpTrf-E1 and phosphate groups are consistent with the previous report demonstrating that rSpTrf-E1 binds to LPS (12). Anionic phosphates on LPS are present on the glucosamine disaccharide in lipid A and also on the polysaccharide core (41) and these phosphates may also form charge-based electrostatic interactions with the positively charged amino acids in rSpTrf-E1.

rSpTrf Interactions with Liposome Membranes Containing PA

rSpTrf-E1 Clusters PA in Liposome Membranes

Recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), causes PA to cluster in liposome membranes as observed by the changes in the distribution of NBD-PA. This suggests that rSpTrf-E1 is bivalent, binds two PA molecules, and once bound through electrostatic interactions, it transforms from

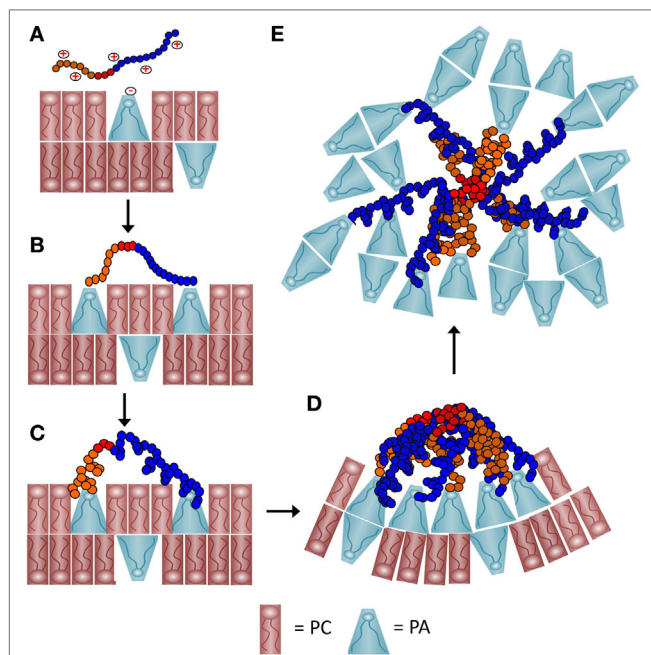


FIGURE 9 | A schematic representation of a proposed process of how recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1), may cause phosphatidic acid (PA) clustering and PA extraction from liposomes. (A) The positively charged amino acids (red+) in the Gly-rich region (orange) and the His-rich region (blue) of rSpTrf-E1 interact with the negatively charged (red-) phosphate head group of PA (blue cone-shaped lipid) through initial electrostatic attractions. Phosphatidylcholine (PC) (red rectangular lipid) is 90% of the lipids in the liposomes. **(B)** The positively charged amino acids from both the Gly-rich and His-rich regions of rSpTrf-E1 each bind to the phosphate head group on PA. The C terminal region of the Gly-rich (C-Gly) region (red) does not bind to PA. **(C)** Binding between rSpTrf-E1 and PA causes the protein to undergo a structural transformation from disordered to α helical. **(D)** The C-Gly region of α helical rSpTrf-E1 interacts with other C-Gly regions in other rSpTrf-E1 proteins causing protein multimerization or aggregation that brings PA into clusters. Clusters of cone-shaped PA induce liposome membrane curvature that leads to budding, invagination and fusion (not shown). **(E)** The C-Gly region continues to multimerize rSpTrf-E1 proteins into larger aggregates that extract PA from the liposomes and result in disordered PA clusters that are separated from the liposomes.

disordered to α helical structure (Figures 9A–C). This structural change may lead to or be concurrent with multimerization among rSpTrf-E1 proteins bound to PA on the lipid membrane that would bring PA into visible clusters (Figure 9D). The interaction time required for changes to become evident would depend on the number of rSpTrf-E1 proteins bound to PA on a particular liposome, the number of PA molecules in that membrane, and the fluidity of the membrane. Based on the conical shape of PA, its enrichment into clusters would be expected to promote membrane curvature (Figure 9D) (42) leading to the morphological changes observed as budding, fusion, and invagination. Membrane curvature reported here for liposomes is in agreement with PA involvement in membrane curvature in cells including (i) mitochondrial fusion and fission (43, 44); (ii) vesicle formation by generating membrane curvature in the Golgi complex (45), and (iii) membrane dynamics and vesicle trafficking along the secretory pathway including membrane fusion and exocytosis (46–48). The level of PA is elevated in vertebrate macrophages upon activation and functions in signal transduction to induce endocytosis, fusion of perinuclear vesicles with the plasma membrane, and immune activation of these cells (49). Enrichment of cone-shaped PA at sites of closely apposed membranes facilitates fusion to complete the formation of phagosomes, endosomal vesicles, and the process of exocytosis. Our observations of PA clustering induced by rSpTrf-E1, the positions of those clusters at regions of membrane curvature and at intersections of contact between two liposomes are consistent with the activities of PA in intact cells.

rSpTrf-E1 Extracts PA from Liposome Membranes

The second level of interactions between rSpTrf-E1 and liposomes containing PA is the apparent extraction of PA from the membranes after 2 h of incubation. This phenomenon may be the outcome of the Gly-rich and His-rich regions of rSpTrf-E1 each binding to a PA molecule followed by the transformation of the protein to α helical, the diffusion along the membrane of rSpTrf-E1-2PA complexes into close association with each other, and the multimerization of the proteins into larger complexes that is mediated by the rC-Gly region (Figure 9D). This would initially appear as large clusters of PA followed by continued multimerization of rSpTrf-E1-2PA not only within but between liposomes to generate complexes large enough to extract PA from liposome membranes (Figure 9E). This would require overcoming the PA acyl chain associations within the membrane and their extraction into the aqueous buffer, after which the PA acyl chains would likely associate with each other. A possible transition from membrane clusters of PA to extracted clusters is consistent with the image in Figure S3C in Supplementary Material. The final outcome of this process is disorganized clusters of PA that are distinct from the residual liposomes (Figure 7; Figure S3 in Supplementary Material).

rSpTrf-E1 Causes Liposome Leakage

The change in distribution of PA in membranes after mixing with rSpTrf-E1 correlates with both the appearance of dark regions within lumens of liposomes loaded with dextran-488 and the slow leakage of luminal contents. The change in liposome membrane

permeability requires a 20-min interaction time with rSpTrf-E1 or the rHis-rich fragment before leakage becomes evident. Although the rGly-rich fragment binds to PA, it does not induce leakage, indicating that the rHis-rich fragment and the His-rich region of rSpTrf-E1 are responsible for altering the characteristics of the liposome membrane to induce leakage. Leakage by the rHis-rich fragment also indicates that protein multimerization after PA binding is not required for the process because the rC-Gly region, which drives multimerization (12), is not included in the rHis-rich fragment (**Figure 1A**). This suggests that rSpTrf-E1 may have two activities that alter liposomes: (i) those that lead to membrane destabilization and leakage and (ii) those that lead to membrane curvature and changes in liposome morphology. However, these two activities may occur simultaneously in which PA binding leads to (i) membrane destabilization and eventual luminal leakage and (ii) PA clustering that leads to membrane curvature and PA extraction. For example, apparent luminal leakage associated with membrane curvature followed by invagination is illustrated for the bean-shaped GUV in **Figure 3B**, f–h (white arrow). It is not known whether membrane destabilization and leakage observed for liposomes in the presence of rSpTrf-E1 has an equivalent *in vivo* for natSpTrf activity.

CONCLUSION

We report that rSpTrf-E1 associates with the phospholipids PA and PtdIns(4)P. Although these results suggest a means by which this recombinant protein may associate with sea urchin coelomocytes and/or bacterial membranes, it is not known whether PA binding is an important interaction between natSpTrf proteins and membranes of intact cells. This is because there is no information on the phospholipid composition of sea urchin coelomocytes or the marine bacteria, *V. diazotrophicus*, to which rSpTrf-E1 and natSpTrf proteins are known to bind (12). PA is present in small quantities in most internal cellular membranes and is critical for many physiological functions including (i) serving as the precursor for phospholipid synthesis, (ii) involvement in important stress signaling pathways in plants and animals, and (iii) activities in enzyme activation, protein recruitment, cell stress response, and cell signaling (37, 47, 50–52). PA is elevated on the cytoplasmic side of the plasma membrane in vertebrate phagocytes (49) particularly during phagocytosis (39) and can readily translocate between membrane leaflets depending on pH and charge neutralization of the phosphate head group (33). Whether it accumulates on the surface of sea urchin coelomocytes in association with natSpTrf proteins is not known. PA binding by rSpTrf-E1 may represent the ability of this protein to bind lipids, proteins, and PAMPs with the common attribute of exposed phosphates. However, this does not rule out the possibility of receptors for natSpTrf proteins on small phagocytes and vesicle membranes. If exposed phosphates are a common binding target for a subset of natSpTrf proteins and are present on foreign target cells including LPS and PA, it may be possible for some natSpTrf proteins to bind both bacteria and coelomocytes, which may link bacteria with immune cells through natSpTrf multimerization, thus promoting phagocytosis. Furthermore, if PA clustering is induced by natSpTrf protein multimerization on the coelomocyte

surface, this may also aid in driving membrane curvature and endocytosis or phagocytosis.

The extraordinary protein diversity of the natSpTrf proteins that has been reported for sea urchins (14–16) suggests that subsets of these proteins may engage in different levels of phospholipid (or exposed phosphate) binding based on their amino acid sequence compositions. Depending on the element patterns of the mature proteins and putative editing of the mRNAs [reviewed in Ref. (6, 8)], the number of positively charged amino acids varies greatly among these proteins. Consequently, some natSpTrf proteins may not bind to free phosphates on lipids or other molecules, others may bind to different categories of lipids perhaps including the series of phosphatidyl inositols that are phosphorylated at all combinations of sites on the inositol ring. The results presented here infer more complex biological processes for this immune response protein family in sea urchins than previously considered, particularly if each natSpTrf protein variant has multiple and overlapping binding targets that includes not only a range of PAMPs but also a subset of macromolecules with free phosphates including membrane lipids.

AUTHOR CONTRIBUTIONS

CML was involved in all aspects of the research. AB generated the FRET data. RS and SG generated liposomes, were involved with the liposome experiments, and provided confocal microscopy imaging and image processing. LCS supervised and directed the research. CML, RS, SG, and LCS wrote, edited, and revised the manuscript. All authors approved the submitted manuscript.

ACKNOWLEDGMENTS

The authors are grateful to Anika Armstrong and Barney M. Bishop (George Mason University) for assistance with CD and to Martin Flajnik, Leon Grayfer, Ioannis Eleftherianos, and Robert Donaldson for comments on an early manuscript draft. The authors appreciate the thoughtful questions and comments from the reviewers.

FUNDING

This work was supported by a graduate stipend from the Wilbur V. Harlan Trust of the Department of Biological Sciences at George Washington University, an award from the Cosmos Club of Washington, DC, and two Columbian College of Arts and Sciences Summer Dissertation Fellowships to CML, awards from the George Washington University Facilitating Fund and the Columbian College Facilitating Fund to RS and SG, an undergraduate summer scholarship from the Wilbur V. Harlan Trust to AB, and awards from the National Science Foundation (IOS-1146124, IOS-1550474) to LCS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fimmu.2017.00481/full#supplementary-material>.

MOVIE S1 | Giant unilamellar vesicles (GUVs) show fusion, budding, and invagination in the presence of recombinant SpTransformer protein, rSpTransformer-E1 (rSpTrf-E1). The movie is composed of a series of confocal microscopy images captured every 15 s starting at ~12 min after the

addition of rSpTrf-E1 to the upper right edge of the well that likely diffuses across the well from upper right to lower left. The GUVs show budding (first white arrow that appears), fusion, and invagination (subsequent white arrows) in the presence of rSpTrf-E1.

REFERENCES

- Hibino T, Loza-Coll M, Messier C, Majeske AJ, Cohen A, 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
- Rast JP, Smith LC, Loza-Coll M, Hibino T, Litman GW. Genomic insights into the immune system of the sea urchin. *Science* (2006) 314:952–6. doi:10.1126/science.1134301
- Sodergren E, Weinstock GM, Davidson EH, Cameron RA, Gibbs RA, Angerer RC, et al. The genome of the sea urchin, *Strongylocentrotus purpuratus*. *Science* (2006) 314:941–52. doi:10.1126/science.1133609
- 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) 26:134–44. doi:10.1152/physiolgenomics.00011.2006
- 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
- Ghosh J, Buckley KM, Nair SV, Raftos DA, Miller CA, 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
- 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) 11:575. doi:10.1186/1471-2164-11-575
- 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
- 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) 17:900. doi:10.1186/s12864-016-3241-x
- 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
- 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, beta-1,3-glucan, and dsRNA. *BMC Mol Biol* (2007) 8:16. doi:10.1186/1471-2199-8-16
- Lun CM, Schrankel CS, Chou H-Y, 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) 221:889–903. doi:10.1016/j.imbio.2016.03.006
- Buckley KM, Terwilliger DP, Smith LC. Sequence variations in 185/333 messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. *J Immunol* (2008) 181:8585–94. doi:10.4049/jimmunol.181.12.8585
- Brockton V, Henson JH, Raftos DA, Majeske AJ, Kim YO, 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) 121:339–48. doi:10.1242/jcs.012096
- Dheilly NM, Nair SV, Smith LC, Raftos DA. Highly variable immune-response proteins (185/333) from the sea urchin *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. *J Immunol* (2009) 182:2203–12. doi:10.4049/jimmunol.07012766
- Sherman LS, Schrankel CS, Brown KJ, Smith LC. Extraordinary diversity of immune response proteins among sea urchins: nickel-isolated Sp185/333 proteins show broad variations in size and charge. *PLoS One* (2015) 10:e0138892. doi:10.1371/journal.pone.0138892
- 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) 198:2957–66. doi:10.4049/jimmunol.1601795
- Smith LC, Coscia MR. Tuning the host-pathogen relationship through evolution with a special focus on the echinoid Sp185/333 system. *Invertebrate Surviv J* (2016) 13:355–73.
- Majeske AJ, Oren M, Sacchi S, Smith LC. Single sea urchin phagocytes express messages of a single sequence from the diverse Sp185/333 gene family in response to bacterial challenge. *J Immunol* (2014) 193:5678–88. doi:10.4049/jimmunol.1401681
- Dheilly N, Birch D, Nair SV, Raftos DA. Ultrastructural localization of the highly variable immune response protein (He185/333) within coelomocytes and the gut tissues of the sea urchin, *Heliocidaris erythrogramma*. *Immunol Cell Biol* (2011) 89:861–9. doi:10.1038/icb.2011.3
- 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:e62079. doi:10.1371/journal.pone.0062079
- Kessler MS, Samuel RL, Gillmor SD. Polka-dotted vesicles: lipid bilayer dynamics and cross-linking effects. *Langmuir* (2013) 29:2982–91. doi:10.1021/la3042007
- Antimisiaris SG. Preparation of DRV liposomes. In: Weissig V, editor. *Liposomes: Methods and Protocols*. New York, NY: Humana Press (2010). p. 51–75.
- Angelova MI, Dimitrov DS. Liposome electroformation. *Faraday Discuss Chem Soc* (1986) 81:303–11. doi:10.1039/dc9868100303
- Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta* (2005) 1751:119–39. doi:10.1016/j.bbapap.2005.06.005
- Boehm G, Muhr R, Jaenicke R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Eng* (1992) 5:191–5. doi:10.1093/protein/5.3.191
- Boehm G. CDNN. (1998). Available from: <http://gerald-boehm.de/download/cdnn>
- Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* (2004) 32:W668–73. doi:10.1093/nar/gkh371
- Ladokhin AS, Wimley WC, White SH. Leakage of membrane vesicle contents: determination of mechanism using fluorescence reequenching. *Biophys J* (1995) 69:1964–71. doi:10.1016/S0006-3495(95)80066-4
- Ladokhin AS, Selsted ME, White SH. Bilayer interactions of indolicidin, a small antimicrobial peptide rich in tryptophan, proline, and basic amino acids. *Biophys J* (1997) 72:794–805. doi:10.1016/S0006-3495(97)78713-7
- Ladokhin AS, Wimley WC, Hristova K, White SH. Mechanism of leakage of contents of membrane vesicles determined by fluorescence reequenching. *Meth Enzymol* (1997) 278:474–86. doi:10.1016/S0076-6879(97)78025-X
- Pokorny A, Almeida PF. Kinetics of dye efflux and lipid flip-flop induced by delta-lysine in phosphatidylcholine vesicles and the mechanism of graded release by amphipathic, alpha-helical peptides. *Biochemistry* (2004) 43:8846–57. doi:10.1021/bi0497087
- van Meer G, Voelker DR, Feigenson GW. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol* (2008) 9:112–24. doi:10.1038/nrm2330
- Munkholm C, Parkinson DR, Walt DR. Intramolecular fluorescence self-quenching of fluoresceinamine. *J Am Chem Soc* (1990) 112:2608–12. doi:10.1021/ja00163a021
- Sato H, Feix JB. Peptide-membrane interactions and mechanisms of membrane destruction by amphipathic alpha-helical antimicrobial peptides. *Biochim Biophys Acta* (2006) 1758:1245–56. doi:10.1016/j.bbamem.2006.02.021
- Zimmerberg J, Kozlov MM. How proteins produce cellular membrane curvature. *Nat Rev Mol Cell Biol* (2006) 7:9–19. doi:10.1038/nrm1784
- Shin JJ, Loewen CJ. Putting the pH into phosphatidic acid signaling. *BMC Biol* (2011) 9:85. doi:10.1186/1741-7007-9-85

38. Du H, Samuel RL, Massiah MA, Gillmor SD. The structure and behavior of the NA-CATH antimicrobial peptide with liposomes. *Biochim Biophys Acta* (2015) 1848:2394–405. doi:10.1016/j.bbame.2015.07.006
39. Kassas N, Tanguy E, Thahouly T, Fouillen L, Heintz D, Chasserot-Golaz S, et al. Comparative characterization of phosphatidic acid sensors and their localization during frustrated phagocytosis. *J Biol Chem* (2017) 292:4266–79. doi:10.1074/jbc.M116.742346
40. Kooijman EE, Tieleman DP, Testerink C, Munnik T, Rijkers DT, Burger KN, et al. An electrostatic/hydrogen bond switch as the basis for the specific interaction of phosphatidic acid with proteins. *J Biol Chem* (2007) 282:11356–64. doi:10.1074/jbc.M609737200
41. Raetz CR, Whitfield C. Lipopolysaccharide endotoxins. *Annu Rev Biochem* (2002) 71:635–700. doi:10.1146/annurev.biochem.71.110601.135414
42. Kooijman EE, Chupin V, Fuller NL, Kozlov MM, De Kruijff B, Burger KNJ, et al. Spontaneous curvature of phosphatidic acid and lysophosphatidic acid. *Biochemistry* (2005) 44:2097–102. doi:10.1021/bi0478502
43. Choi SY, Huang P, Jenkins GM, Chan DC, Schiller J, Frohman MA. A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. *Nat Cell Biol* (2006) 8:1255–62. doi:10.1038/ncb1487
44. Yang CY, Frohman MA. Mitochondria: signaling with phosphatidic acid. *Int J Biochem Cell Biol* (2012) 44:1346–50. doi:10.1016/j.biocel.2012.05.006
45. Yang JS, Gad H, Lee SY, Mironov A, Zhang L, Bezoussenko GV, et al. COPI vesicle fission: a role for phosphatidic acid and insight into Golgi maintenance. *Nat Cell Biol* (2008) 10:1146–53. doi:10.1038/ncb1774
46. Huang P, Altshuler YM, Hou JC, Pessin JE, Frohman MA. Insulin-stimulated plasma membrane fusion of Glut4 glucose transporter-containing vesicles is regulated by phospholipase D1. *Mol Biol Cell* (2005) 16:2614–23. doi:10.1091/mbc.E04-12-1124
47. Zeniou-Meyer M, Zabari N, Ashery U, Chasserot-Golaz S, Haeberle AM, Demais V, et al. Phospholipase D1 production of phosphatidic acid at the plasma membrane promotes exocytosis of large dense-core granules at a late stage. *J Biol Chem* (2007) 282:21746–57. doi:10.1074/jbc.M702968200
48. Bullen HE, Soldati-Favre D. A central role for phosphatidic acid as a lipid mediator of regulated exocytosis in apcomplexa. *FEBS Lett* (2016) 590:2469–81. doi:10.1002/1873-3468.12296
49. Bohdanowicz M, Schlam D, Hermansson M, Rizzuti D, Fairn GD, Ueyama T, et al. Phosphatidic acid is required for the constitutive ruffling and macropinocytosis of phagocytes. *Mol Biol Cell* (2013) 24:1700–12, S1–7. doi:10.1091/mbc.E12-11-0789
50. Testerink C, Munnik T. Phosphatidic acid: a multifunctional stress signaling lipid in plants. *Trends Plant Sci* (2005) 10:368–75. doi:10.1016/j.tplants.2005.06.002
51. Liu Y, Su Y, Wang X. Phosphatidic acid-mediated signaling. *Adv Exp Med Biol* (2013) 991:159–76. doi:10.1007/978-94-007-6331-9_9
52. Yao J, Rock CO. Phosphatidic acid synthesis in bacteria. *Biochim Biophys Acta* (2013) 1831:495–502. doi:10.1016/j.bbalip.2012.08.018

Disclaimer: This work was prepared while SG was employed at George Washington University. The opinions expressed in this article are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Conflict of Interest Statement: 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.

Copyright © 2017 Lun, Samuel, Gillmor, Boyd and Smith. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Immunometabolic Phenotype Alterations Associated with the Induction of Disease Tolerance and Persistent Asymptomatic Infection of *Salmonella* in the Chicken Intestine

Michael H. Kogut^{1*} and Ryan J. Arsenault²

¹ USDA-ARS, SPARC, College Station, TX, USA, ² Department of Animal and Food Sciences, University of Delaware, Newark, DE, USA

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, USA

Reviewed by:

Guntram A. Grassl,
Hannover Medical School, Germany
Hosni M. Hassan,
North Carolina State University, USA

*Correspondence:

Michael H. Kogut
mike.kogut@ars.usda.gov

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 02 January 2017

Accepted: 15 March 2017

Published: 04 April 2017

Citation:

Kogut MH and Arsenault RJ (2017)
Immunometabolic Phenotype
Alterations Associated with the
Induction of Disease Tolerance and
Persistent Asymptomatic Infection of
Salmonella in the Chicken Intestine.
Front. Immunol. 8:372.
doi: 10.3389/fimmu.2017.00372

The adaptation of *Salmonella enterica* to the eukaryotic host is a key process that enables the bacterium to survive in a hostile environment. *Salmonella* have evolved an intimate relationship with its host that extends to their cellular and molecular levels. Colonization, invasion, and replication of the bacteria in an appropriate host suggest that modification of host functions is central to pathogenesis. Intuitively, this subversion of the cell must be a complex process, since hosts are not inherently programmed to provide an environment conducive to pathogens. Hosts have evolved countermeasures to pathogen invasion, establishment, and replication through two types of defenses: resistance and tolerance. Resistance functions to control pathogen invasion and reduce or eliminate the invading pathogen. Research has primarily concentrated on resistance mechanisms that are mediated by the immune system. On the other hand, tolerance is mediated by different mechanisms that limit the *damage* caused by a pathogen's growth without affecting or reducing pathogen numbers or loads. The mechanisms of tolerance appear to be separated into those that protect host tissues from the virulence factors of a pathogen and those that limit or reduce the damage caused by the host immune and inflammatory responses to the pathogen. Some pathogens, such as *Salmonella*, have evolved the capacity to survive the initial robust immune response and persist. The persistent phase of a *Salmonella* infection in the avian host usually involves a complex balance of protective immunity and immunopathology. *Salmonella* is able to stay in the avian ceca for months without triggering clinical signs. Chronic colonization of the intestinal tract is an important aspect of persistent *Salmonella* infection because it results in a silent propagation of bacteria in poultry stocks due to the impossibility to isolate contaminated animals. Data from our lab promote the hypothesis that *Salmonella* have evolved a unique survival strategy in poultry that minimizes host defenses (disease resistance) during the initial infection and then exploits and/or induces a dramatic immunometabolic reprogramming in the cecum that alters the host defense to disease tolerance. Unfortunately, this disease tolerance results in the ongoing human food safety dilemma.

Keywords: *Salmonella enterica*, chickens, disease resistance, disease tolerance, immunometabolism

INTRODUCTION

Salmonella Infection and Poultry

Foodborne illness is a significant worldwide public health problem that continues to plague the world, costing approximately \$152 billion annually (1). Despite control efforts that cost over a half a billion dollars annually, foodborne illnesses due to *Salmonella* and *Campylobacter* increased during the last 15 years. In 2013, 20% of the 9.4 million episodes of foodborne illnesses were attributed to *Salmonella* and accounted for 26% of the hospitalizations (2). In 2012, the Foodborne Diseases Active Surveillance Network found that *Salmonella* accounted for over 28% of the confirmed foodborne disease cases in the U.S., and cost U.S. residents \$14.6 billion annually, respectively (3). Clearly, efforts to elucidate and implement new and existing methods for control are well justified by the economic cost alone to control *Salmonella*, *Campylobacter*, and other foodborne pathogens. Poultry products have been associated frequently and consistently with the transmission of enteric pathogens, including *Salmonella* and *Campylobacter* (4).

Salmonellosis is a zoonotic disease caused by the Gram-negative facultative anaerobic, enteric bacterium *Salmonella*. With more than 2,500 serotypes having been described, most *Salmonella* serovars are not restricted to particular host species and are able to colonize the alimentary tract of animals without production of disease (5). Not coincidentally, the most common human clinical isolates, *Salmonella enterica* serotypes Typhimurium (STm) and Enteritidis (SE), are the most commonly detected serotypes in poultry (6).

In poultry, *S. enterica* serotypes can be divided into two groups based on their host species range and their disease pathogenesis (5, 7, 8) with *S. enterica* serotypes *S. Gallinarum* and *S. Pullorum* being chicken-specific and the broad host serovars best exemplified by STm and SE. STm and SE are major causes of zoonotic gastroenteritis in a wide range of host species worldwide (5–8).

Both broad host range *Salmonella* serovars (STm and SE) are able to colonize the gastrointestinal tract of chickens a few days of age without clinical disease but induce a rapid (within 4 h) and mild acute inflammatory response (9). After oral infection of fowl, the bacterial colonization is durable in the gut where the two ceca represent a suitable site for colonization. *Salmonella* can be transmitted horizontally within the flock after fecal shedding as well as vertically through the trans-ovarian route. Chicks are more susceptible to salmonellosis than adults. In particular, asymptomatic carriers have a major role in *Salmonella* propagation in poultry and hence in food contamination, since they cannot be easily isolated and identified. The persistence of *Salmonella* in the intestinal tract of chickens is the main cause of disease propagation in poultry (5, 8, 9).

PERSISTENCE OF *Salmonella* INTESTINAL COLONIZATION

Infection with a pathogenic microorganism usually results in the host responding by activating the innate and adaptive immune responses. However, some pathogens, such as *Salmonella*, have evolved the capacity to survive the initial robust immune response and persist (10–12). The persistent phase of infection usually

involves a complex balance of protective immunity and immunopathology. The interactions between the host and pathogen are very complex and likely reflect the coevolution and fine tuning of bacterial virulence mechanisms and host immune responses (13). Until recently, very little is known about the molecular regulatory interactions between the host immune response and virulence mechanisms that lead to *S. enterica* persistence in the avian intestine. The carrier state, corresponding to a persistent colonization of the gut, is established, and *Salmonella* is able to stay in the ceca for months without triggering clinical signs (14). Chronic colonization of the intestinal tract is an important aspect of persistent *Salmonella* infection because it results in a silent propagation of bacteria in poultry stocks due to the impossibility to isolate contaminated animals (15, 16).

The establishment of persistence is in the face of a substantial immune response requiring evasion or modulation of the response by the bacteria. The fact that many *Salmonella* serovars persist within the chicken intestinal tract with little sign of gastrointestinal disease despite eliciting a considerable inflammatory response and that inflammatory responses to *Salmonella* are relatively short-lived (17), strongly suggests there is a degree of regulation of this response.

HOST DEFENSE STRATEGIES

Historically, host defense strategy has been based on the outcome of the immune response's ability to detect and eliminate pathogens through multiple killing mechanisms known as host resistance (18, 19). However, a relatively new immunological concept, tolerance as a host defense strategy has been put forward (19, 20). Tolerance is the ability of the host to limit the damage caused by both the pathogen and the host immune response, i.e., immunopathology (20). Tolerance, as a host defense strategy, has been ignored in veterinary infectious disease studies (18). It is important to point out that infection tolerance is not immune tolerance which is defined as "unresponsiveness of the immune system to substances or tissue that has the capacity to elicit an immune response" (21).

Unlike immune responses that have measurable outputs to evaluate effectiveness, disease tolerance lacks clear-cut outputs (18). However, measurement of local cell metabolic processes and function, redox status, concentrations of metabolites, and organelle function of parenchymal cells and tissues (host's cells/tissues that do not have a direct impact on pathogens) would be beneficial in evaluating stress and damage responses. Since a pathogen and the induced immunopathology can theoretically affect any physiological system, disease tolerance would involve a number of processes that will reduce host susceptibility to damage. Therefore, any physiological mechanism that typically maintains homeostasis and functional integrity of host tissues could contribute to disease tolerance. Mechanistically, limiting tissue damage is regulated by a number of evolutionarily conserved stress and/or damage responses. These responses confer tissue damage control, by providing cellular adaptation to environmental changes (22). For example, stress responses maintain cellular functions by activating metabolic processes in response to local alterations in oxygen tension (hypoxia), redox status

(oxidative stress), osmolarity, and metabolite concentrations (ADP/ATP, glucose). All are essential mechanisms of cell and tissue homeostasis (23). Damage responses attempt to preserve cellular functions while minimizing damage to macromolecules (DNA, lipids, and proteins) and/or organelles (mitochondria, Golgi, and endoplasmic reticulum) (19, 23). The concept of tolerance as a host defense mechanism has led to an excellent recent editorial (24). The authors ask a very provocative question of what effect does therapeutics based on reducing the symptoms induced by a pathogen (tolerance) instead of reducing pathogen numbers have on evolution of the host population and the pathogen? By not reducing pathogen numbers, will there be an effect on pathogen transmission and spread and the potential development of disease carriers or will the limitation of disease symptoms allow the host immune system to concentrate on controlling pathogen numbers?

Salmonella-CHICKEN INFECTION BIOLOGY

Salmonella can be carried by poultry with virtually no ill effects on the host; whereas, in humans, the same bacteria cause pathological inflammation (25, 26). The induction of this severe inflammation appears to be essential for the salmonellae organisms to procure critical nutrients and respiratory substrates from the host allowing the pathogen to out-compete the commensal microbiota that rely on anaerobic fermentation (27–29). Thus, the interactions between the host response and *Salmonella* infections in the intestinal tract of poultry appear to be directed toward disease tolerance characterized by the asymptomatic nature of infection. Therefore, the chicken and bacteria appear to have evolved a relationship that minimizes both the normal host response and the normal bacterial virulence. However, this tolerant state is “detrimental to food safety” in humans (30).

In a recent review, Wigley (9) described how *Salmonella* infection in chickens facilitated our understanding of avian immunology over the last 20+ years. At the end of his review, Wigley (9) asked a “few key questions that still needed to be fully answered.” We have used two of the questions for the basis of our studies into the persistence of colonization of *Salmonella* in the intestine of chickens. Namely, (1) what mechanisms trigger the persistence of *Salmonella* in the cecum and (2) how is the intestinal response regulated to prevent excessive damage to the host? The *Salmonella*-chicken dynamics provide a unique system where the pathogen appears to evade the immune system, alters the local intestinal phylogeny of recognition and signaling pathways, and takes residence amongst the cecal microbiota.

Based on the findings by us and others, we propose that *Salmonella* infection in the chicken can be separated into three distinct stages of host defense strategies characterized by the cecal immune effector cells, immune gene expression, and immunometabolic responses at different times postinfection:

1. Stage 1, Disease Resistance: characterized by an acute heterophil-mediated pro-inflammatory response and anabolic metabolism 1–2 days postinfection.
2. Stage 2, Disease Tolerance: exemplified by a profound increase in cecal T regulatory cells and an anti-inflammatory response and a conversion to a catabolic phenotype 4 days postinfection.
3. Stage 3, Homeostasis: a return to a homeostatic metabolic phenotype with a more IL-10-mediated regulatory immune response 5–10 days postinfection.

STAGE 1, DISEASE RESISTANCE

Salmonella invasion of the chicken intestine induces an inflammatory process resulting in the expression of pro-inflammatory cytokines and chemokines by epithelial cells lining the intestine (17, 31–33). The outcome of this activation of innate immunity is a major influx of heterophils (granulocytes) to the intestine that limits bacterial invasion (34, 35) but does not lead to a pathological inflammation that is seen in humans (17, 36). However, this heterophil response does not have a significant protective response against the salmonellae bacteria that remain in the luminal side of the cecal epithelium. Interestingly, this inflammatory response is largely resolved by 3–4 days postinfection (35, 37, 38) characterized by the reduction of pro-inflammatory cytokines mRNA transcription in the cecum to non-infected control levels yet *Salmonella* can persist in the intestine and be shed in the feces for several weeks (17).

Accompanying intestinal inflammation are extreme alterations in tissue metabolism, most of which are due to the incoming heterophils and other inflammatory cells and can include the increase in fatty acid, protein synthesis, glycolysis and the production of reactive oxygen intermediates (38, 39). Energy-demanding processes, such as migration, phagocytosis, and the generation of an oxidative burst, that accompany the recruitment of the heterophils to the site of infection, trigger transcriptional and translational changes in tissue phenotype (predominately the metabolic signaling pathway of mTOR phosphorylation) that shifts fundamental changes to the local intestinal tissue to anabolic metabolism (37–39). Further, the presence of the PMNs and subsequent metabolic requirements exhaust the microenvironmental oxygen to quantities nearing anoxia (39). This localized oxygen depletion leads to the stabilization, and thus the activation of the transcription factor, hypoxia-inducible factor- α (HIF1 α), and activation of the HIF1 α signaling pathway that resolves inflammation, and potentially provides a more tolerant local setting for the bacteria (40, 41). Under these oxygen-deprived conditions, HIF1 α activation inhibits mTOR activity resulting in a potent anti-inflammatory microenvironment through the production and stimulation of T regulatory cells would regulate tissue damage (42, 43).

The initial inflammatory response (disease resistance) is sufficient to help control invasion and elicit the development of a protective acquired immune response that can lead to systemic and eventual clearance of gastrointestinal infection.

STAGE 2, DISEASE TOLERANCE

Immunological Phenotype

It has been demonstrated by numerous groups that early cecal pro-inflammatory (disease resistance) signals following initial

infection with STm or SE was dramatically downregulated 2–4 days after infection that is linked with the development of an anti-inflammatory, Th2 response (15, 17, 32, 34, 44) to increased expression of IL-10 and TGF- β , which suggests the end of the disease resistance and the start of a disease tolerant state were being initiated.

It would seem likely that regulation of inflammatory immune responses, presumably by regulatory T cells (Tregs), allows *Salmonella* to persist within the gut for a number of weeks without disease to the bird. Such a “tolerogenic” response would have little or no impact on the bird itself but has public health consequences in allowing persistence for several weeks, particularly given broiler chickens are typically slaughtered at around 5 weeks of age. Subsequently, we have found an expansion of the CD4+ CD25+ T cell (Treg) population in the cecum of *Salmonella*-infected chickens (45). Functionally, the cecal Tregs had increased suppressive activity for T effector cells and had a profound increase in IL-10 mRNA transcription. In the murine model of ST infection, the ability of the bacteria to persist or be cleared has been found to be dependent on the presence and function of Tregs (46).

Mechanistically, in a series of experiments using a chicken-specific kinome array, the plasticity of the local cecal immune phenotype where the initial inflammatory response against a *Salmonella* infection is then followed by a striking alteration in the immune microenvironment 2 days later during the establishment of a persistent *Salmonella* infection (35, 37, 38). We used the power of a species-specific kinome array to delineate the mechanisms that alter the host avian inflammatory responses and uncover host signaling events that are manipulated by the bacteria in order to establish a persistent infection. First, we found that the establishment of a persistent *Salmonella* cecal colonization in chickens activates both the canonical (Smad-dependent) and non-canonical (Smad-independent) TGF- β signaling pathways (35). TGF- β functions by controlling immune responses by suppressing non-Treg function and promoting Treg function. These results are suggestive of a change in the cecal mucosal phenotype from pro-inflammatory to tolerance is, in part, mediated by the increased expression of TGF- β that activates both Smad-dependent and -independent TGF- β pathways that increases the differentiation and function of Tregs while decreasing the function of pro-inflammatory immune cells. Second, during the establishment of a persistent *Salmonella* cecal infection, we found the activation of the non-canonical Wnt signaling pathways (35). Non-canonical Wnt signaling controls nuclear localization of nuclear factor of activated T cell (NFAT) transcriptional factor. NFAT regulates the interaction of the innate immune cells with acquired immunity to promote anti-inflammatory programs and is essential for both development and function of Tregs (47, 48). The transformation in the avian host response from resistance to tolerance during the establishment of *Salmonella* persistence was further confirmed by a study showing two select host immune signaling pathways were altered; namely, the T cell receptor and JAK-STAT signaling pathways (38). Both signaling pathways were shown to have alterations in the phosphorylation of multiple peptides that resulted in the inactivation of an active immune response in the local cecal environment. The response

was characterized by the dephosphorylation of phospholipase c- γ 1 that induced the dephosphorylation (inhibits activation) of NF- κ B signaling, thus preventing activation of immune response genes and the phosphorylation of NFAT signaling which activates anti-inflammatory cytokine production as described above. Further, interferon-gamma production that is central in the resolution of *Salmonella* infections in the cecum of avian species (16, 32, 49) was also found to be inhibited in the cecum of SE-infected chickens through the disruption of the JAK-STAT signaling pathway (dephosphorylation of JAK2, JAK3, and STAT4). The JAK-STAT signaling pathway transmits information from extracellular chemical signals to the nucleus resulting in DNA transcription and expression of genes involved in immunity, proliferation, differentiation, and apoptosis (50, 51). Taken together, by 4 days postinfection, the immune phenotype in the cecum of *Salmonella*-infected chickens has undergone a dramatic alteration in host responsiveness where the host does not appear to recognize the bacterium as a pathogen resulting in a persistent cecal colonization.

Metabolic Phenotype

Concurrently to the alterations in the local immune response during the tolerance phase, profound metabolic phenotype alterations occurred in the cecal tissue of *Salmonella*-infected chickens from the early resistance response (4–48 h postinfection) which is pro-inflammatory, fueled by glycolysis and mTOR-mediated protein synthesis to the later tolerance phase (4 days postinfection) where the local environment has undergone an immune-metabolic reprogramming to an anti-inflammatory state driven by adenosine monophosphate-activated protein kinase (AMPK)-directed oxidative phosphorylation (37). Therefore, metabolism appears to provide a potential measurement that characterizes a state of infection tolerance. Additionally, these results provide further evidence of what Olive and Sassetti (52) describe as a pathogen’s ability to “sense the metabolic environment of the host, adapting to changing nutrient availability.” Further, these phosphorylation alterations at the gut level during the first 3 weeks after infection of day-old broilers with ST appear to lead to key metabolic changes that affected fatty acid and glucose metabolism through the 5'-AMPK and the insulin/mTOR signaling pathway in the *skeletal muscle* were altered (53). Supplemental proof for the effects of fatty acid and glucose metabolism on long-term persistence of *Salmonella* was recently demonstrated using the murine macrophage model (54, 55). ST preferred living in alternatively activated macrophages that require the activation of the transcription factor, peroxisome proliferator-activating receptor δ (PPAR δ), which regulates fatty acid metabolism (54). Thus, the bacteria prefer macrophages that employ oxidative metabolism for energy instead of glycolysis due to the factor that disruption of glycolysis is a signal of the activation of the NLRP3 inflammasome and the subsequent initiation of inflammatory cell death, pyroptosis (55).

STAGE 3, HOMEOSTASIS

Immunologically, the third stage of an avian *Salmonella* infection occurs shortly after day 4 postinfection with the expression of a

disease tolerance state. The number of Tregs in the cecum of the infected birds remains constant suggesting an immune regulation state further evidenced by the increased transcription of IL-10 and TGF- β (37, 38, 44, 45). The underlying question here is whether *Salmonella* is no longer “sensed” by the immune system as foreign invaded and has become a component of the cecal microbiome. Experiments to answer this question are ongoing in our laboratories.

Metabolically, the local microenvironment appears to go through a final reprogramming during this third stage of infection moving from a catabolic state in stage 2 to a more homeostatic status. This was verified in our kinome studies by the fact that we observed no differences in the metabolic signaling pathways in the ceca from the *Salmonella*-infected and non-infected chickens (37, 38, 44).

PERSPECTIVE

The data from our lab and others soundly support the hypothesis that *Salmonella* have evolved a unique survival strategy in poultry that minimizes host defenses (disease resistance) during the initial infection and then exploits and/or induces a dramatic immunometabolic reprogramming in the cecum that alters the host defense to disease tolerance (summarized in **Table 1**). The ability to induce a state of disease tolerance is unique to the poultry-*Salmonella* interactome in that it allows the bacterium to establish a long-term persistent infection in the cecum while allowing the host to control disease pathology. Unfortunately, it also results in the ongoing human food safety dilemma. It should be pointed out that the energy balance reported in **Table 1** is not backed by direct evidence in these experiments but is an assumption based on the fact that AMP is elevated when AMPK is activated and ATP is elevated when mTOR is activated.

These studies have used the emerging field of immunometabolism at the tissue level to identify potential mechanisms by which the host can tolerate a *Salmonella* infection. Recently, an immunometabolic mechanism for disease tolerance to a murine STm infection was found to involve the microbiome and the insulin-signaling pathway (56). Taken together, identifying potential

TABLE 1 | Immunometabolic alterations in the chicken cecum during the establishment of a persistent infection.

Parameter	Stage 1 Days 1–2 postinfection Disease resistance	Stage 2 Days 3–4 postinfection Disease tolerance	Stage 3 Days 5–28 postinfection Homeostasis
Cytokines	Pro-inflammatory (17, 31–33)	IL-10, TGF- β (17, 32, 44)	IL-17, IL-10
Immune cells	Heterophils (34, 35), M1 macrophages (55)	Th2 T-helper cells (15), regulatory T cells (Tregs) (45), M2 macrophages (54)	Tregs, APCs
Signaling	Toll-like receptor, NOD receptor, HIF1 (40, 41), NF- κ B (38)	Smad, Wnt (35), nuclear factor of activated T cell (47, 48), JAK-STAT (38)	Maintenance
Metabolism	Increased fatty acid synthesis, protein synthesis, glycolysis, reactive oxygen intermediates (38, 39)	Oxidative phosphorylation, fatty acid catabolism (53)	Aerobic
Metabolic signaling	mTOR (37–39)	Adenosine monophosphate-activated protein kinase (37)	Energy neutral
Energy balance	AMP:ATP	AMP:ATP	AMP:ATP
Oxygen state	Hypoxia (43)	Normoxia	Normoxia
Metabolic state	Anabolic	Catabolic	Balanced
Immune state	Inflammation	Anti-inflammatory (tolerance)	Non-inflammatory

molecular mechanisms of disease tolerance as a host defense can not only “provide a perspective into the evolutionary forces that have driven coevolution” (56) of host–pathogen interactions but also provide the discovery of new therapeutic targets to control foodborne pathogens.

AUTHOR CONTRIBUTIONS

MK and RA conducted the experiments and made substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

- Scharff RL. Health related cost from foodborne illness in the United States. *Produce Safety Project*. Georgetown University (2010).
- Crim SM, Iwamoto M, Huang JY, Griffin PM, Gilliss D, Cronquist AB, et al. Incidence and trends of infection with pathogens transmitted commonly through food – Foodborne Diseases Active Surveillance Network, 10 U.S. sites, 2006–2013. *MMWR Morb Mortal Wkly Rep* (2014) 63:328–32.
- CDC. *Foodborne Diseases Active Surveillance Network (FoodNet): FoodNet Surveillance Report for 2012*. Atlanta, GA: United States Department of Health and Human Services, CDC (2014).
- Vanderplas S, Dubois-Duphin R, Beckers V, Thonart P, Thewis A. *Salmonella* in chicken: current and developing strategies to reduce contamination at farm level. *J Food Prot* (2010) 73:774–85. doi:10.4315/0362-028X-73.4.774
- Barrow PA. The paratyphoid *Salmonellae*. *Rev Sci Tech* (2000) 19:351–75. doi:10.20506/rst.19.2.1225
- Stevens MP, Humphrey TJ, Maskell DJ. Molecular insight into farm animal and zoonotic *Salmonella* infections. *Philos Trans R Soc Lond B Biol Sci* (2009) 364:2709–23. doi:10.1098/rstb.2009.0094
- Barrow PA, Freitas Neto OC. Pullorum disease and fowl typhoid-new thoughts on an old disease: a review. *Avian Pathol* (2011) 40:1–13. doi:10.1080/03079457.2010.542575
- Barrow PA, Bumstead N, Marston K, Lovell MA, Wigley P. Fecal shedding and intestinal colonization of *Salmonella enterica* in in-bred chickens: the effect of host-genetic background. *Epidemiol Infect* (2004) 132:117–26. doi:10.1017/S0950268803001274
- Wigley P. *Salmonella enterica* in the chicken: how it has helped our understanding of immunology in a non-biomedical model species. *Front Immunol* (2014) 5:482. doi:10.3389/fimmu.2014.00482
- Ribet D, Cossart P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect* (2015) 17:173–83. doi:10.1016/j.micinf.2015.01.004
- Byndios MX, Tsoilis RM. Chronic bacterial pathogens: mechanisms of persistence. *Microbiol Spectr* (2016) 4:515–28. doi:10.1128/microbiolspec.VMBF-0020-2015
- Monack DM. *Helicobacter* and *Salmonella* persistent infection strategies. *Cold Spring Harb Perspect Med* (2013) 3:a10348. doi:10.1101/cshperspect.a10348

13. Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, et al. Comparative genome analysis of *Salmonella enteritidis* PT4 and *Salmonella gallinarum* 287/19 provides insights into evolutionary and host adaptation pathways. *Genome Res* (2008) 18:1624–37. doi:10.1101/gr077404.108
14. Barrow PA, Huggins MB, Lovell MA, Simpson JM. Observations on the pathogenesis of experimental *Salmonella typhimurium* infection in chickens. *Res Vet Sci* (1987) 42:194–9.
15. Chausse A-M, Grepinet O, Bottreau E, Robert V, Hennequet-Antier C, Lalmanach A-C, et al. Susceptibility to *Salmonella* carrier-state: a possible Th2 response in susceptible chicks. *Vet Immunol Immunopathol* (2014) 159:16–28. doi:10.1016/j.vetimm.2014.03.001
16. Sadeyn J-R, Trottereau J, Velge P, Marty J, Beaumont C, Barrow PA, et al. *Salmonella* carrier state in chicken: comparison of immune response genes between susceptible and resistant animals. *Microbes Infect* (2004) 6:1278–86. doi:10.1016/j.micinf.2004.07.005
17. Withange GS, Wigley P, Kaiser P, Mastroeni P, Brooks H, Powers C, et al. Cytokine and chemokine responses associated with clearance of a primary *Salmonella enterica* serovars Typhimurium infection in the chicken and in protective immunity to challenge. *Infect Immun* (2005) 73:173–82. doi:10.1128/IAI.73.8.5173-5182.2005
18. Schneider DS, Ayres JS. Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. *Nat Rev Immunol* (2008) 8:889–95. doi:10.1038/nri2432
19. Medzhitov R, Schneider DS, Soares MP. Disease tolerance as a defense strategy. *Science* (2012) 335:936–41. doi:10.1126/science.1214935
20. Ayres JS, Schneider DS. Tolerance of infections. *Annu Rev Immunol* (2012) 30:271–94. doi:10.1146/annurev-immunol-020711-075030
21. Suzuki J, Camill R, Zhibin C. Immune tolerance induction by integrating innate and adaptive immune regulators. *Cell Transplant* (2010) 19:253–68. doi:10.3727/096368909X480314
22. Hayes JD, Dinkova-Kostova AT. The Nrf2 regulatory network provides an interface between redox and intermediary metabolism. *Trends Biochem Sci* (2014) 19:199–218. doi:10.1016/j.tibs.2014.02.002
23. Soares MP, Ribiero AM. Nrf2 as a master regulator of tissue damage control and disease tolerance to infection. *Biochem Soc Trans* (2015) 43:663–8. doi:10.1042/BST20150054
24. Doeschl-Wilson AB, Kyriazakis I. Should we aim for genetic improvement in host resistance or tolerance to infectious pathogens? *Front Genet* (2012) 3:272. doi:10.3389/fgene.2012.00272
25. Galan JE. *Salmonella* interactions with host cells: type III secretion at work. *Annu Rev Cell Dev Biol* (2001) 17:53–86. doi:10.1146/annurev.cellbio.17.1.53
26. Figueira R, Holden D. Functions of the *Salmonella* pathogenicity island 2 (SPI-2) type III secretion system effectors. *Microbiology* (2012) 158:1147–61. doi:10.1099/mic.0.058115-0
27. Stechler B, Robbiani R, Walker M, Westendorf A, Barthel M, Kremer M, et al. *Salmonella enterica* serovars Typhimurium exploits inflammation to compete with the intestinal microbiota. *PLoS Biol* (2007) 5:2177–89. doi:10.1371/journal.pbio.0050244
28. Winter SE, Thienemitt P, Winter MG, Butler BP, Huseby DL, Crawford RW, et al. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* (2010) 467:426–9. doi:10.1038/nature09415
29. Behnsen J, Perez-Lopez A, Nuccio S-P, Raffatellu M. Exploiting host immunity: the *Salmonella* paradigm. *Trends Immunol* (2015) 36:112–20. doi:10.1016/j.it.2014.12.003
30. Calenge F, Beaumont C. Toward integrative genomics study of genetic resistance to *Salmonella* and *Campylobacter* intestinal colonization. *Front Genet* (2012) 3:261. doi:10.3389/fgene.2012.00261
31. Withange GS, Kaiser P, Wigley P, Powers C, Mastroeni C, Brooks H, et al. Rapid expression of chemokines and proinflammatory cytokines in newly hatched chickens infected with *Salmonella enterica* serovar Typhimurium. *Infect Immun* (2004) 72:2152–9. doi:10.1128/IAI.72.4.2152-2159.2004
32. Setta AM, Barrow PA, Kaiser P, Jones MA. Early immune dynamics following infection with *Salmonella enterica* serovars Enteritidis, Infantis, Pullorum, and Gallinarum: cytokine and chemokine gene expression profile and cellular changes of chicken cecal tonsils. *Comp Immunol Microbiol Infect Dis* (2012) 35:397–410. doi:10.1016/j.cimid.2012.03.004
33. Matulova M, Varmuzova K, Sisak F, Havlickova H, Babak V, Stejskal K, et al. Chicken innate immune response to oral infection with *Salmonella enterica* serovar enteritidis. *Vet Res* (2013) 44:37. doi:10.1186/1297-9716-44-37
34. Kogut MH, Tellez GI, McGruder ED, Hargis BM, Williams JD, Corrier DE, et al. Heterophils are decisive components in the early responses of chickens to *Salmonella enteritidis* infections. *Microb Pathog* (1994) 16:141–51. doi:10.1006/mpat.1994.1015
35. Kogut MH, Chiang HI, Swaggerty CL, Pevzner IY, Zhou H. Gene expression analysis of toll-like receptor pathways heterophils from genetic chicken lines that differ in their susceptibility to *Salmonella enteritidis*. *Front Genet* (2012) 3:121. doi:10.3389/fgene.2012.00121
36. Foster N, Lovell MA, Marston KL, Hulme SD, Frost AJ, Bland P, et al. Rapid protection of gnotobiotic pigs against experimental salmonellosis following induction of polymorphonuclear leukocytes by avirulent *Salmonella enterica*. *Infect Immun* (2003) 71:2182–91. doi:10.1128/IAI.71.4.2182-2191.2003
37. Kogut MH, Genovese KJ, He H, Arsenault RJ. AMPK and mTOR: sensors and regulators of immunometabolic changes during *Salmonella* infection in the chicken. *Poult Sci* (2016) 95:345–53. doi:10.3382/ps/pev349
38. Kogut MH, Swaggerty CL, Byrd JA, Selvaraj R, Arsenault RJ. Chicken-specific kinome array reveals that *Salmonella enterica* serovars enteritidis modulates host immune signaling pathways in the cecum to establish a persistent infection. *Int J Mol Sci* (2016) 17:1207. doi:10.3390/ijms17081207
39. Campbell EL, Bruyninck W, Kelly CJ, Glover LE, McNamee EN, Bowers BE, et al. Transmigrating neutrophils shape the mucosal microenvironment through localized oxygen depletion to influence resolution of inflammation. *Immunity* (2014) 40:66–77. doi:10.1016/j.immuni.2013.11.020
40. Campbell EL, Kao DJ, Colgan SP. Neutrophils and the inflammatory tissue microenvironment in the mucosa. *Immunol Rev* (2016) 273:112–20. doi:10.1111/imr.12456
41. Campbell EL, Colgan SP. Neutrophils and inflammatory metabolism in antimicrobial functions of the mucosa. *J Leukoc Biol* (2015) 98:517–22. doi:10.1189/jlb.3MR1114-556R
42. Wouters BG, Koritzinsky M. Hypoxia signaling through mTOR and the unfolded protein response in cancer. *Nat Rev Cancer* (2008) 8:851–64. doi:10.1038/nrc2501
43. Clambet ET, McNamee EN, Westrich JA, Glover LE, Campbell CL, Jedlicka P, et al. Hypoxia-inducible factor-1 alpha-dependent induction of FoxP3 drives regulatory T-cell abundance and function during inflammatory hypoxia of the mucosa. *Proc Natl Acad Sci U S A* (2012) 109:E2784–93. doi:10.1073/pnas.1202366109
44. Kogut MH, Arsenault RL. A role for the non-canonical Wnt- β -catenin and TGF- β signaling pathways in the induction of tolerance during the establishment of a *Salmonella enterica* serovar enteritidis persistent cecal infection in chickens. *Front Vet Sci* (2015) 2:33. doi:10.3389/fvets.2015.00033
45. Shanmugasundaram R, Kogut MH, Arsenault RJ, Swaggerty CL, Cole KY, Reddish MJ, et al. Effect of *Salmonella* infection on cecal tonsil regulatory T cell properties in chickens. *Poult Sci* (2015) 94:1828–35. doi:10.3382/ps/pev161
46. Johanns TM, Erelt JM, Rowe JH, Way SS. Regulatory T cell suppressive potency dictates the balance between bacterial proliferation and clearance during persistent *Salmonella* infection. *PLoS Pathog* (2010) 6:e1001043. doi:10.1371/journal.ppat.1001043
47. Zanon I, Granucci F. Regulation and dysregulation of innate immunity by NFAT signaling downstream of pattern recognition receptors (PRRs). *Eur J Immunol* (2012) 42:1924–31. doi:10.1002/eji.201242580
48. Oh-hora M, Rao A. The calcium/NFAT pathway: role in development and function of regulatory T cells. *Microbes Infect* (2009) 11:612–9. doi:10.1016/j.micinf.2009.04.008
49. Kogut MH, Rothwell L, Kaiser P. IFN-gamma priming of chicken heterophils upregulates the expression of pro-inflammatory and Th1 cytokine mRNA following receptor-mediated phagocytosis of *Salmonella enterica* serovar enteritidis. *J Interferon Cytokine Res* (2005) 25:73–81. doi:10.1089/jir.2005.25.73
50. Heim MH. The JAK-STAT pathway: cytokine signaling from the receptor to the nucleus. *J Recept Signal Transduct Res* (1999) 19:75–120. doi:10.3109/10799899909036638
51. Murray PJ. The JAK-STAT signaling pathway: input and output integration. *J Immunol* (2007) 178:2623–39. doi:10.4049/jimmunol.178.5.2623

52. Olive AJ, Sassetti CM. Metabolic crosstalk between host and pathogen: sensing, adapting and competing. *Nat Rev Microbiol* (2016) 14:221–34. doi:10.1038/nrmicro.2016.12
53. Arsenault RJ, Napper S, Kogut MH. *Salmonella enterica* serotype Typhimurium infection causes metabolic changes in chicken muscle involving AMPK, fatty acid and insulin/mTOR signaling. *Vet Res* (2013) 44:35–50. doi:10.1186/1297-9716-44-35
54. Eisele NA, Ruby T, Jacobson A, Manzanillo PS, Cox JS, Lam L, et al. *Salmonella* require the fatty acid regulator PPAR α for the establishment of a metabolic environment essential for long-term persistence. *Cell Host Microbe* (2013) 14:171–82. doi:10.1016/j.chom.2013.07.010
55. Sanman LE, Cian Y, Eisele NA, Ng TM, van der linden WA, Monack DM, et al. Disruption of glycolytic flux is a signal for inflammasome signaling and pyroptotic cell death. *Elife* (2016) 5:e13663. doi:10.7554/eLife.13663
56. Schieber AMP, Lee YM, Chang MW, LeBlanc M, Collins B, Downes M, et al. Disease tolerance mediated by microbiome *E. coli* involves inflammasome and IGF-1 signaling. *Science* (2015) 350:558–62. doi:10.1126/science.aac6468

Conflict of Interest Statement: 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.

Copyright © 2017 Kogut and Arsenault. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evidence of an Antimicrobial Peptide Signature Encrypted in HECT E3 Ubiquitin Ligases

Ivan Lavander Candido-Ferreira^{1,2*}, Thales Kronenberger^{3†}, Raphael Santa Rosa Sayegh^{1,4†}, Isabel de Fátima Correia Batista⁵ and Pedro Ismael da Silva Junior^{1*}

¹ Special Laboratory for Applied Toxinology (LETA), Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Butantan Institute, São Paulo, São Paulo, Brazil, ² Biosciences Institute, University of São Paulo, São Paulo, São Paulo, Brazil, ³ Department of Parasitology, Biomedical Sciences Institute, University of São Paulo, São Paulo, São Paulo, Brazil, ⁴ Department of Biochemistry, Institute of Chemistry, University of São Paulo, São Paulo, São Paulo, Brazil, ⁵ Laboratory of Biochemistry and Biophysics, Butantan Institute, São Paulo, São Paulo, Brazil

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, USA

Reviewed by:

Lydia E. Matesic,
University of South Carolina, USA
Yunhao Tan,
Harvard Medical School, USA

*Correspondence:

Ivan Lavander Candido-Ferreira
ivan.lavander.ferreira@usp.br;
Pedro Ismael da Silva Junior
pisjr@butantan.gov.br

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 19 August 2016

Accepted: 16 December 2016

Published: 09 January 2017

Citation:

Candido-Ferreira IL, Kronenberger T,
Sayegh RSR, Batista IFC
and da Silva Junior PI (2017)
Evidence of an Antimicrobial
Peptide Signature Encrypted in
HECT E3 Ubiquitin Ligases.
Front. Immunol. 7:664.
doi: 10.3389/fimmu.2016.00664

The ubiquitin-proteasome pathway (UPP) is a hallmark of the eukaryotic cell. In jawed vertebrates, it has been co-opted by the adaptive immune system, where proteasomal degradation produces endogenous peptides for major histocompatibility complex class I antigen presentation. However, proteolytic products are also necessary for the phylogenetically widespread innate immune system, as they often play a role as host defense peptides (HDPs), pivotal effectors against pathogens. Here, we report the identification of the arachnid HDP oligoventin, which shares homology to a core member of the UPP, E3 ubiquitin ligases. Oligoventin has broad antimicrobial activity and shows strong synergy with lysozymes. Using computational and phylogenetic approaches, we show high conservation of the oligoventin signature in HECT E3s. *In silico* simulation of HECT E3s self-proteolysis provides evidence that HDPs can be generated by fine-tuned 26S proteasomal degradation, and therefore are consistent with the hypothesis that oligoventin is a cryptic peptide released by the proteolytic processing of an Nedd4 E3 precursor protein. Finally, we compare the production of HDPs and endogenous antigens from orthologous HECT E3s by proteasomal degradation as a means of analyzing the UPP coupling to metazoan immunity. Our results highlight the functional plasticity of the UPP in innate and adaptive immune systems as a possibly recurrent mechanism to generate functionally diverse peptides.

Keywords: HECT ligases, host defense, immune evolution, innate immunity, Nedd4, synergy, ubiquitination, ubiquitin-proteasome system

INTRODUCTION

The ubiquitin-proteasome pathway (UPP) is central to the eukaryotic cell, being involved virtually in every intracellular pathway, including protein posttranslational modifications, fine-tuned proteolysis, autophagy, cell cycle regulation, programmed cell death, cell signaling, transcriptional regulation, gene expression, protein and mRNA turn over, cancer development, viral budding, and immune evasion by pathogens (1–10). Precise posttranslational modifications of proteins by ubiquitin or ubiquitin-like proteins involve the multistep, hierarchical transfer of ubiquitin to a substrate by ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2), and ubiquitin-protein ligases (E3) (1–10). Repeating this process generates polyubiquitin chains, which function as a

signal for degradation via the 26S proteasome (1–10). In addition to the temporal, spatial, and context-specific regulation of the UPP, the large number of UPP targets is tightly regulated by the specificity-conferring components, E3s (3–7). As part of jawed vertebrates' immune systems, E3s specify host defense signal transduction pathways, transcriptional regulation, and targeted proteolysis of cytosolic proteins for production of endogenous antigens, which are then presented to cytotoxic T cells mediated by major histocompatibility complex (MHC) class I receptors (1–10). Remarkably, bacteria have also evolved E3 ligases mimicking eukaryotic ones as an adaption to evade host defense (10–15).

In contrast to the adaptive immune system, which is restricted to jawed vertebrates and is based on humoral and cellular responses with specificity for antigens (1–10, 16), the phylogenetically widespread and more ancient innate defense relies on pattern recognition molecules conferring specificity against pathogens, complex signaling pathways leading to phagocytosis, encapsulation, and production of effector molecules with broad activity against pathogens (3, 4, 16–21). Recently, emerging roles in the innate immune system have been attributed to ubiquitin (22–29). Ubiquitin degradation produces host defense peptides (HDPs) (22–28), pivotal players in the innate defense against microbial pathogens (17–19). Additionally, more than 70 immunosuppressive peptides have been discovered originating from ubiquitin degradation *in vitro* (22, 30, 31). However, production of antimicrobial and immunomodulatory peptides is a complex, multilayered process (17–19, 22, 26, 28, 30, 31): it involves the canonical expression of transcriptionally regulated gene-encoded peptides (17–19) and fast production of cryptic peptides [that is, release of protein fragments with distinct properties from that of the original protein (19, 22, 26, 28)]. Therefore, the UPP orchestrates biologically active peptide production in both the innate and the adaptive metazoan immune systems by generating the recently discovered ubiquitin-encrypted HDPs (22–28), E3-mediated transcriptional regulation of host defense gene expression (3, 4, 6, 7, 10), and production of short fragments for MHC class I-mediated antigen presentation (3, 4, 8, 9, 16).

However, ubiquitin is highly constrained, with only three amino acids varying between yeast and human primary sequences (1–7, 22). Such constraint greatly reduces its potential to generate diverse HDPs. In contrast, E3s are larger, highly diversified, evolutionarily less conserved, and hold high affinity toward their molecular targets (1–7), which probably makes them a more enriched substrate than ubiquitin for encrypted HDPs. Surprisingly, little is known about the physiological roles that proteasomal degradation-derived fragments play, except for MHC class I antigens (3, 4, 8, 9, 16). Here, we hypothesize that similar to ubiquitin, E3 ligases harbor a defensive arsenal that can be released by fine-tuned proteolysis and thereby represent a novel scaffold for cryptic peptide discovery.

Among organisms that lack the adaptive immune system, arachnids' innate defenses are remarkably enriched for HDPs (19–21). Identification of many classes of antimicrobial peptides that appeared early in evolution (17–21), such as glycine-rich peptides (19, 32, 33), tachyplesin-like HDPs (19, 34), defensins (19–21), lysozymes (20), and hemocyanin-derived antifungal

peptides (19, 35) are consistent with their phylogenetic position at the base of extant arthropods' phylogeny (19, 20, 34–36). Thus, investigating the arachnid innate immune peptidome is a promising approach to identify ancient host defense effectors, including cryptic HDPs.

Here, we describe the discovery and characterization of oligoventin, an arachnid HDP isolated from the eggs of the Brazilian armed spider *Phoneutria nigriventer* (Ctenidae, Araneomorphae). Based on bioinformatics analysis, we suggest that oligoventin is a cryptic peptide derived from the proteasomal degradation of E3s. Bayesian phylogenetic analysis indicates that oligoventin appeared early in evolution and its production is likely not restricted to arachnids. Moreover, oligoventin inhibits growth of yeast, Gram-positive and Gram-negative bacteria and also exhibits synergy with lysozymes against *Micrococcus luteus* A270, consistent with the proposed function as a host defense effector. Furthermore, investigating mouse and human immune epitopes derived from the proteasomal degradation of E3s uncovered eight sequences, which are indeed involved in MHC class I antigen presentation. Our results provide, to our knowledge, the first evidence that proteasomal-mediated protein degradation evolved independently to produce functional short-sized peptides in the adaptive immunity of jawed vertebrates and possibly in the innate defense of arachnids, thus highlighting a recurrent role of the UPP in the generation of functionally diverse peptides in metazoan immune systems.

MATERIALS AND METHODS

Animals

Adult female *P. nigriventer* spiders laid eggs in captivity. Eggs were separated from silk and stored at -20°C for later use. These animals were collected under license Permanent Zoological Material no. 11024-3-IBAMA and Special Authorization for Access to Genetic Patrimony no. 001/2008.

Microorganisms

Fungal and bacterial strains were obtained from various sources. *Escherichia coli* SBS363 and *M. luteus* A270 were from the Pasteur Institute (Paris); *Candida albicans* (MDM8) was from the Department of Microbiology from the University of São Paulo (Brazil); *E. coli* ATCC 25922, *Pseudomonas aeruginosa* ATCC 27853, *Serratia marcescens* ATCC4112, *Staphylococcus aureus* ATCC 29213, and *Staphylococcus epidermidis* ATCC 12228 were from the American Type Culture Collection (ATCC). The following human clinical yeast isolates, which can be agents of candidiasis, obtained from the Oswaldo Cruz Institute (Brazil) were also used: *Trichosporon* sp. IOC 4569, *Candida krusei* IOC 4559, *Candida glabrata* IOC 4565, *C. albicans* IOC 4558, *Candida parapsilosis* IOC 4564, *Candida tropicalis* IOC 4560, and *Candida guilliermondii* IOC 4557.

Activity-Guided Isolation of Host Defense Effectors from *P. nigriventer* Eggs

Purification of antimicrobials was carried out following the strategies described elsewhere (37, 38). In brief, eggs were

suspended in 20 mL of glacial acetic acid and homogenized. The insoluble material was removed by centrifugation at 16,000 g for 30 min. The supernatant was partially purified by applying it in two Sep-Pak C₁₈ (Light tC₁₈—Water Associates) cartridges connected in series equilibrated in 0.05% trifluoroacetic acid, and protein-concentrated fractions were eluted in three steps using 5, 40, and 80% of acetonitrile in acidified water. Only the protein-concentrated fraction eluted in 40% acetonitrile was directly used for HPLC purifications. Fractionation was carried out using a reversed-phase high-performance liquid chromatography (RP-HPLC) semipreparative C₁₈ column (Jupiter, 10 × 250 mm), equilibrated in 2% acetonitrile, and 0.05% trifluoroacetic acid. Elution was successfully performed with a linear 2–60% gradient of solution B [0.10% (v/v) trifluoroacetic acid in acetonitrile] in acidified water {solution A [0.05% (v/v) trifluoroacetic acid in water]} run for 60 min at a flow rate of 1.5 mL/min. Effluent absorbance was monitored at 225 nm. Fractions with antimicrobial activity were further purified using an analytical Jupiter C₁₈ column (250 mm × 4.6 mm) at a flow rate of 1.0 mL/min in 60 min with distinct gradients: from 33.5 to 43.5% of acetonitrile in acidified water for fraction containing the HDP of 1.4 kDa (Figure S1A in Supplementary Material), 22.5 to 32.5% for the oligoventin-enriched fraction (Figure 1B), 27.5 to 37.5% for the anti-*M. luteus* lysozyme (Figure S1C in Supplementary Material), and 34.5 to 44.5% for the anti-*C. albicans* lysozyme (Figure S1D in Supplementary Material). A symmetrical peak on the HPLC system, amino acid sequencing, and mass spectrometry analysis ascertained the purity of the peptide or protein. Fractions were lyophilized in a SpeedVac Concentrator.

Growth Inhibition Assays

During the purification procedure, antimicrobial activities were detected or monitored by liquid growth inhibition assays as described in Ref. (38–40), using the Gram-negative bacteria *E. coli* SBS363 and Gram-positive bacteria *M. luteus* A270 that were cultured in poor broth nutrient medium (PB: 1.0 g peptone in 100 mL of water containing 86 mM NaCl at pH 7.4; 217 mOsm), and the yeast strain *C. albicans* MDM8, cultured in poor dextrose broth (1/2 PDB: 1.2 g potato dextrose in 100 mL of H₂O at pH 5.0; 79 mOsm) used at half strength. Determination of antimicrobial activity was performed using fivefold micro titer broth dilution assay in 96-well sterile plates at a final volume of 100 mL. Mid-log phase culture was diluted to a final concentration of 1×10^5 colony forming units/mL. Dried fractions were dissolved in 200 µL of ultrapure water and 20 µL applied in to each well and added to 80 µL of the bacterium/yeast dilution. A total of 100 µL of sterile water and PB or PDB were used as quality controls. Tetracycline and/or amphotericin B were also used as controls for growth inhibition. The microtiter plates were incubated for 18 h at 30°C; growth inhibition was determined by measuring absorbance at 595 nm.

Minimal inhibitory concentrations (MICs) were determined using the purified peptide against Gram-negative, Gram-positive, fungal, and yeast strains. MIC determination was performed using a fivefold microtiter broth dilution assay of stock solution, and serial dilution in 96-well sterile plates at a final volume of 100 µL where 20 µL of stock solution was applied in to each well

at serial dilution twofold microtiter broth dilution and added to 80 µL of the bacterium/yeast dilution. MIC is defined as the minimal concentration of peptide that caused 100% growth inhibition (33–35, 38, 39). In an attempt to test how rich broth medium affects oligoventin activity, growth inhibition assays were carried out using RPMI-1640 (Sigma-Aldrich) with MOPS 0.165 mol/L (RPMI without bicarbonate 10.4 g/L; MOPS [3-(*n*-morpholino) propanesulfonic acid] 34.53 g/L) at pHs 7.0 and 5.0 against *C. albicans* MDM8.

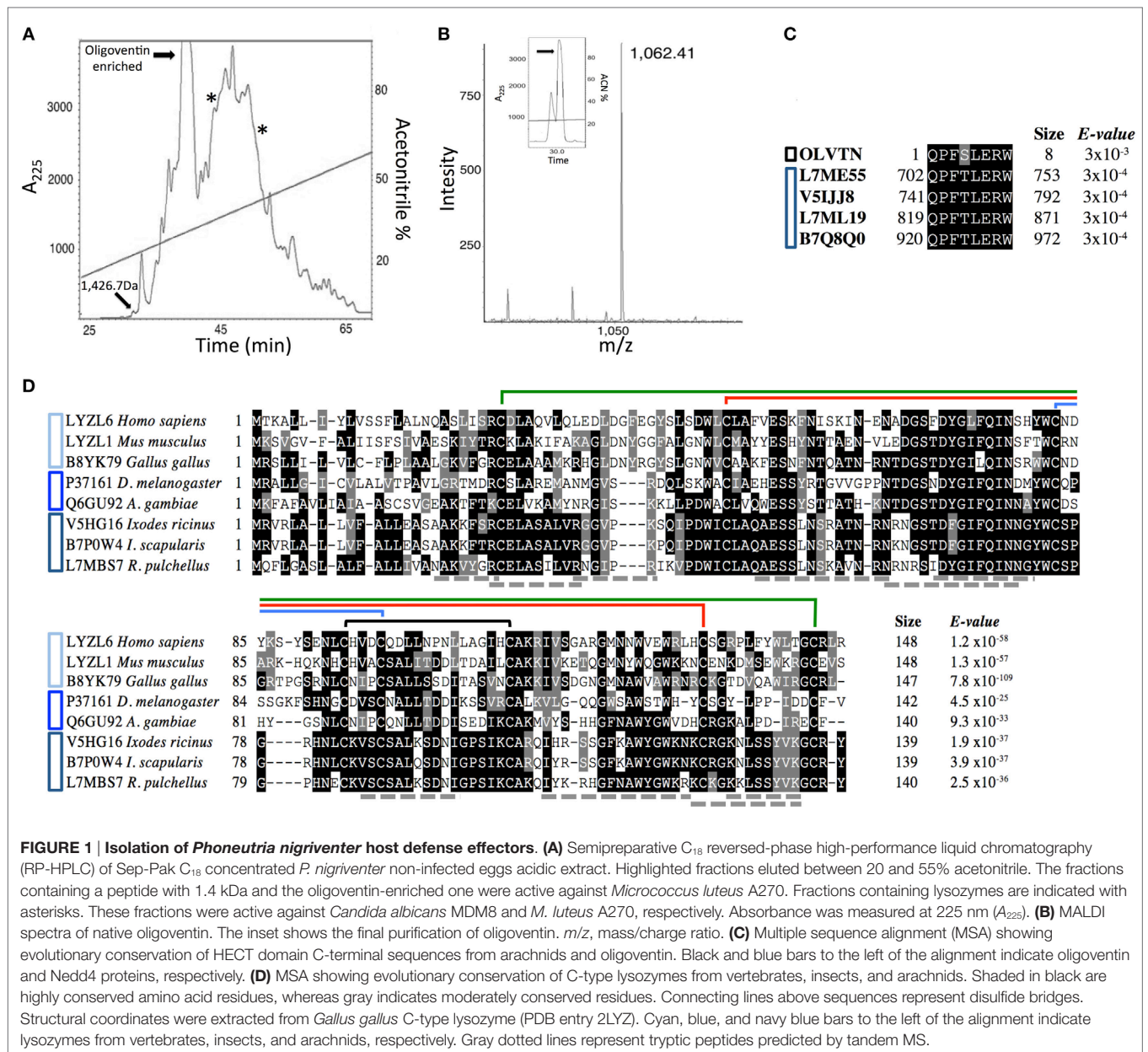
Synergy was measured by checkerboard titration assays using a fivefold microtiter broth dilution assay of stock solution and serial dilution in 96-well sterile plates at a final volume of 100 µL. Oligoventin was diluted along the rows of a microtiter tray and the avian lysozyme was diluted along the columns. A total of 80 µL of *M. luteus* was added to each well, diluted to a final concentration of 1×10^5 colony forming units/mL. The fractional inhibitory concentration (FIC) was determined after 18 h of incubation of the plates at 30°C. Synergy was defined as an FIC index of 0.5 or less as it represents at least fourfold decrease in the MIC of each compound (41), calculated according to the following formula: FIC index = $[A]/MIC_A + [B]/MIC_B$, where [A] was the concentration of drug A in a well that represented the lowest inhibitory concentration in its row, MIC_A was the MIC of drug A alone, [B] was the concentration of drug B in a well that represented the lowest inhibitory concentration in its row, and MIC_B was the MIC of drug B alone. Growth inhibition was determined by measuring absorbance at 595 nm. Bioassays were done in triplicate.

Oligoventin Toxicity to Erythrocytes

The hemolytic activity of oligoventin was tested in duplicate using human erythrocytes. A 2.5% (v/v) suspension of erythrocytes washed in PBS was incubated with oligoventin ranging from 0.4 to 188.8 µM in a 96-well plate for 3 h with intermittent shaking. The absorbance in the supernatant was measured at 415 nm. Hemolysis caused by PBS and 1% (v/v) Triton X-100 were used as 0 and 100% controls, respectively.

Molecular Mass Characterization and Sequence Determination

The fractions enriched for peptides were spotted (0.5 µL) onto the sample slide, dried on the bench, and crystallized with 0.5 µL of matrix solution [5 mg/mL (w/v) CHCA (α-cyano-4-hydroxycinnamic acid), in 50% acetonitrile and 0.1% TFA] (Sigma). The samples were analyzed on an Ettan MALDI-ToF/Pro spectrometer (Amersham Biosciences) operating in reflectron and positive ion mode. To determine the amino acid sequence of peptides, Edman degradation was performed in a PPSq 21 Automated Protein Sequencer (Shimadzu Co., Japan). Lysozymes were analyzed by SDS polyacrylamide gel electrophoresis (12.5% SDS-PAGE). In-gel lysozymes were destained, dehydrated in 100% acetonitrile for 10 min, and lyophilized. Freeze-dried purified protein was dissolved (1 mg/mL) in denaturant buffer [6 M GdmCl (guanidinium chloride), 0.25 M Tris/HCl, and 1 mM EDTA, pH 8.5]. A total of 20 µL of 2-mercaptoethanol (Sigma) was added to the mixture, followed by vortex-mixing and incubating



at 37°C for 2 h. After incubation, 100 µL of 4-vinylpyridine was added to the solution, followed by incubation at room temperature (26°C) for 2 h. The reduction and alkylation of the protein were confirmed by mass spectrometry. Reduced and alkylated proteins were digested with trypsin (Boehringer Mannheim) and tryptic peptides were analyzed by tandem mass spectrometry (MS/MS) in a Q-TOF Ultima API (Micromass) spectrometer operating in positive ion mode. In the mass spectrometer, doubly charged ions of sufficient abundance were selected for MS/MS fragmentation. MS/MS peak list files were submitted to an in-house version of MASCOT server (Matrix Science, USA) and screened against the Uniprot database. Representative resulting spectra and corresponding tables are provided in Table S1 in Supplementary Material.

Arachnid Genome Screening

BLAST searches were done against the *Ixodes scapularis* IscaW1 genome (42), a well-annotated arachnid genome resource, using oligoventin sequence as the query. Peptide-protein matching was adjusted with the following stringent settings: word size: 2; filter off; e-value 20,000; composition-based statistic off; PAM30 scoring matrix.

HECT Domain Homology Modeling and Surface Mapping of Conserved Residues

The 3D model of the HECT domain was generated using the online server HHPred (43) for template identification and Modeller 9v15 (44) for the model construction. PDB entry 2ONI

was used as template (71% similarity). The quality of the final structure was accessed using MolProbity (45) showing just one residue out of the Ramachandran allowed region (Ala 61) and 99.02% of the residues placed on the favorable region. Rate of amino acid evolution among Nedd4 HECT E3s was calculated from 50 homologs using maximum-likelihood phylogenetic analysis and mapped onto the protein structure using default parameters with Consurf (46).

Computational Simulations of Proteasomal Degradation

Human proteasome cleavage predictions were simulated with a stringent 0.7 threshold for four sequences (Table S3 in Supplementary Material) using the neural network algorithm Netchop3.1 online server “<http://www.cbs.dtu.dk/services/NetChop/>.” We used the C-term 3.0 network, which is trained on 1,260 naturally occurring MHC class I ligands (8, 47).

Phylogenetic Reconstruction of Eukaryotic Nedd4 Diversification

Bayesian phylogenetic inference was carried out with modifications from Ref. (48). Orthologous sequences containing the HECT domain (PF00632) were collected from a large number of eukaryotes. A complete list of Uniprot identifiers was used for the acquisition of 4,392 HECT-containing protein sequences. Redundancy in our dataset was minimized by employing a clustering methodology using the CD-HIT software (49) with a threshold of >90%. One representative protein from each cluster was retrieved for further analysis. This yielded 318 protein sequences relative to 370 eukaryotic species (Table S2 in Supplementary Material), which were aligned with T-coffee (50). The final alignment was manually edited in GeneDoc (51) resulting in 2,470 sites. The choice of the best-fit model of evolution was performed with ProtTest3 (52) using the Akaike Information Criterion, which led us to choose the WAG model. The phylogenetic reconstruction was inferred by the Bayesian inference method implemented in the Beast v1.7.0 software (53, 54). The starting tree was randomly generated, and the proportion of invariable sites and g-distributed rate variation across sites were estimated. The substitution rate categories were set in four categories, and we modeled the molecular clock accordingly to relaxed clock model available (55). The clades were supported by posterior probabilities obtained by Bayesian analysis. For Bayesian method generations, the burn-in was determined in Tracer (54) through log-likelihood scores, and data were summarized in TreeAnnotator (54) after trees that were out of the convergence area had been discarded. A total of 10,000,000 trees were generated, from which 25,000 were burned out of the final tree. The visualization and the final tree edition were performed using FigTree v1.3.1 “<http://tree.bio.ed.ac.uk/software/figtree/>.” Finally, proteins belonging to each phylogenetic cluster were dissected for revealing the oligoventin-orthologous sequence, from which sequence logos belonging to each clade were generated in Weblogo (56).

Epitope Comparison

The Immune Epitope Database (57) was screened using Nedd4 E3 HECT proteins as queries and yielded nine MHC ligands.

RESULTS

Identification of Host Defense Effectors

We combined solid-phase purification with assay-guided RP-HPLC runs to isolate host defense effectors from an acidic extract from *P. nigriverter* non-infected eggs. Four fractions with antimicrobial activity were found (Figure 1A). Mass analysis indicates that they were enriched for innate immune defense effectors, namely, HDPs and lysozymes (Figure S1 in Supplementary Material). To isolate these host defense effectors, analytical RP-HPLC runs coupled with bioassays yielded two fractions with molecules ranging in size from 0.8 to 1.7 kDa (Figures S1A–C in Supplementary Material), which were active against the Gram-positive bacteria *M. luteus* A270, and two lysozyme-like molecules ranging in size from 14 to 16 kDa (Figures S1D–F in Supplementary Material). A peptide with a mass of 1.4 kDa (Figure S1A in Supplementary Material) was purified to homogeneity, and Edman degradation showed that this peptide was N-terminally blocked. Further investigation by tandem mass spectrometry is needed to sequence this putative HDP. The fraction enriched for oligoventin (Figure 1A) was further purified. N-terminal sequencing by Edman degradation revealed an oligopeptide with eight residues: QPFSLERW, which we named oligoventin. Matrix-assisted laser desorption/ionization—time of flight mass spectrometry (MALDI-ToF-MS) analysis of the resulting fraction shows that the oligoventin molecular weight (MW) observed 1,061.4 Da ($M + H^+ = 1,062.4$, Figure 1B) corresponds to the MW calculated (1,061.5 Da).

Fractions enriched for lysozymes (Figures S1D–F in Supplementary Material) were active against the Gram-positive bacteria *M. luteus* A270 and the yeast *C. albicans* MDM8, respectively, and were further fractionated by C_{18} RP-HPLC. Reduction, alkylation, and trypsinization of these fractions followed by comparison of the resulting tryptic peptides by LC-ESI-MS/MS suggest that both of these antimicrobial factors are C-type lysozymes (Table S1 in Supplementary Material). Oligoventin and the putative HDP with 1.4 kDa show similar relative abundances. Altogether, these four antimicrobials represent less than 4% of the total protein content (Figure S2 in Supplementary Material).

Oligoventin Antimicrobial Activity and Synergy with Lysozymes

Oligoventin presents antimicrobial activity and MICs in the micromolar range (Table 1), markedly against Gram-positive bacteria: *M. luteus* A270 and the multi-resistant *S. aureus* ATCC 29213 and *S. epidermidis* ATCC 12228 strains. It is also active against the yeast *C. albicans* MDM8 and the Gram-negative bacteria *S. marcescens* ATCC 4112. However, it has antimicrobial activity in concentrations higher than those of rondonin (35), a cryptic peptide derived from the oxygen-carrier protein hemocyanin (19, 35). Oligoventin is active at concentrations

TABLE 1 | Minimal inhibitory concentrations (MICs) of HDPs from arachnids.

Microorganisms	MIC (μM)	
	Rondonin (35)	Oligoventin
Gram-positive bacteria		
<i>Micrococcus luteus</i> A270	ND	47.2–94.5
<i>Staphylococcus aureus</i> ATCC 29213	ND	94.5–188.9
<i>Staphylococcus epidermidis</i> ATCC 12228	ND	94.5–188.9
Gram-negative bacteria		
<i>Escherichia coli</i> SBS363	NT	ND
<i>Escherichia coli</i> ATCC 25922	NT	ND
<i>Pseudomonas aeruginosa</i> ATCC 27853	NT	ND
<i>Serratia marcescens</i> ATCC 4112	NT	23.53–47.2
Fungi		
<i>Trichosporon</i> sp. IOC 4569	1.1–2.1	ND
Yeasts		
<i>Candida albicans</i> MDM8	16.75–33.5	94.5–188.9
<i>Candida krusei</i> IOC 4559	16.75–33.5	ND
<i>Candida glabrata</i> IOC 4565	8.37–16.5	ND
<i>Candida albicans</i> IOC 4558	8.37–16.5	ND
<i>Candida parapsilosis</i> IOC 4564	16.75–33.5	ND
<i>Candida tropicalis</i> IOC 4560	8.37–16.5	ND
<i>Candida guilliermondii</i> IOC 4557	16.75–33.5	ND

The highest concentration tested was 189 μM . ND, activity was not detected in the range of concentrations tested. NT, antimicrobial activity was not tested or previously not reported (35). Rondonin is an hemocyanin-encrypted antifungal HDP from the tarantula *Acanthoscurria rondoniae* (19, 35).

MICs are expressed as the interval of two concentrations, where the first is the highest concentration tested at which microorganisms from each strain grew and the second is the lowest concentration tested that caused 100% growth inhibition (33–35, 38, 39).

ranging from 47 to 188.9 μM , while rondonin is active from 16.5 to 33.5 μM , except for the *Trichosporon* sp. IOC 4569 fungi strain, which is 2.1 μM (35). In contrast, oligoventin inhibits growth of Gram-positive, Gram-negative, and yeast strains, thus having a broader spectrum of antimicrobial activity compared to that of rondonin.

C-type lysozymes, classical players of innate immunity, are ubiquitous and constitutively expressed in leukocytes, providing an immediate defensive barrier to invading pathogens (20, 41, 58). Because context-specific co-expression of lysozymes with other host defense elements such as HDPs suggests synergy between these factors (41, 59), we investigated whether oligoventin has a mutually potentiating effect on the *G. gallus* C-lysozyme, as C-type lysozymes from arachnids are highly similar to avian lysozymes (Figure 1D). We determined the MIC of avian C-lysozyme as 0.01–0.02 μM against *M. luteus* A270. When both factors were tested together against this strain, we found a 0.37 FIC index, indicating strong synergy (41) between these molecules (Table S2 in Supplementary Material). MIC of oligoventin is potentiated 8-fold by lysozymes, exerting its effect at a concentration as low as 11.81 μM , whereas oligoventin potentiates lysozymes by decreasing its MIC 3.5-fold, causing growth inhibition at 0.006 μM . Moreover, oligoventin exhibits 4.3% hemolytic activity against human erythrocytes at 188.9 μM and 2% at 94.5 μM (Figure S3 in Supplementary Material). However, both oligoventin and lysozymes did not display antimicrobial activity against *C. albicans* MDM8 when cultured in RPMI-1640 (Sigma-Aldrich) at pH 5.0 or 7.0, even when oligoventin concentration was increased

twofold relative to its MIC (376 μM). Similarly, although we determined lysozymes as having a MIC of 1.3–2.6 μM against *C. albicans* MDM8 in poor dextrose broth medium, this classical HDP also lacked antimicrobial activity in RPMI-1640, even when lysozyme concentration was increased more than 50-fold relative to its MIC (data not shown).

E3 Ubiquitin Ligases As Oligoventin Precursor Proteins

To determine the oligoventin precursor protein, we BLAST screened the blacklegged tick *I. scapularis* genome IscaW1.4 from Vectorbase (42), a high-quality genomic resource for arachnids. Results showed high homology (88%) between oligoventin and the C-terminal HECT (Homologous to the E6-AP Carboxyl Terminus) domain sequence between residues 920–927 of Nedd4 (neural precursor cell expressed, developmentally downregulated 4-like) E3 ubiquitin ligases (Figure 1C; Figures S4 and S5 in Supplementary Material). Consistently, BLAST screening the Uniprot database with adjustments for peptide–protein matching revealed similar results for orthologous Nedd4 from two arachnids (*Ixodes ricinus*, Uniprot ID: V5IJJ8, and *Rhipicephalus pulchellus*, Uniprot IDs: L7ML19, L7ME55) and one crustacean (*Daphnia pulex*, Uniprot ID: E9GKW9). These results suggest oligoventin is a cryptic peptide released by the proteolysis of a *P. nigriventer* Nedd4 protein ortholog. Alternative hypotheses of oligoventin precursor proteins are indicated in Table S3 in Supplementary Material.

3D molecular modeling of the blacklegged tick Nedd4 E3 (Uniprot ID: B7Q5Q0) HECT domain (Figure 2A) shows the oligoventin-homologous epitope folding as a β -sheet and its close position to the catalytic cysteine site and the PY motif. PY motifs are internal regulatory motifs that are recognized by WW domains. Notice that the internal PY motif from HECT domains is different from the PPxY motifs found in substrates targeted also by the WW domains of catalytic HECT E3 ligases [reviewed in Ref. (60)] and herein we refer only to the regulatory PY motif from HECT domains.

Surface mapping projection of orthologous sequences onto the 3D model shows high-sequence conservation in the oligoventin-homologous, the catalytic site, and the regulatory PY motif C-terminal sequences (Figures 2B,C). Such high-sequence conservation suggests that the oligoventin-encrypted site is functionally important across different taxa.

In Silico Proteolytic Processing of E3 Ubiquitin Ligases by the 26S Proteasome

We used bioinformatics approaches to verify if oligoventin generation could occur by HECT E3s proteasome-mediated proteolysis. A small dataset (Figure S5 in Supplementary Material) consisting of only the C-terminal Nedd4 HECT domain of ubiquitin ligase sequences for two invertebrates (the arachnid *I. scapularis* and the crustacean *D. pulex*) and two vertebrates (mouse and human) were used as input in NetChop3.1, a neural network algorithm trained to predict 26S proteasomal cleavage sites both for constitutive and immunoproteasomes (8, 47). This approach is appropriate because tissue-specific proteasomes,

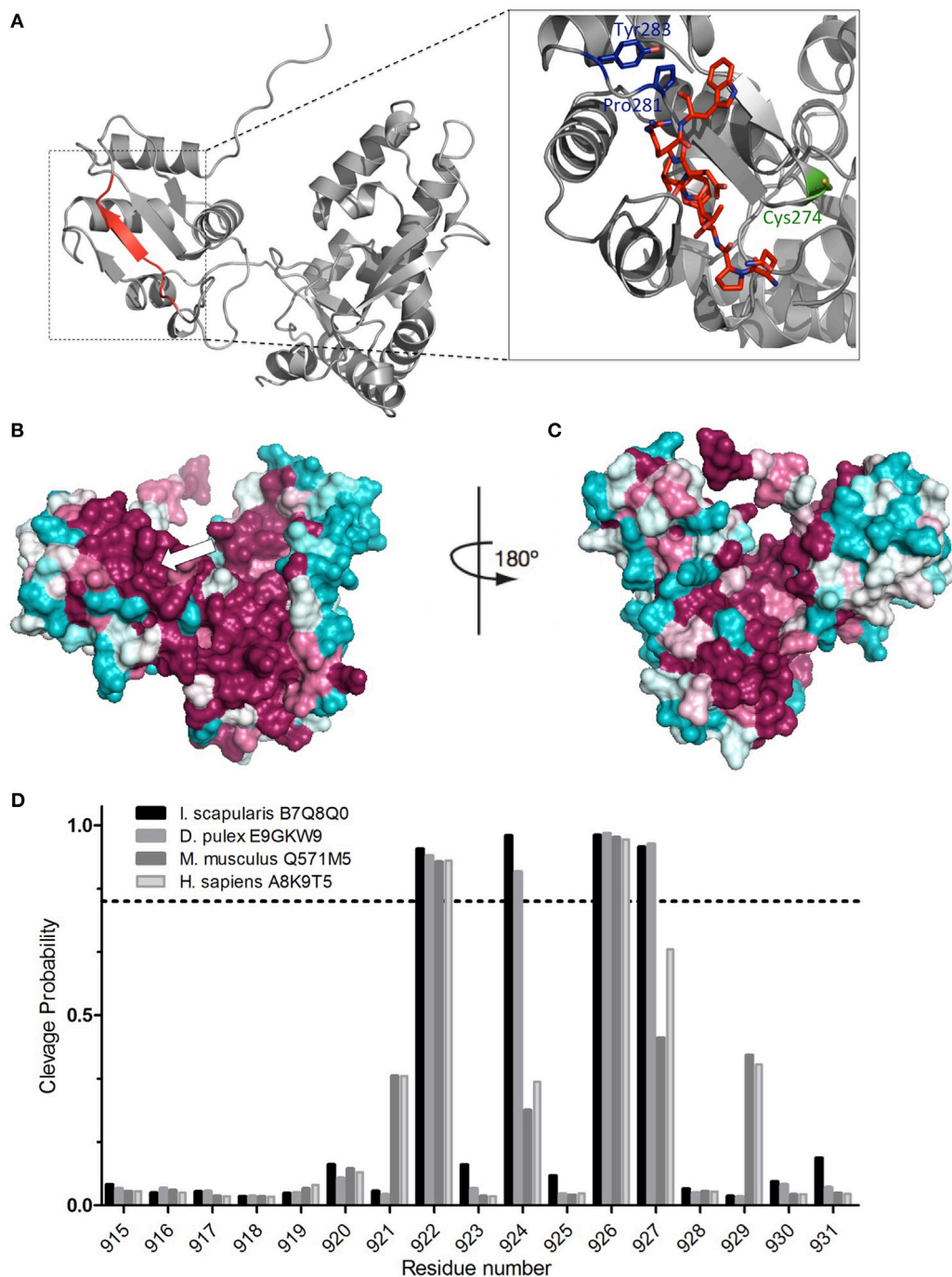


FIGURE 2 | Conservation of the oligoventin-encrypted site and *in silico* proteasomal degradation of E3s. (A) Homology modeling of the HECT domain from the arachnid B7Q8Q0 Nedd4 E3 ubiquitin ligase. Oligoventin is encrypted in a site between residues 920 and 927 (colored in red), which folds as a β -sheet. The catalytic site (Cys 274) and the PY motif (Pro281, Tyr 285) are colored in green and blue, respectively. **(B,C)** Surface mapping of conserved sites onto the arachnid Nedd4 HECT domain three-dimensional model. The arrow indicates the oligoventin-encrypted site. Teal represents the most variable residues and burgundy the most highly conserved. Model figures were generated using Pymol. **(D)** Proteasomal cleavage sites predicted for the HECT domain of four E3 sequences from human, mouse, the blacklegged tick *Ixodes scapularis*, and the crustacean *Daphnia pulex*. The dashed line shows the 0.7 threshold.

namely, constitutive, immune-, or thymus-specific proteasomes, are structurally rearranged in ways that combine different regulatory and catalytic domains, thus yielding different products

(3–6, 8, 9, 47, 61). Therefore, simulating several proteasomes in a single prediction method magnifies the possibility of accurately mapping multiple cleavage sites onto a template protein (8, 47).

Figure 2D shows the site positioned within the amino acid residues 920–927 in which oligoventin shows marked homology (**Figure 1C**) is enriched for cleavage sites in all sequences evaluated. In contrast, the flanking residues lack cleavage sites (Figure S6 and Table S4 in Supplementary Material). These results suggest high conservation of cleavage sites within metazoan E3s. Therefore, our approach indicates that Ned4 26S proteasome-mediated proteolysis might explain oligoventin production from an E3 ubiquitin ligase precursor protein.

Phylogenetic Analysis Reveals an Ancient HDP Signature in Metazoan Ned4s

In spite of the HECT domain deep conservation (**Figures 2B,C**), HECT E3s are pervasive within eukaryotic genomes, comprising more than 33 protein families highly diversified in animals, which have undergone wide architectural rearrangement (1–7). Therefore, our structural conservation data do not inform to what extent the different HECT-containing E3 families may contribute to oligoventin production. To address this issue, we conducted a comprehensive Bayesian phylogenetic analysis on 318 orthologous sequences recovered from all major eukaryotic clades (Table S5 in Supplementary Material), which yielded a tree topology (**Figure 3A**) consistent with previous findings (2). Our results indicate that Ned4 proteins are enriched for the oligoventin motif, but not WWP, Itchy, Smurf, and fungal Ned4 HECT ligases. **Figure 3B** illustrates sequence logos for each corresponding encrypted site within sampled orthologous sequences. We found that Ned4s are enriched for a motif composed of Q(P/M/L)F(S/T)(L/I)E(R/K/Q)W. In arthropods positioned near the base of extant ecdysozoans, namely, crustaceans and arachnids, the motif is more pronounced, with two conservative amino acid substitutions, one at the fourth position (S/T) and another at the seventh (K/R). The oligoventin sequence signature is also encrypted to some extent in Ned4 from insects and vertebrates, with a single non-conservative change at the seventh position in chordates.

Possible Convergent Evolution of the UPP in Metazoan Immune Systems

Because all nucleated cells from jawed vertebrates present their own antigens derived from cytosolic proteins to cytotoxic T cells through MHC class I (3, 4, 8, 9, 16, 47, 57), we investigated if human and mouse Ned4 proteins (7, 60) were involved in antigen presentation as a means of comparing the contribution of HECT E3s to the adaptive and innate immune systems. Screening the Immune Epitope Database (57) for Ned4-derived antigens yielded nine peptides, ranging in size from 8 to 16 residues (MWs between 0.9 and 1.2 kDa), indeed involved in antigen presentation as revealed by MHC ligand assays (**Table 2**). These results strongly support the hypothesis of convergent evolution of the UPP function in immunity. Therefore, it seems that the UPP was co-opted multiple times in immune systems. Whereas in mammals, Ned4 E3s play a role in the adaptive immune system as a core member of the UPP (2, 7, 9, 16, 47, 57, 60, 62) and as precursors for MHC class I antigens in mammals (e.g., **Table 2**), it is possible that a function for E3s has also evolved in the innate

defense as precursors of HDPs in a pathway likely dependent on the ubiquitin–proteasome system, at least in arachnids (**Figure 4**).

DISCUSSION

Our analysis suggests that a novel player in the ancient yet diverse innate immune system from arachnids (17, 19–21, 34, 35), oligoventin, shares homology to the catalytic HECT domain of metazoan Ned4 E3 ubiquitin ligases. Computational dissection of hundreds of HECT E3 ubiquitin ligases indicates that production of oligoventin-like HDPs might be limited to Ned4 orthologs, as the oligoventin signature is encrypted in metazoan Ned4s, but not in Ned4s from fungi nor in the closely related metazoan WWP/Itch, Smurf and HECW HECT-containing E3s, which might represent paralogs of this family of ligases. However, it is also possible that other classes of E3s can be involved in releasing additional HDPs. In fact, the results summarized in **Table 2** show that other HECT-containing E3s also produce functional peptides, such as HECW1 and HECW2; therefore, it is likely that other E3s can mediate the production of functional peptides, consistent with the low numbers of HECT-containing ligases in arachnid genomes (63). Indeed, while Ned4 or Ned4-2-deficient mice show a variety of phenotypes including embryonic and neonatal lethality (60), recent studies provided evidence that other classes of E3s (RING-containing) play a role in arachnid host defense, at least against the bacterial pathogen *Anaplasma phagocytophilum* (64–66), as suggested by silencing the E3 ligase XIAP in *I. scapularis* ticks (64). However, the underlying mechanism of the role of E3 in tick host defense remains elusive (64–66), and it would be interesting to test whether XIAPs are precursors of HDPs or are involved in the regulation of innate immune pathways.

The proposed homology between an antimicrobial peptide and the catalytic HECT domain of an E3 ubiquitin ligase immediately suggests a mechanism of oligoventin production by E3s self-proteolysis. E3 ubiquitin ligases are involved in the last step of ubiquitination, flagging substrates with ubiquitin for proteasomal degradation (1–8). Thus, E3s are the components that confer specificity to the UPP (3–6). E3s containing the HECT domain first form an intermediate thioester bond between the catalytic cysteine and ubiquitin before transferring this moiety to a lysine residue in the target substrate (4–6). In fact, the intermediate thioester formation with ubiquitin is critical to HECT E3s in *cis* self-ubiquitination activity (5, 6). Hence, a parsimonious mechanism of generating the HDP oligoventin would be HECT E3s self-ubiquitination coupled with proteasomal degradation. Indeed, our *in silico* simulation of the proteolysis of E3 ligases is consistent with the idea that oligoventin is generated as a product of the proteasomal degradation of E3s. It might be, therefore, straightforward to recruit the UPP to generate diverse HDPs from targeted cytosolic proteins, as it is for production of MHC class I antigens (3, 4, 8, 9, 16, 62).

Nevertheless, we cannot rule out the possibility that oligoventin can be generated by non-self-ubiquitination of oligoventin precursors coupled with proteasomal degradation [that is, E3s in *trans* ubiquitination (5, 6)], or ubiquitin- or proteasomal-independent proteolysis [e.g., by selective macroautophagy (26–29)]. In fact, Ned4 proteins preferentially conjugate the K63 linkage

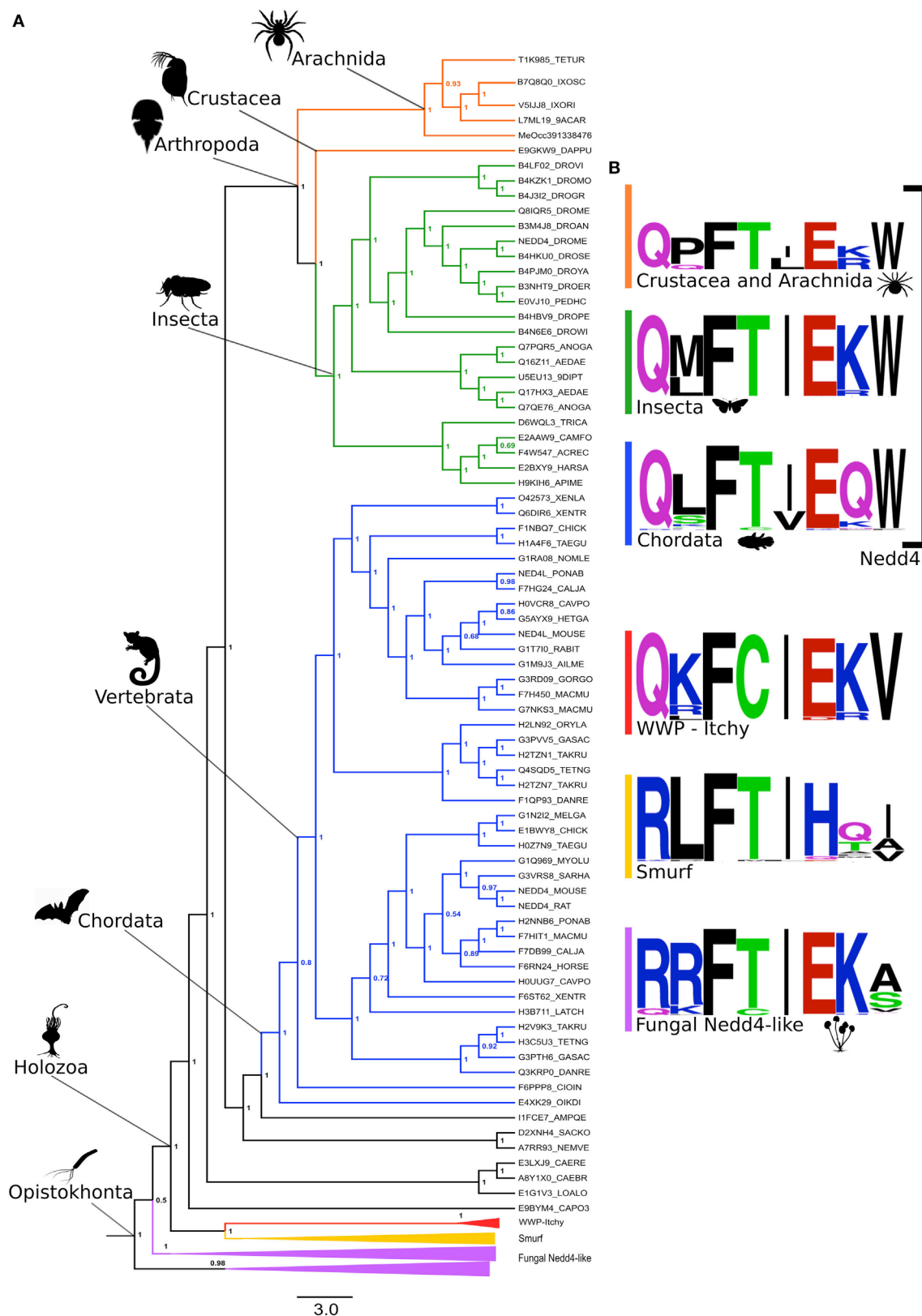


FIGURE 3 | Oligoventin signature conserved within metazoans suggests Nedd4 E3 ubiquitin ligases as the precursor protein candidate of oligoventin. (A) Bayesian phylogenetic inference of 318 HECT-containing E3s. **(B)** Oligoventin-orthologous sequence signatures for major protein subfamilies comprised of Nedd4, WWP-Itchy, Smurf, and fungal Nedd4 HECT E3 ligases groups. Silhouettes from organisms are from Phylopic (<http://phylopic.org/>).

TABLE 2 | Endogenous antigens derived from E3 ligases from the Nedd4 family involved in MHC-mediated antigens presentation.

Epitope ID	Peptide	MW	Antigen name	Species	Method	MHC allele name	MHC Class
191812	SGLCNEDHL	986.4	E3 ubiquitin-protein ligase 4-like isoform 1	<i>Mus musculus</i>	Cellular MHC/mass spectrometry	H2-Db	I
214505	LPFEKSQL	960.5	NEDD4-like E3 ubiquitin-protein ligase WWP2 isoform 3	<i>Homo sapiens</i>	Cellular MHC/mass spectrometry	HLA-B*08:01	I
225136	YFDEKELEL	1,184.5	NEDD4-like E3 ubiquitin-protein ligase WWP2	<i>H. sapiens</i>	Cellular MHC/mass spectrometry	HLA-C*04:01	I
241145	ARAPAPYKR	1,184.5	E3 ubiquitin-protein ligase HECW2	<i>H. sapiens</i>	Mass spectrometry	HLA-B*27:05	I
241758	GGSAATATAASEQSPG	1,028.5	NEDD4-like E3 ubiquitin-protein ligase WWP2	<i>M. musculus</i>	Mass spectrometry	HLA-DQ8	II
422506	KSRPIIKRF	1,446.6	E3 ubiquitin-protein ligase HECW1	<i>H. sapiens</i>	Cellular MHC/mass spectrometry	HLA-B*57:01	I
422718	SGLCNEDHL + OX(C ₄)	1,143.7	E3 ubiquitin-protein ligase NEDD4-like isoform 1	<i>M. musculus</i>	Cellular MHC/mass spectrometry	H2-class I	I
424855	GFLRLK MAY	986.4	E3 ubiquitin-protein ligase NEDD4-like isoform 6	<i>H. sapiens</i>	Cellular MHC/mass spectrometry	HLA-A*29:02	I
429266	ASDPYVKLSLY	1,097.5	E3 ubiquitin-protein ligase NEDD4-like	<i>H. sapiens</i>	Cellular MHC/mass spectrometry	HLA-A*01:01	I

MW, molecular weight.

Antigens were retrieved from the Immune Epitope Database (57).

ubiquitin chain on substrate proteins, which alters the signaling properties or trafficking pattern of these modified proteins, instead of the K48 linkage ubiquitin chain that usually directs ubiquitinated proteins to proteasomal degradation (64, 67). Therefore, it is more likely that oligoventin precursors are ubiquitinated by other E3s conjugating K48 linkages. Furthermore, it can be argued that alternative mechanisms can explain production of short HDPs such as oligoventin. For example, long non-coding RNAs that produce short-sized biologically active peptides (68, 69) or proteasomal-independent production of encrypted HDPs (26–28) might underlie the generation of oligoventin. However, the biochemical, phylogenetic, and computational evidence presented here supports a model in which E3s release oligoventin by proteasomal proteolysis.

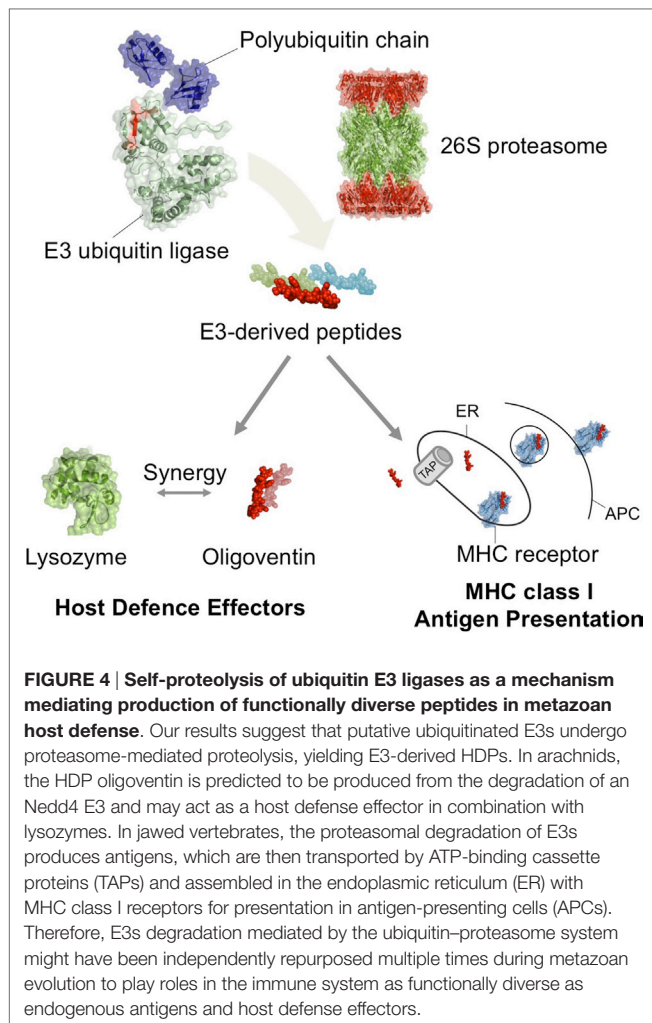
Oligoventin production by proteasomal degradation suggests that this peptide can be stored in intracellular compartments, similar to MHC class I antigens (8, 9, 47, 70) and then be directed to the extracellular space where it might play its functional role. Indeed, previous studies suggest that spider hemocytes (immune cells similar to mammalian macrophages) preferentially export antimicrobial effectors, such as acanthoscurrins and gomesins (33, 34), through exocytosis (71), contrasting with vertebrate macrophages, which preferentially display phagocytic activity against pathogens (71, 72). Therefore, the model in which oligoventin is derived from E3s proteasomal-mediated proteolysis could provide an extraordinary example of convergent evolution in which highly conserved orthologous pathways (e.g., TAP-dependent) direct proteasomal products to their extracellular site, despite functional divergence of those products. However, as some ubiquitin-encrypted HDPs are produced in autophagosomes (26–28), it is also possible that oligoventin can be produced by proteasomal-independent selective degradation of cytosolic Nedd4 proteins during macroautophagy of invading pathogens.

Together with the discovery of oligoventin, the presence of two lysozymes and a putative HDP of 1.4 kDa reveals hallmarks of an ancient immune defense system (17–21). Although oligoventin

alone shows a relatively weak antimicrobial activity (Table 1), it has potent synergy with lysozymes, thus providing evidence of putative E3-derived peptides playing an important role in modulating the innate defenses of *P. nigriventer*. Moreover, because some arachnids lack inducible production of HDPs (17, 19–21, 34, 71), it is possible that oligoventin is constitutively expressed during early development, as it was discovered in non-infected eggs. Alternatively, antimicrobial factors such as HDPs, lysozymes, and antibodies are either maternally deposited or upregulated by parental imprinting within several taxa, including, but not limited to, cnidarians (73), insects (74, 75), amphibians (76), and amniotes (77), such as birds, rats, and humans, then it is possible that oligoventin has a maternal origin and could play a role in regulating early microbial colonization during development.

Future research can benefit from MALDI imaging mass spectrometry (78, 79) to investigate E3-mediated production of HDPs *in vivo*. Combining MALDI imaging with the recently developed ubiquitin variants (80), which systematically modulate HECT E3 ligase activity, as well as proteasome inhibitors (81), will be useful to probe the mechanism of HDP production proposed herein. Furthermore, the i5k initiative (82) aims to sequence 63 arachnid genomes, including *P. nigriventer* itself and two closely related species from the Ctenidae family: *Phoneutria fera* and *Cupiennius salei* (83). Therefore, we expect that the community-based efforts to annotate their genomes will provide the sequence data needed to test if oligoventin indeed maps to ctenid HECT E3 ligases.

Oligoventin synergy with C-type lysozymes suggests this HDP is a lysozyme-partner effector. When combined with lysozymes, oligoventin inhibits clinically isolated Gram-positive bacteria growth *in vitro* at concentrations as low as 11.8 μ M. Thus, further research should test whether oligoventin is expressed in other tissues in addition to the eggs, which would indicate that it engages in constitutive innate immunity together with lysozymes at different developmental stages. Furthermore, the fact that synergy between antimicrobials might reduce the cost of defense (41, 59, 84) indicates that arachnids might be able to defend themselves



against a wide range of pathogens from a relatively limited repertoire of host defense effectors. Synergy between two distinct classes of antimicrobials is usually explained by different modes of action between the effector molecules (41, 59, 84). Hence, lysozyme-induced bacterial peptidoglycan disruption (41, 58) might facilitate oligoventin binding to its target, possibly explaining their synergy. Because oligoventin is a neutrally charged antimicrobial peptide, varying its net charge from -1 to 0.8 in pHs 10.0 and 4.0 , respectively, we suggest that it is unlikely that it binds directly to anionic membranes, although it could bind to membrane receptors. However, neutrally charged peptides usually act by binding to intracellular targets such as catalytic enzymes and nucleic acids, in contrast to cationic antimicrobial peptides directly disrupting membranes (40, 85, 86). Therefore, future studies aiming to understand oligoventin's mode of action might reveal the underlying mechanism of synergy of oligoventin with lysozymes.

Despite oligoventin's small size, antimicrobial activity against clinical strains, synergy with lysozymes, and lack of hemolytic activity, its discovery will likely be of limited interest to drug development as many short-sized peptides such as gomesin (34) and rondonin (35), among others (84–89), outcompete oligoventin's

attractiveness as a blueprint for next-generation antimicrobial drugs. However, the discovery of oligoventin highlights the potential to screen E3s for novel peptide-based drug discovery (90). At present, the identification of oligoventin provides two main insights: first, the discovery of E3-derived peptides as a possibly new class of biologically active peptides. Second, it sheds new light on comparative immunology by illustrating a remarkable case of independent evolution of UPP function in animal host defense. Therefore, the most relevant result of our study is the evidence suggesting that E3 degradation might have been independently repurposed leading to the production of MHC class I antigens at least in humans and mouse and possibly HDPs in arachnids, respectively. Our findings indicate that the UPP was independently coupled to different immune pathways during the evolution of metazoans as a possibly convergent adaptation of metazoan immunity to produce functionally diverse peptides.

In conclusion, our data support the prediction that Nedd4s play a role in a new innate immune-related cellular pathway dependent on the UPP. The evidence presented suggests an emergent function of HECT E3s as novel precursors of HDPs in the ancient arachnid innate immune system. If confirmed, it will highlight the functional plasticity of the UPP and expand the currently known function of E3s (3, 4, 6, 7, 60, 62, 91). Thus, our results are consistent with the hypothesis that the UPP has been independently co-opted several times during evolution and gained multiple immune-related functions. Further experimentation is therefore necessary to robustly test the postulated role of Nedd4 proteins in immunity suggested by the data presented here and also to further test the precise molecular origin of oligoventin.

DATA ACCESSIBILITY

The datasets supporting this article have been uploaded as part of the electronic supplementary material. The accession number for oligoventin is B3EWR9.

AUTHOR CONTRIBUTIONS

ILC-F, RSRS, and PIdSJ designed experiments; ILC-F, TK, RSRS, IdFCB, and PIdSJ carried out experiments; ILC-F and TK carried out bioinformatics analysis; RSRS and PIdSJ gave conceptual advice; ILC-F and PIdSJ wrote the manuscript with input from all the authors.

ACKNOWLEDGMENTS

The authors are thankful to current and former members of LETA-CAT/CEPID, CeTICS, and LEEV from the Butantan Institute for technical assistance, support, and advice. Thanks are also extended to F.Q. Camargo, L. Kühlen, and L. Field for critical reading of the manuscript, and to FEBRACE for the early and ongoing support to our projects.

FUNDING

This work received funding from São Paulo Research Foundation (FAPESP) Grants 13/07467-1 to CeTICS-CEPID

and 2014/03644-9 to TK, as well as from the Brazilian National Counsel of Technological and Scientific Development (CNPq) Grant 472744/2012-7. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Hochstrasser M. Origin and function of ubiquitin-like protein conjugation. *Nature* (2009) 7237:422. doi:10.1038/nature07958
- Grau-Bové X, Sebé-Pedrós A, Ruiz-Trillo I. A genomic survey of HECT ubiquitin ligases in eukaryotes reveals independent expansions of the HECT system in several lineages. *Genome Biol Evol* (2013) 5:833–47. doi:10.1093/gbe/evt052
- Bhoj VG, Chen ZJ. Ubiquitylation in innate and adaptive immunity. *Nature* (2009) 7237:430–7. doi:10.1038/nature07959
- Jiang X, Chen ZJ. The role of ubiquitylation in immune defence and pathogen evasion. *Nat Rev Immunol* (2011) 12:35–48. doi:10.1038/nri3111
- Weissman AM, Shabek N, Ciechanover A. The predator becomes the prey: regulating the ubiquitin system by ubiquitylation and degradation. *Nat Rev Mol Cell Biol* (2011) 23:605–20. doi:10.1038/nrm3173
- Rotin D, Kumar S. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* (2009) 10:398–409. doi:10.1038/nrm2690
- Ingham RJ, Gish G, Pawson T. The Nedd4 family of E3 ubiquitin ligases: functional diversity within a common modular architecture. *Oncogene* (2004) 23:1972–84. doi:10.1038/sj.onc.1207436
- Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* (2005) 57:33–41. doi:10.1007/s00251-005-0781-7
- Neeffes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* (2011) 11:823–36. doi:10.1038/nri3084
- Ashida H, Kim M, Sasakawa C. Exploitation of the host ubiquitin system by human bacterial pathogens. *Nat Rev Microbiol* (2014) 12(6):399–413. doi:10.1038/nrmicro3259
- Anderson DM, Frank DW. Five mechanisms of manipulation by bacterial effectors: a ubiquitous theme. *PLoS Pathog* (2012) 8(8):1–4. doi:10.1371/journal.ppat.1002823
- Hicks SW, Galán JE. Hijacking the host ubiquitin pathway: structural strategies of bacterial E3 ubiquitin ligases. *Curr Opin Microbiol* (2010) 13(1):41–6. doi:10.1016/j.mib.2009.11.008
- Huïbregtse J, Rohde JR. Hell's BELs: bacterial E3 ligases that exploit the eukaryotic ubiquitin machinery. *PLoS Pathog* (2014) 10(8):8–11. doi:10.1371/journal.ppat.1004255
- Lin DY, Diao J, Chen J. Crystal structures of two bacterial HECT-like E3 ligases in complex with a human E2 reveal atomic details of pathogen-host interactions. *Proc Natl Acad Sci U S A* (2012) 109(6):1925–30. doi:10.1073/pnas.1115025109
- Maculins T, Fiskin E, Bhogaraju S, Dikic I. Bacteria-host relationship: ubiquitin ligases as weapons of invasion. *Cell Res* (2016) 26(4):499–510. doi:10.1038/cr.2016.30
- Litman GW, Cannon JP, Dishaw LJ. Reconstructing immune phylogeny: new perspectives. *Nat Rev Immunol* (2005) 11:866–79. doi:10.1038/nri1712
- Bulet P, Stöcklin R, Menin L. Anti-microbial peptides: from invertebrates to vertebrates. *Immunol Rev* (2004) 198:169–84. doi:10.1111/j.0105-2896.2004.0124.x
- Ferrandon D, Imler JL, 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
- Kuhn-Nentwig L, Nentwig W. The immune system of spider. In: Nentwig W, editor. *Spider Ecophysiology*. Berlin, Heidelberg: Springer-Verlag (2013). p. 81–91. doi:10.1007/978-3-642-33989-9_7
- Bechsgaard J, Vanthournout B, Funch P, Vestbo S, Gibbs RA, Richards S, et al. Comparative genomic study of arachnid immune systems indicates loss of β GRPs and the IMD pathway. *J Evol Biol* (2015) 29:277–91. doi:10.1111/jeb.12780
- Palmer WJ, Jiggins FM. Comparative genomics reveals the origins and diversity of arthropod immune systems. *Mol Biol Evol* (2015) 8:2111–29. doi:10.1093/molbev/msv093
- Majetschak M. Extracellular ubiquitin: immune modulator and endogenous opponent of damage-associated molecular pattern molecules. *J Leukoc Biol* (2011) 89:205–19. doi:10.1189/jlb.0510316
- Kieffer AE, Goumon Y, Ruh O, Chasserot-Golaz S, Nullans G, Gasnier C, et al. The N- and C-terminal fragments of ubiquitin are important for the antimicrobial activities. *FASEB J* (2003) 17:776–8. doi:10.1096/fj.02-0699jfe
- Kim JY, Lee SY, Park SC, Shin SY, Choi SJ, Park Y, et al. Purification and antimicrobial activity studies of the N-terminal fragment of ubiquitin from human amniotic fluid. *Biochim Biophys Acta* (2007) 1774:1221–6. doi:10.1016/j.bbapap.2007.06.013
- Seo JK, Lee MJ, Go HJ, Kim GD, Jeong HD, Nam BH, et al. Purification and antimicrobial function of ubiquitin isolated from the gill of Pacific oyster, *Crassostrea gigas*. *Mol Immunol* (2013) 53:88–98. doi:10.1016/j.molimm.2012.07.003
- Ponpuak M, Davis AS, Roberts EA, Delgado MA, Dinkins C, Zhao Z, et al. Delivery of cytosolic components by autophagic adaptor protein p62 endows autophagosomes with unique antimicrobial properties. *Immunity* (2010) 32(3):329–41. doi:10.1016/j.immuni.2010.02.009
- Münz C. Selective macroautophagy for immunity. *Immunity* (2010) 32:298–9. doi:10.1016/j.immuni.2010.03.002
- Ponpuak M, Deretic V. Autophagy and p62/sequestosome 1 generate neo-antimicrobial peptides (cryptides) from cytosolic proteins. *Autophagy* (2011) 7(3):336–7. doi:10.4161/auto.7.3.14500
- Kraft C, Peter M, Hofmann K. Selective autophagy: ubiquitin-mediated recognition and beyond. *Nat Cell Biol* (2010) 12(9):836–41. doi:10.1038/ncb0910-836
- Pasikowski P, Goździewicz T, Stefanowicz P, Artym J, Zimecki M, Szewczuk Z. A novel immunosuppressive peptide originating from the ubiquitin sequence. *Peptides* (2011) 32:2418–27. doi:10.1016/j.peptides.2011.10.002
- Kluczyk A, Cydzik M, Biernat M, Bachor R, Pasikowski P, Stefanowicz P, et al. Dimeric analogs of immunosuppressive decapeptide fragment of ubiquitin. *J Pept Sci* (2012) 18:456–65. doi:10.1002/psc.2416
- Baumann T, Kämpfer U, Schürch S, Schaller J, Largiadèr C, Nentwig W, et al. Ctenidins: antimicrobial glycine-rich peptides from the hemocytes of the spider *Cupiennius salei*. *Cell Mol Life Sci* (2010) 16:2787–98. doi:10.1007/s00018-010-0364-0
- Lorenzini DM, da Silva PI Jr, Fogaça AC, Bulet P, Daffre S. Acanthoscurrin: a novel glycine-rich antimicrobial peptide constitutively expressed in the hemocytes of the spider *Acanthoscurria gomesiana*. *Dev Comp Immunol* (2003) 9:781–91. doi:10.1016/S0145-305X(03)00058-2
- Silva PI Jr, Daffre S, Bulet P. Isolation and characterization of gomesin, an 18-residue cysteine-rich defense peptide from the spider *Acanthoscurria gomesiana* hemocytes with sequence similarities to horseshoe crab antimicrobial peptides of the tachyplesin family. *J Biol Chem* (2000) 43:33464–70. doi:10.1074/jbc.M001491200
- Riciluca KC, Sayegh RS, Melo RL, Silva PI Jr. Rondonin an antifungal peptide from spider (*Acanthoscurria rondoniae*) haemolymph. *Results Immunol* (2012) 2:66–71. doi:10.1016/j.rinim.2012.03.001
- Giribet G, Edgecombe GD. Reevaluating the arthropod tree of life. *Annu Rev Entomol* (2012) 57:167–86. doi:10.1146/annurev-ento-120710-100659
- Conlon JM. Purification of naturally occurring peptides by reversed-phase HPLC. *Nat Protoc* (2007) 2(1):191–7. doi:10.1038/nprot.2006.437

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fimmu.2016.00664/full#supplementary-material>.

38. Bulet P. Strategies for the discovery, isolation, and characterization of natural bioactive peptides from the immune system of invertebrates. *Methods Mol Biol* (2008) 494:9–29. doi:10.1007/978-1-59745-419-3_2
39. Wiegand I, Hilpert K, Hancock RE. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* (2008) 3(2):163–75. doi:10.1038/nprot.2007.521
40. Wang G. Antimicrobial peptides: discovery, design and novel therapeutic strategies. *Adv Mol Cell Microbiol* (2010) 18:116–40. doi:10.1079/9781845936570.0000
41. Yan H, Hancock RE. Synergistic interactions between mammalian antimicrobial defense peptides. *Antimicrob Agents Chemother* (2001) 45:1558–60. doi:10.1128/AAC.45.5.1558-1560.2001
42. Gulia-Nuss M, Nuss AB, Meyer JM, Sonenshine DE, Roe RM, Waterhouse RM, et al. Genomic insights into the *Ixodes scapularis* tick vector of Lyme disease. *Nat Commun* (2016) 7:10507. doi:10.1038/ncomms10507
43. Söding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* (2005) 33:W244–8. doi:10.1093/nar/gki408
44. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Min-yi S, et al. Comparative protein structure modeling using modeller. *Curr Protoc Bioinformatics* (2006) 5:Unit5.6. doi:10.1002/0471140864.ps0209s50
45. Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* (2010) 66:12–21. doi:10.1107/S0907444909042073
46. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* (2010) 38:W529–33. doi:10.1093/nar/gkq399
47. Keşmir C, Nussbaum AK, Schild H, Detours V, Brunak S. Prediction of proteasome cleavage motifs by neural networks. *Protein Eng* (2002) 15:287–96. doi:10.1093/protein/15.4.287
48. Geraldo MT, Valente GT, Braz AS, Martins C. The discovery of Foxl2 paralogs in chondrichthyan, coelacanth and tetrapod genomes reveals an ancient duplication in vertebrates. *Heredity* (2013) 111:57–65. doi:10.1038/hdy.2013.19
49. 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
50. Notredame C, Higgins DG, Heringa J. T-coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* (2000) 302:205–17. doi:10.1006/jmbi.2000.4042
51. Nicholas KB, Nicholas HB Jr, Deerfield DW. GeneDoc: analysis and visualization of genetic variation. *EMBNEW NEWS* (1997) 4:14.
52. 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
53. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* (2007) 7:214. doi:10.1186/1471-2148-7-214
54. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* (2012) 29:1969–73. doi:10.1093/molbev/mss075
55. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* (2006) 5:e88. doi:10.1371/journal.pbio.0040088
56. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* (2004) 14:1188–90. doi:10.1101/gr.849004
57. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* (2014) 43:D405–12. doi:10.1093/nar/gku938
58. Chipman DM, Sharon N. Mechanism of lysozyme action. *Science* (1969) 165:454–65. doi:10.1126/science.165.3892.454
59. McCafferty DG, Cudic P, Yu MK, Behenna DC, Kruger R. Synergy and duality in peptide antibiotic mechanisms. *Curr Opin Chem Biol* (1999) 3:672–80. doi:10.1016/S1367-5931(99)00025-3
60. Scheffner M, Kumar S. Mammalian HECT ubiquitin-protein ligases: biological and pathophysiological aspects. *Biochim Biophys Acta* (2014) 1843(1):61–74. doi:10.1016/j.bbamcr.2013.03.024
61. Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, et al. *Science* (2007) 316(5829):1349–53. doi:10.1126/science.1141915
62. Strehl B, Seifert U, Kruger E, Heink S, Kuckelkorn U, Kloetzel PM. Interferon-gamma, the functional plasticity of the ubiquitin-proteasome system, and MHC class I antigen processing. *Immunol Rev* (2005) 207(1):19–30. doi:10.1111/j.0105-2896.2005.00308.x
63. Choy A, Severo MS, Sun R, Girke T, Gillespie JJ, Pedra JH. Decoding the ubiquitin-mediated pathway of arthropod disease vectors. *PLoS One* (2013) 8(10):e78077. doi:10.1371/journal.pone.0078077
64. Severo MS, Choy A, Stephens KD, Sakhon OS, Chen G, Chung DW, et al. The E3 ubiquitin ligase XIAP restricts *Anaplasma phagocytophilum* colonization of *Ixodes scapularis* ticks. *J Infect Dis* (2013) 208(11):1830–40. doi:10.1093/infdis/jit380
65. Hajdušek O, Sîma R, Ayllón N, Jalovecká M, Perner J, de la Fuente J, et al. Interaction of the tick immune system with transmitted pathogens. *Front Cell Infect Microbiol* (2013) 3:26. doi:10.3389/fcimb.2013.00026
66. Severo MS, Sakhon OS, Choy A, Stephens KD, Pedra JH. The “ubiquitous” reality of vector immunology. *Cell Microbiol* (2013) 15(7):1070–8. doi:10.1111/cmi.12128
67. Nathan JA, Kim HT, Ting L, Gygi SP, Goldberg AL. Why do cellular proteins linked to K63-polyubiquitin chains not associate with proteasomes? *EMBO J* (2013) 32(4):552–65. doi:10.1038/emboj.2012.354
68. Ruiz-Orera J, Messegue X, Subirana JA, Alba MM. Long non-coding RNAs as a source of new peptides. *Elife* (2014) 3:e03523. doi:10.7554/eLife.03523
69. Andrews SJ, Rothnagel JA. Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet* (2014) 15:193–204. doi:10.1038/nrg3520
70. van Montfort N, Camps MG, Khan S, Filippov DV, Weterings JJ, Griffith JM, et al. Antigen storage compartments in mature dendritic cells facilitate prolonged cytotoxic T lymphocyte cross-priming capacity. *Proc Natl Acad Sci U S A* (2009) 106(16):6730–5. doi:10.1073/pnas.0900969106
71. Fukuzawa AH, Vellutini BC, Lorenzini DM, Silva PI, Mortara RA, da Silva JM, et al. The role of hemocytes in the immunity of the spider *Acanthoscurria gomesiana*. *Dev Comp Immunol* (2008) 32(6):716–25. doi:10.1016/j.dci.2007.11.002
72. Flannagan RS, Cosío G, Grinstein S. Antimicrobial mechanisms of phagocytes and bacterial evasion strategies. *Nat Rev Microbiol* (2009) 7(5):355–66. doi:10.1038/nrmicro2128
73. Fraune S, Augustin R, Anton-Erxleben F, Wittlieb J, Gelhaus C, Klimovich VB, et al. In an early branching metazoan, bacterial colonization of the embryo is controlled by maternal antimicrobial peptides. *Proc Natl Acad Sci U S A* (2010) 107:18067–72. doi:10.1073/pnas.1008573107
74. Jacobs CG, van der Zee M. Immune competence in insect eggs depends on the extraembryonic serosa. *Dev Comp Immunol* (2013) 41:263–9. doi:10.1016/j.dci.2013.05.017
75. Sadd BM, Schmid-Hempel P. Facultative but persistent trans-generational immunity via the mother's eggs in bumblebees. *Curr Biol* (2007) 17:R1046–7. doi:10.1016/j.cub.2007.11.007
76. Walke JB, Harris RN, Reinert LK, Rollins-Smith LA, Woodhams DC. Social immunity in amphibians: evidence for vertical transmission of innate defenses. *Biotropica* (2011) 43:396–400. doi:10.1111/j.1744-7429.2011.00787.x
77. Grindstaff JL, Brodie ED, Ketterson ED. Immune function across generations: integrating mechanism and evolutionary process in maternal antibody transmission. *Proc Biol Sci* (2003) 270:2309–19. doi:10.1098/rspb.2003.2485
78. Kuhn-Nentwig L, Kopp LS, Nentwig W, Haenni B, Streitberger K, Schürch S, et al. Functional differentiation of spider hemocytes by light and transmission electron microscopy, and MALDI-MS-imaging. *Dev Comp Immunol* (2014) 1:59–67. doi:10.1016/j.dci.2013.10.008
79. Cornett DS, Reyzer ML, Chaurand P, Caprioli RM. MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. *Nat Methods* (2007) 10:828–33. doi:10.1038/nmeth1094
80. Zhang W, Wu KP, Sartori MA, Kamadurai HB, Ordureau A, Jiang C, et al. System-wide modulation of HECT E3 ligases with selective ubiquitin variant probes. *Mol Cell* (2016) 62:1–16. doi:10.1016/j.molcel.2016.02.005
81. Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, et al. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. *Cell* (1994) 78:761–71. doi:10.1016/S0092-8674(94)90462-6

82. Consortium i5K. The i5K Initiative: advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *J Hered* (2013) 104:595–600. doi:10.1093/jhered/est050.104:595–600
83. Platnick NI. The world spider catalog, version 15. *Am Museum Nat Hist* (2014). Available from: <http://research.amnh.org/entomology/spiders/catalog/index.html>
84. Rahnamaeian M, Cytryńska M, Zdybicka-Barabas A, Dobslaff K, Wiesner J, Twyman RM, et al. Insect antimicrobial peptides show potentiating functional interactions against Gram-negative bacteria. *Proc Biol Sci* (2015) 282:20150293. doi:10.1098/rspb.2015.0293
85. Brogden KA. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nat Rev Microbiol* (2005) 3(3):238–50. doi:10.1038/nrmicro1098
86. Guilhelmelli F, Vilela N, Albuquerque P, Derengowski, Lda S, Silva-Pereira I, Kyaw CM. Antibiotic development challenges: the various mechanisms of action of antimicrobial peptides and of bacterial resistance. *Front Microbiol* (2013) 4:353. doi:10.3389/fmicb.2013.00353
87. Cho JH, Sung BH, Kim SC. Buforins: histone H2A-derived antimicrobial peptides from toad stomach. *Biochim Biophys Acta* (2009) 1788(8):1564–9. doi:10.1016/j.bbame.2008.10.025
88. Chaparro E, da Silva PI. Lacrain: the first antimicrobial peptide from the body extract of the Brazilian centipede *Scolopendra viridicornis*. *Int J Antimicrob Agents* (2016) 48(3):277–85. doi:10.1016/j.ijantimicag.2016.05.015
89. Ayroza G, Ferreira IL, Sayegh RS, Tashima AK, da Silva Junior PI. Juruin: an antifungal peptide from the venom of the Amazonian Pink Toe spider, *Avicularia juruensis*, which contains the inhibitory cystine knot motif. *Front Microbiol* (2012) 3:324. doi:10.3389/fmicb.2012.00324
90. Watt PM. Screening for peptide drugs from the natural repertoire of biodiverse protein folds. *Nat Biotechnol* (2006) 24:177–83. doi:10.1038/nbt1190
91. Liu Y. Ubiquitin ligases and the immune response. *Annu Rev Immunol* (2004) 22:81–127. doi:10.1146/annurev.immunol.22.012703.104813

Conflict of Interest Statement: 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.

Copyright © 2017 Candido-Ferreira, Kronenberger, Sayegh, Batista and da Silva Junior. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



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

Arjun Balakrishnan¹, Markus Schnare² and Dipshikha Chakravorty^{3*}

¹ Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore, India, ² Institute for Immunology, University of Marburg, Marburg, Germany, ³ Centre for Biosystems Science and Engineering, Indian Institute of Science, Bangalore, India

OPEN ACCESS

Edited by:

Larry J. Dishaw,
University of South Florida St.
Petersburg, USA

Reviewed by:

Mikhail A. Gavrilin,
Ohio State University, USA
Kenneth Reid,
University of Oxford, UK
Paola Italiani,
National Research Council, Italy

*Correspondence:

Dipshikha Chakravorty
dipa@mcbl.iisc.ernet.in

Specialty section:

This article was submitted to
Molecular Innate Immunity,
a section of the journal
Frontiers in Immunology

Received: 18 August 2016

Accepted: 11 October 2016

Published: 24 October 2016

Citation:

Balakrishnan A, Schnare M and
Chakravorty D (2016) 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. 7:455.
doi: 10.3389/fimmu.2016.00455

Macrophages as immune cells prevent the spreading of pathogens by means of active phagocytosis and killing. We report here the presence of an antimicrobial protein, bactericidal/permeability-increasing protein (BPI) in human macrophages, which actively participates in engulfment and killing of Gram-negative pathogens. Our studies revealed increased expression of BPI in human macrophages during bacterial infection and upon stimulation with various pathogen-associated molecular patterns, viz., LPS and flagellin. Furthermore, during the course of an infection, BPI interacted with Gram-negative bacteria, resulting in enhanced phagocytosis and subsequent control of the bacterial replication. However, it was observed that bacteria which can maintain an active replicating niche (*Salmonella* Typhimurium) avoid the interaction with BPI during later stages of infection. On the other hand, *Salmonella* mutants, which cannot maintain a replicating niche, as well as *Shigella flexneri*, which quit the endosomal vesicle, showed interaction with BPI. These results propose an active role of BPI in Gram-negative bacterial clearance by human macrophages.

Keywords: innate immunity, Gram-negative bacteria, macrophage evolution, bacterial niche, phagocytic receptor, antimicrobial protein

INTRODUCTION

Innate immune responses refer to the first line of non-specific defense mechanisms that get activated immediately upon encounter with the pathogen. Once a pathogen comes in contact with the host innate immune cells, various sets of genes are upregulated, whose products play an important role in defense mechanisms. These defense mechanisms are classically categorized as O₂-dependent and O₂-independent modes of bacterial killing. O₂-dependent bactericidal activity is mediated by the NADPH phagocyte oxidase and inducible nitric oxide synthase pathways, whereas O₂-independent bactericidal activity is mediated by antimicrobial peptides and proteins. Bactericidal/permeability-increasing protein (BPI) is a 55-kDa antimicrobial protein with multiple functions including bacterial killing, bacterial opsonization, and LPS neutralization (1). BPI is primarily known to be expressed

in human neutrophils and epithelial cells. Previous studies have shown that among innate immune cells, murine BPI is expressed only in dendritic cells and neutrophils but not in macrophages (2). Based on these results, no further studies have been carried out to understand the expression of BPI in macrophages. However, murine macrophages unlike human macrophages are strong producers of nitric oxide and can kill invading pathogens by oxidative stress (3–5). We assumed that in contrast to murine macrophages, human macrophages in compensation for the relatively weak nitric oxide production may employ stronger and more diverse antimicrobial protein production to restrict the spread of invading pathogens. To evaluate this hypothesis, we studied the expression of BPI in human monocytes and macrophages as BPI is the principal O_2 -independent bactericidal agent that acts against Gram-negative bacteria in human neutrophils (6).

In this report, we have analyzed BPI expression in murine and human macrophages under various inflammatory conditions. We investigated the potential role of BPI as an antibacterial agent in human macrophages. Surprisingly, we show that BPI is expressed in human macrophages. In addition to its role as an antibacterial agent, BPI expressed in human macrophages can mediate uptake of Gram-negative bacteria. Gram-negative bacteria which can maintain an active replicating niche in human macrophages avoid the interaction with BPI during later stages of an infection. Together, these results suggest an active role of BPI in Gram-negative bacterial internalization and restricting Gram-negative bacterial replication in human macrophages.

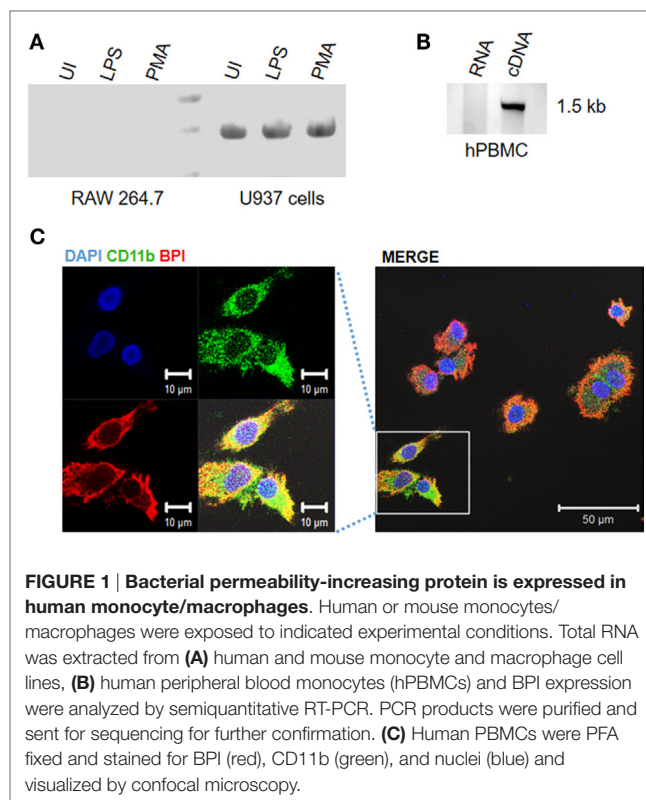
RESULTS

BPI Expression in Human Monocytes

To understand whether BPI is differentially expressed in human and murine macrophages, BPI expression in murine and human macrophage cell lines were compared in the presence of LPS and PMA (Figure 1A). Surprisingly, BPI mRNA was detected in a human monocyte cell line under resting condition. Furthermore, BPI was found to be expressed in PMA stimulated U937 cells as well, indicating BPI expression in differentiated macrophages (Figure 1A). To validate BPI expression in human PBMCs, BPI-full length product from human PBMCs was amplified and sequenced (Figure 1B). Sequencing results showed >97% sequence identity to BPI-encoding DNA sequence (Figure S1A in Supplementary Material). Additionally, immunostaining with a BPI-specific antibody in human PBMCs revealed predominant localization of BPI toward cell surface (Figure 1C). CD11b staining of human PBMCs further confirmed the presence of BPI in human PBMCs derived macrophages as well as the colocalization of BPI with the surface molecule CD11b (Figure 1C). These results suggest that BPI is expressed in human but not murine macrophages.

Regulation of BPI Expression in Human Monocytes

To understand whether BPI expression varies during the course of infection, BPI expression in human monocytes was analyzed under different inflammatory conditions. To this end, U937 cells



were infected with different bacteria [*Salmonella* Typhimurium (STM), *Staphylococcus aureus* (SA), and *Salmonella* Typhi (STY)] or were incubated with different PAMPS (LPS and flagellin); afterward, total RNA was isolated and BPI expression was quantified by real-time PCR (Figure 2A). Furthermore, in order to analyze the protein expression of BPI, the cells were fixed and stained with an anti-BPI antibody and investigated by flow cytometry (Figure 2C). These analyses demonstrated that the expression of BPI in human monocytes remained unchanged under all the tested inflammatory conditions.

Bactericidal/permeability-increasing protein is known to be released into inflammatory exudates (7). To understand whether BPI is secreted by human monocytes or macrophages, U937 cells were PMA stimulated or infected with different pathogens (STM and STY). Cell culture supernatant was collected 24 h, post-treatment, and BPI levels were determined by ELISA (Figure S1B in Supplementary Material). There was no detectable level of BPI in cell culture supernatant, indicating that BPI is not secreted by human monocytes or macrophages.

Regulation of BPI Expression in Differentiated Macrophages

Differentiation of monocytes to macrophages is known to induce an inflammatory and antibacterial response in macrophages (8). To determine whether the increased antibacterial activity of human macrophages is due to increased expression of BPI, U937 cells were differentiated into macrophages by treating monocytes with PMA (50 nM) for 24 h. As expected, differentiated macrophages

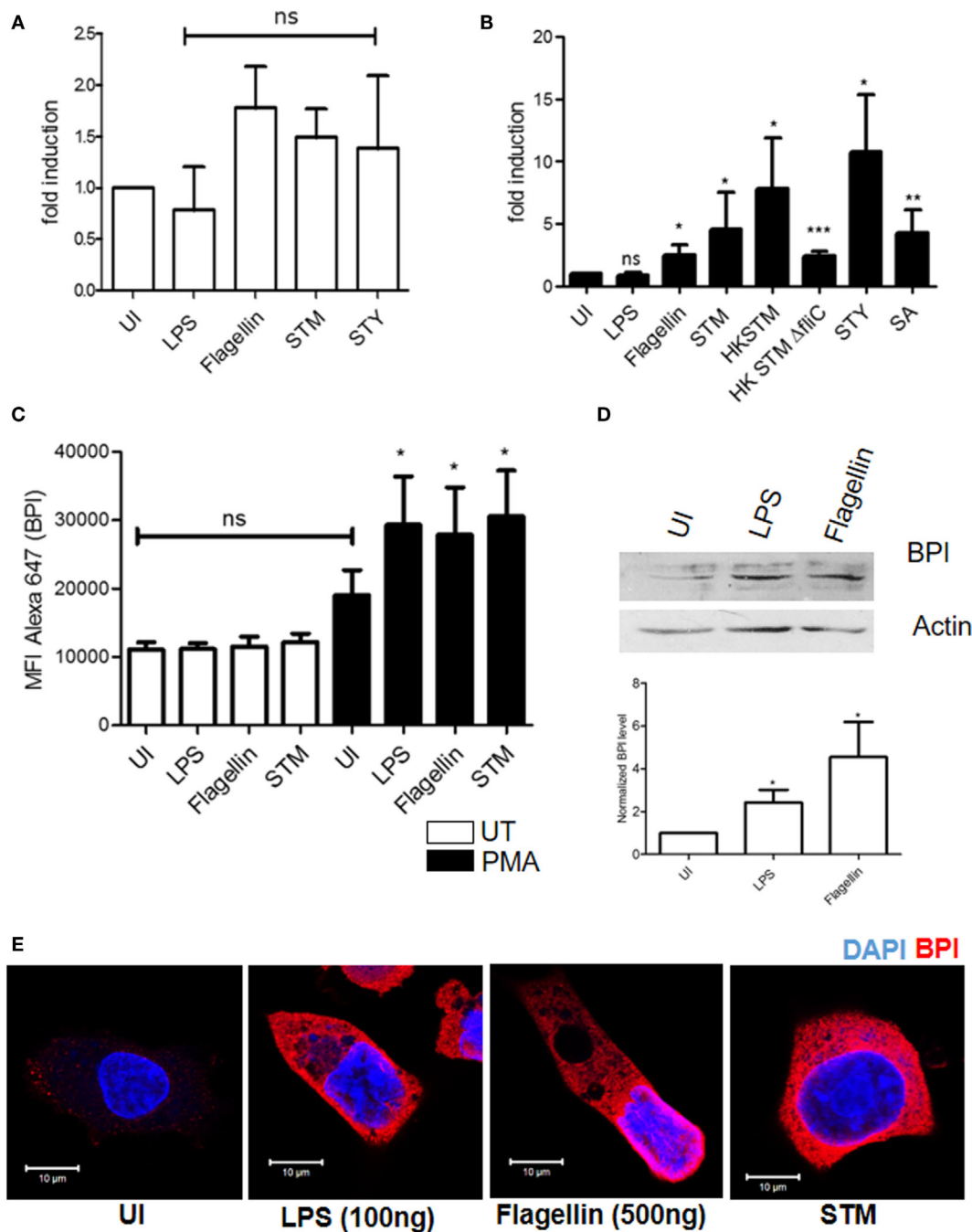


FIGURE 2 | Regulation of BPI expression in human monocytes/macrophages. U937 monocytes/macrophages were exposed to indicated inflammatory stimuli/pathogen. **(A,B)** Total RNA was isolated and BPI expression was investigated by real-time PCR [$n = 5$ (SD)]. **(C)** BPI expression was checked by flow cytometry [$n = 3$ (SD)]. **(D)** Total protein was isolated from U937 macrophage, and BPI levels were analyzed by western blot ($n = 4$). Bottom: quantitative evaluation of the relative BPI expression normalized to β -actin intensity. **(E)** Cells were PFA fixed and stained for BPI (red) and nuclei (blue). Cells were imaged by confocal microscopy ($n = 6$). Key: *** $p < 0.001$, ** $p < 0.005$, * $p < 0.05$; ns, not significant.

showed an increased antibacterial activity toward STM compared to U937 monocytes (Figure S2B in Supplementary Material). Interestingly, BPI mRNA expression significantly increased in differentiated macrophages upon treatment with various PAMPs (LPS and flagellin) or infection with various pathogens (STM, STY,

and SA) (Figure 2B). BPI expression was increased up to three-fold in the presence of flagellin compared to untreated control. In contrast, the BPI expression remained unchanged in the presence of LPS. To evaluate the contribution of flagellin in inducing BPI expression, U937 macrophages were incubated with heat-killed

strains of STM [flagellin-deficient *Salmonella* Typhimurium 14028 (STM Δ fliC)] and a non-motile Gram-positive pathogen (SA). Heat-killed STM showed increased BPI expression when compared to either untreated controls or HK STM Δ fliC and live bacteria. Nevertheless, BPI expression was significantly increased by STM Δ fliC and SA compared to untreated control. These results indicate that several PAMPs other than flagellin can also contribute to the BPI expression during infection. Under uninfected conditions, there was no significant increase in BPI mRNA expression in differentiated macrophages compared to U937 monocytes (Figure S2A in Supplementary Material).

In order to evaluate BPI expression at the protein level, U937 macrophages were incubated in the presence of various PAMPs (LPS and flagellin) and bacteria (STM) for 24 h. Thereafter, the cells were fixed and the BPI expression was checked by flow cytometry as well as confocal microscopy (Figures 2C,E). BPI expression was significantly increased in the presence of bacteria as well as PAMPs. LPS induced BPI expression at the protein level, even though there was no significant induction at the RNA level. Western blot analysis showed 2.5-fold increase in BPI expression after LPS treatment and 4.5-fold after flagellin treatment (Figure 2D). These results confirm that BPI is induced in differentiated macrophages during the course of infection.

BPI Enhances Bactericidal Activity of Human Macrophages

Bactericidal/permeability-increasing protein is known to inhibit the growth of Gram-negative bacteria (9). In order to understand whether BPI expressed in U937 macrophages is functionally active, an antibacterial assay was carried out. U937 macrophages were treated with various PAMPs, which were shown to induce BPI expression (LPS 100 ng and flagellin 500 ng). Twenty-four hours post-treatment, cells were infected with STM 14028 at a multiplicity of infection (MOI) of 10. Bacterial replication was quantified by plating cell lysates 2 and 16 h post-infection. Conditions which induced BPI expression significantly affected bacterial growth in U937 macrophages (Figure S3 in Supplementary Material). To understand the contribution of BPI in inhibiting bacterial growth, bacterial replication was assessed after knocking down BPI in U937 macrophages. To knock down BPI, U937 macrophages were transfected with BPI dsRNA. Twenty-four hours post-transfection, cells were infected with STM 14028 at MOI of 10. The efficacy of BPI knockdown in dsRNA transfected cells was validated by western blot analysis (Figure 3C) and confocal microscopy (Figure S4 in Supplementary Material). Bacterial replication was quantified by plating cell lysates 2 and 16 h post-infection (Figure 3A). Bacterial replication significantly increased in cells where BPI expression was strongly reduced due to dsRNA knockdown compared to untransfected controls. To confirm that BPI expressed in primary human macrophages can inhibit bacterial replication, bacterial replication was assessed in primary human PBMCs after knocking down BPI. Bacterial replication significantly increased in BPI low expressing cells due to dsRNA transfection compared to scrambled dsRNA control (Figure 3B). These data show that BPI contributes significantly in limiting bacterial replication in human macrophages.

Antibacterial activity of BPI is specific toward Gram-negative bacteria due to the specific interaction between BPI and Gram-negative bacterial LPS (6, 10). To validate the specificity of BPI knockdown, Gram-positive bacterial replication was assessed after knocking down BPI in U937 macrophages. U937 macrophages were transfected with BPI dsRNA; 24-h post-transfection, cells were infected with *S. aureus* at an MOI of 10. Bacterial replication was quantified by plating infected cell lysates after 2 and 16 h post-infection (Figure 3A). Replication of *S. aureus* remained unaffected after knocking down BPI in U937 macrophages indicating the specificity of BPI activity to clear Gram-negative pathogens in human macrophages. In order to understand the importance of human BPI in inhibiting bacterial replication, human BPI was amplified from macrophages by PCR, cloned, and expressed in murine macrophages lacking endogenous BPI. Expression of human BPI in RAW 264.7 cells was confirmed by western blotting (Figure 3E). Overexpression of human BPI in murine macrophages increased their antibacterial activity, suggesting that human BPI expressed in human macrophages significantly contributes in bacterial killing (Figure 3D). These results suggest that human BPI actively contributes to the clearance of Gram-negative bacteria in macrophages.

BPI Mediates Phagocytic Uptake of Gram-Negative Bacteria by Human Macrophages

In 1997, the role of BPI as an opsonin in human neutrophils was published (11). We observed a significant interaction of surface BPI with Gram-negative bacteria during early time points of infection (Figure 4A). Interestingly, flow cytometry analysis and confocal microscopic analysis of non-permeabilized U937 macrophages showed the presence of BPI on the cell surface. Furthermore, the cell surface-associated BPI interacted clearly with STM (Figure S5 in Supplementary Material). Based on these observations, we hypothesized that BPI expressed on human macrophages might act as a receptor that enhances the phagocytic activity of macrophages toward Gram-negative bacteria. To validate this hypothesis, phagocytosis of Gram-negative bacteria by macrophages was quantified after knocking down BPI in U937 macrophages. U937 macrophages were transfected with BPI dsRNA; 24-h post-transfection, cells were infected with STM 14028 at MOI of 10. Thereafter, phagocytosis of STM 14028 by macrophages was calculated by plating the macrophage cell lysate 30 min post-infection (Figure 4B). We found that the uptake of STM 14028 was significantly decreased upon BPI knockdown in U937 macrophages compared to untransfected control. Interestingly, BPI knockdown did not affect uptake of *S. aureus* (Gram-positive bacteria) by U937 macrophages (Figure 4B). To confirm the role of BPI in Gram-negative bacterial phagocytosis by primary human macrophages, bacterial phagocytosis was assessed in human PBMCs after knocking down BPI. Bacterial uptake was significantly affected upon BPI knockdown in macrophages derived from human PBMCs (Figure 4C). These results suggest that surface expressed BPI contributes significantly in Gram-negative bacterial phagocytosis by human macrophages.

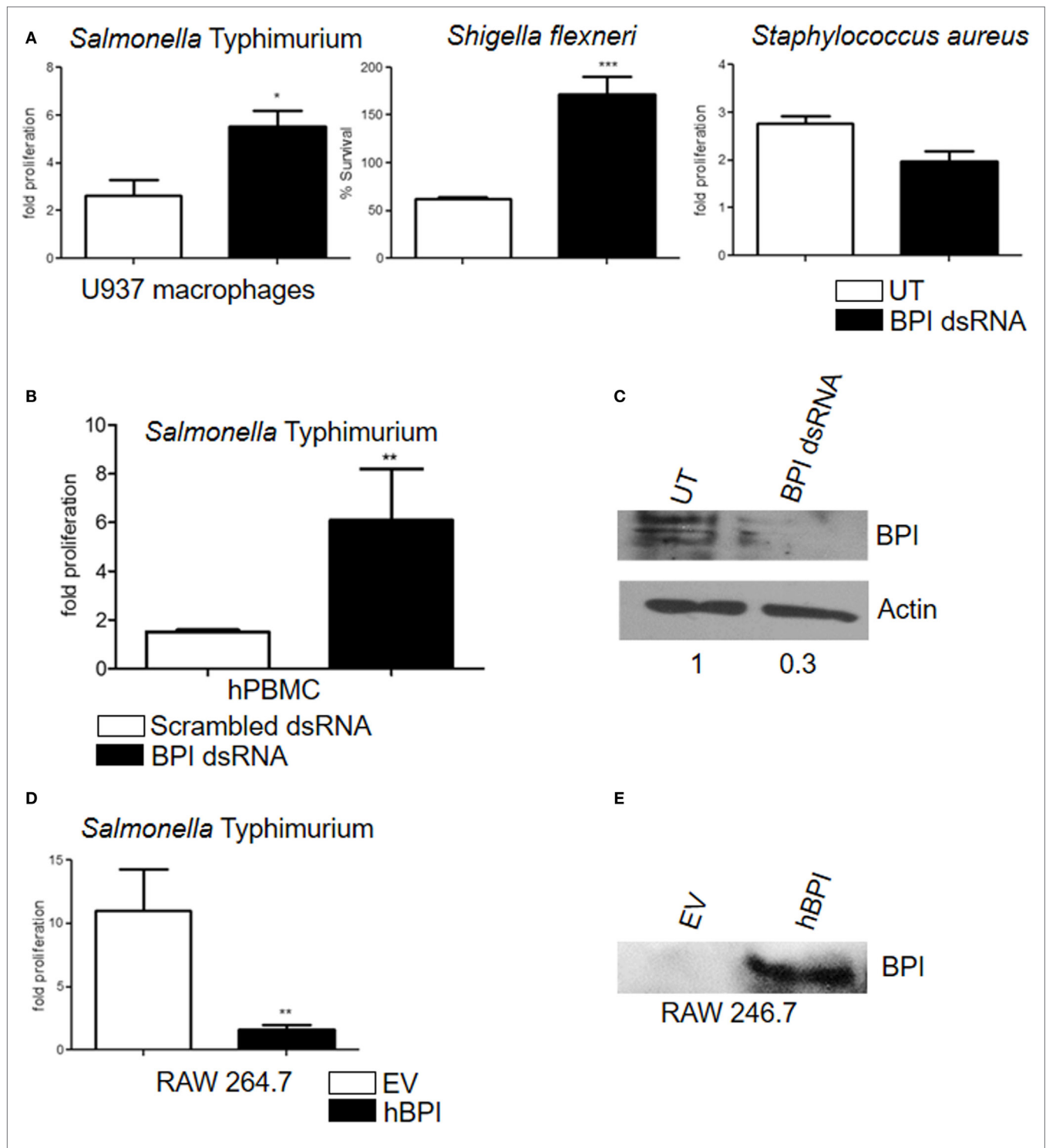


FIGURE 3 | Human macrophages expressed BPI and kills intracellular bacteria. (A) Fold proliferation and percentage survival of bacteria after knocking down of BPI in U937 macrophages. Statistical significance was calculated with respect to untransfected control [$n = 5$ (SD)]. **(B)** Fold proliferation of STM after knocking down of BPI in human PBMCs-derived macrophages. Statistical significance was calculated with respect to scrambled dsRNA-transfected control [$n = 6$ (SD)]. **(C)** Total protein was isolated from BPI dsRNA-transfected and untransfected control U937 macrophages, and BPI levels were checked by western blot ($n = 3$). **(D)** RAW 264.7 macrophages were transfected with either pcDNA empty vector (pcDNA EV) or pcDNA carrying the expression sequence of human BPI (pcDNA hBPI). Twenty-four hours post-transfection, the cells were infected with STM and bacterial proliferation was quantified [$n = 4$ (SD)]. **(E)** Total protein was isolated from pcDNA EV-transfected and pcDNA hBPI-transfected RAW 264.7 macrophages, and BPI levels were quantified by western blot ($n = 3$). Key: *** $p < 0.001$, ** $p < 0.005$, * $p < 0.05$; ns, not significant.

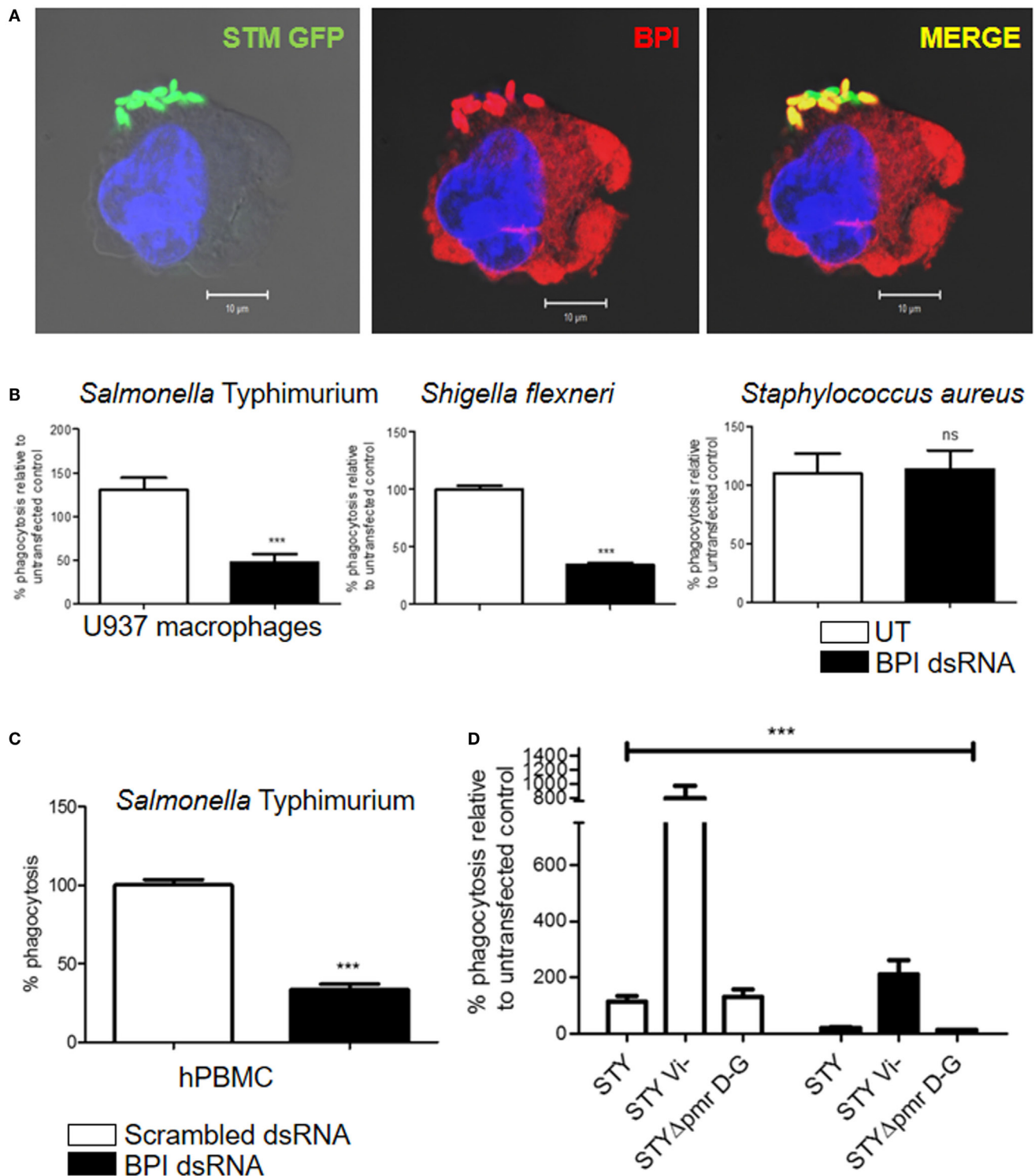


FIGURE 4 | BPI enhances bacterial uptake in human macrophages. (A) U937 macrophages were infected with STM-GFP (green) at an MOI of 10. Thirty minutes post-infection, cells were fixed with paraformaldehyde and stained for BPI (red) and nuclei (blue) ($n = 4$ experiments). **(B,D)** Percentage phagocytosis of bacteria after knocking down of BPI in U937 macrophages. U937 macrophages were infected with the indicated bacteria at an MOI of 10. Bacterial entry was quantified by plating cell lysates after 30 min post-infection [$n = 6$ (SD)]. For experiments with *Salmonella Typhi* (STY), percentage phagocytosis of BPI dsRNA transfected cells was compared to untransfected control and was normalized to STY (WT). **(C)** Percentage phagocytosis of STM in hPBMC-derived macrophages. Statistical significance was calculated with respect to scrambled dsRNA-transfected control [$n = 6$ (SD)]. Key: *** $p < 0.001$, ** $p < 0.005$, * $p < 0.05$; ns, not significant.

Many pathogenic bacteria are known to possess outer membranous structures that help to evade phagocytosis by macrophages. STY has Vi-polysaccharide that resists opsonophagocytosis mediated by complement receptors (12). The Vi-polysaccharide is also known to inhibit the TLR4-mediated innate immune response (13). BPI and TLR4 recognize lipid A moieties on the surface of bacteria (14). We hypothesized that the presence of Vi capsular polysaccharide might inhibit BPI-mediated phagocytosis of STY. In order to understand the importance of Vi-polysaccharide in BPI-mediated phagocytosis, we checked the percentage phagocytosis of Vi-negative *Salmonella* Typhi (STY Vi⁻) by macrophages in the presence or absence of BPI. BPI was knocked down in U937 macrophages as explained above. U937 macrophages were infected with STY, STY Vi⁻, and STY $\Delta pmrDG$. Percentage phagocytosis was calculated by plating the cell lysate 30 min post-infection. Percentage phagocytosis was significantly higher for *Salmonella* devoid of Vi-polysaccharides (STY Vi⁻) compared to STY, but bacterial phagocytosis was significantly decreased irrespective of the presence or the absence of Vi polysaccharide after knocking down BPI in human macrophages (Figure 4D). STY $\Delta pmrDG$ was used as a negative control in this experiment as the *pmr* operon is important for structural modifications in LPS but is not important for preventing phagocytosis by macrophages [Figure 4D; (15)]. These results suggest that capsular polysaccharide, although very important to inhibit phagocytosis by macrophages, in general, will not affect the BPI-mediated phagocytosis of Gram-negative bacteria.

***Salmonella* Typhimurium Evades BPI Interaction during Later Stages of Infection**

We next analyzed the intracellular interaction of BPI with STM by confocal microscopy during the course of infection. Therefore, U937 macrophages were infected with STM14028 at an MOI of 50 and bacterial colocalization with BPI at different time points was analyzed. A region of interest (ROI) was drawn around each bacterium based on GFP signal and % colocalization of a bacterium and BPI at ROI was quantified. STM showed significantly higher colocalization with BPI at early time points of infection (15 min to 1 h). Interestingly, during later time points of infection (2–6 h) by which bacteria maintain a proper niche inside the macrophages, *Salmonella*-containing vesicles (SCVs), a significant lesser colocalization with BPI, could be observed (Figure 5A). The time course of STM replication in macrophages with knocked down BPI showed that STM replication was higher in knocked down conditions within 6 h post-infection compared to untransfected controls. These data are in accordance with previous reports, which suggest that replication of STM takes place starting 6 h post-infection by which the bacteria maintain an actively replicating niche inside the macrophage (16) (Figure 5B). STM replication was significantly higher in BPI KD conditions from 6 to 24 h compared to untransfected control, even though we see a significant decrease in the bacterial entry in BPI KD cells compared to untransfected controls (Figure 5C).

We next tried to understand the importance of maintaining an actively replicating niche by STM (SCV) to avoid interaction

with BPI in human macrophages. U937 macrophages were infected with GFP-tagged bacteria, either replicating STM, paraformaldehyde fixed STM (PFA STM), *Escherichia coli* DH5 α (which cannot replicate inside macrophages), or *Shigella flexneri* (SHG; which quits endosomal vesicle). Two hours post-infection, the cells were fixed and the interaction of BPI with the bacteria was analyzed by confocal microscopy (Figure 6A). Recruitment of BPI to the bacteria was analyzed by two methods. First, we checked the interaction of BPI with the bacteria by analyzing the percentage colocalization of BPI and GFP at ROI as explained above (Figure 6B). Second, we analyzed the mean fluorescent intensity (MFI) of BPI at ROI to understand the recruitment of BPI to the bacteria (Figure 6C). There was a significant increase in the recruitment of BPI measured by MFI as well as % colocalization of BPI with PFA STM compared to STM. This may indicate that *Salmonella* actively inhibits the recruitment of BPI to SCV during later stages of infection. Bacteria which cannot replicate inside macrophages (*E. coli*) and bacteria which quit the endosomal vesicle (SHG) showed significantly higher interaction with BPI compared to STM (Figures 6A–C). BPI was found to be localized along the surface of SHG (Figure 6A, inset). In order to evaluate the survival of SHG in human macrophages, we checked the entry as well as the replication of SHG in human macrophages after knocking down BPI as explained above. SHG entry significantly decreased in U937 macrophages upon knockdown of BPI compared to untransfected control (Figure 4B). Interestingly, SHG which usually gets cleared in human macrophages was able to replicate in human macrophages upon knockdown of BPI (Figure 3A). Percentage survival was calculated by normalizing bacterial CFU from 18 h to the CFU count 2 h post-infection. SHG percentage survival was increased from 50 to 150% indicating the importance of BPI in clearing cytosolic bacteria in human macrophages. BPI levels were detected by western blotting after infection with STM, SHG, and *E. coli* in U937 macrophages (Figure S6 in Supplementary Material). There was no significant difference in BPI levels in STM infected cells compared to SHG infected cells. These results indicate that the differential interaction of BPI with STM and SHG is not due to differential expression or degradation of BPI.

***Salmonella* Typhimurium Maintains an Actively Replicating Niche in Order to Evade BPI Interaction**

Salmonella Typhimurium is known to maintain an actively replicating niche inside the macrophage by modifying the endosomal membrane, thereby preventing their fusion with the late lysosome (SCV) (16). SCV actively modifies its membrane-associated proteins, and these modifications are important for the survival of STM in macrophages. In our present study, we observed that BPI interacts significantly stronger with cytosolic bacteria (SHG) and PFA-fixed STM compared to live STM. These observations led us to hypothesize that SCV might actively avoid the interaction of BPI with the bacteria inside macrophages. To evaluate this hypothesis, we checked the interaction of BPI with STM $\Delta sifA$, a *Salmonella* mutant, which cannot maintain an actively replicating niche (SCV) (17). Two hours post-infection,

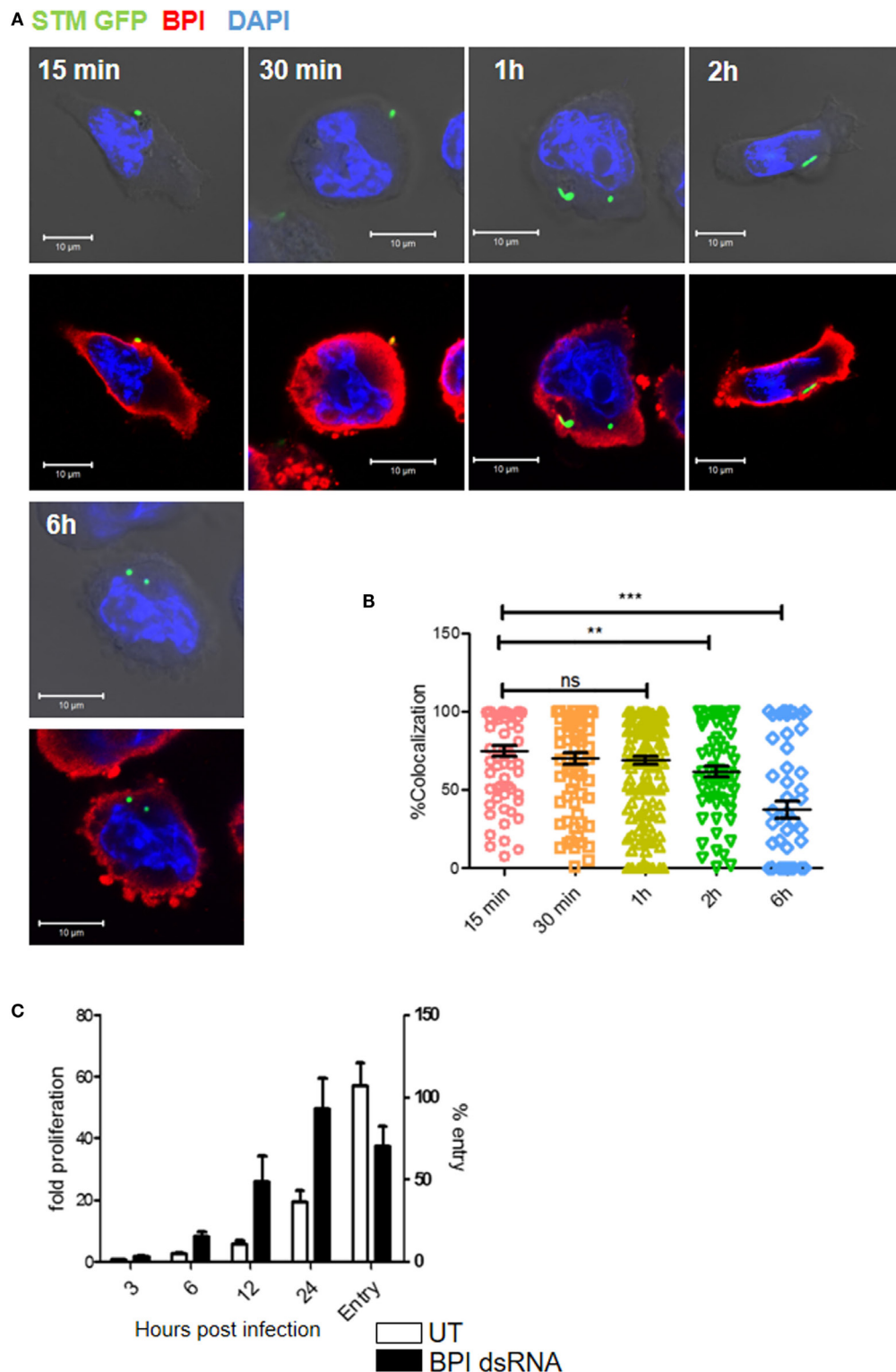


FIGURE 5 | Time-course analysis of BPI interaction with *Salmonella* Typhimurium in U937 macrophages during the course of infection. (A) U937 macrophages were infected with STM-GFP (green) at an MOI of 50. Cells were fixed with PFA at the indicated time points and stained for BPI (red) and nuclei (blue) ($n = 3$). Panel shows representative bright field and merged images for each time point. **(B)** Quantification of colocalization of BPI with STM-GFP [$n = 3$ (SD)]. **(C)** U937 macrophages were infected with STM-GFP at an MOI of 10. Bacterial entry was quantified by plating the cell lysates after 30 min post-infection. Bacterial replication was quantified by plating the cell lysates at indicated time points and was normalized to the CFU at 1 h post-infection. Fold proliferation and percentage survival were calculated [$n = 3$ (SD)]. Key: *** $p < 0.001$, ** $p < 0.005$, * $p < 0.05$; ns, not significant.

cells were fixed, bacteria and cells were stained with DAPI. An ROI was drawn around each bacterium marked upon DAPI staining (**Figures 7A,B**). MFI of BPI as well as % colocalization of BPI and bacteria was evaluated at ROI (**Figures 7C,D**). LAMP 2 was used as a marker to confirm whether STM Δ sifA maintains in

an intracellular vesicle or not (**Figures 7A,B**). STM Δ sifA mostly remained in the cytoplasm compared to STM wild type, which was in LAMP2-positive compartments as analyzed by LAMP2 colocalization (**Figure 7E**). STM Δ sifA showed an increased colocalization with BPI compared to STM wild type, indicating

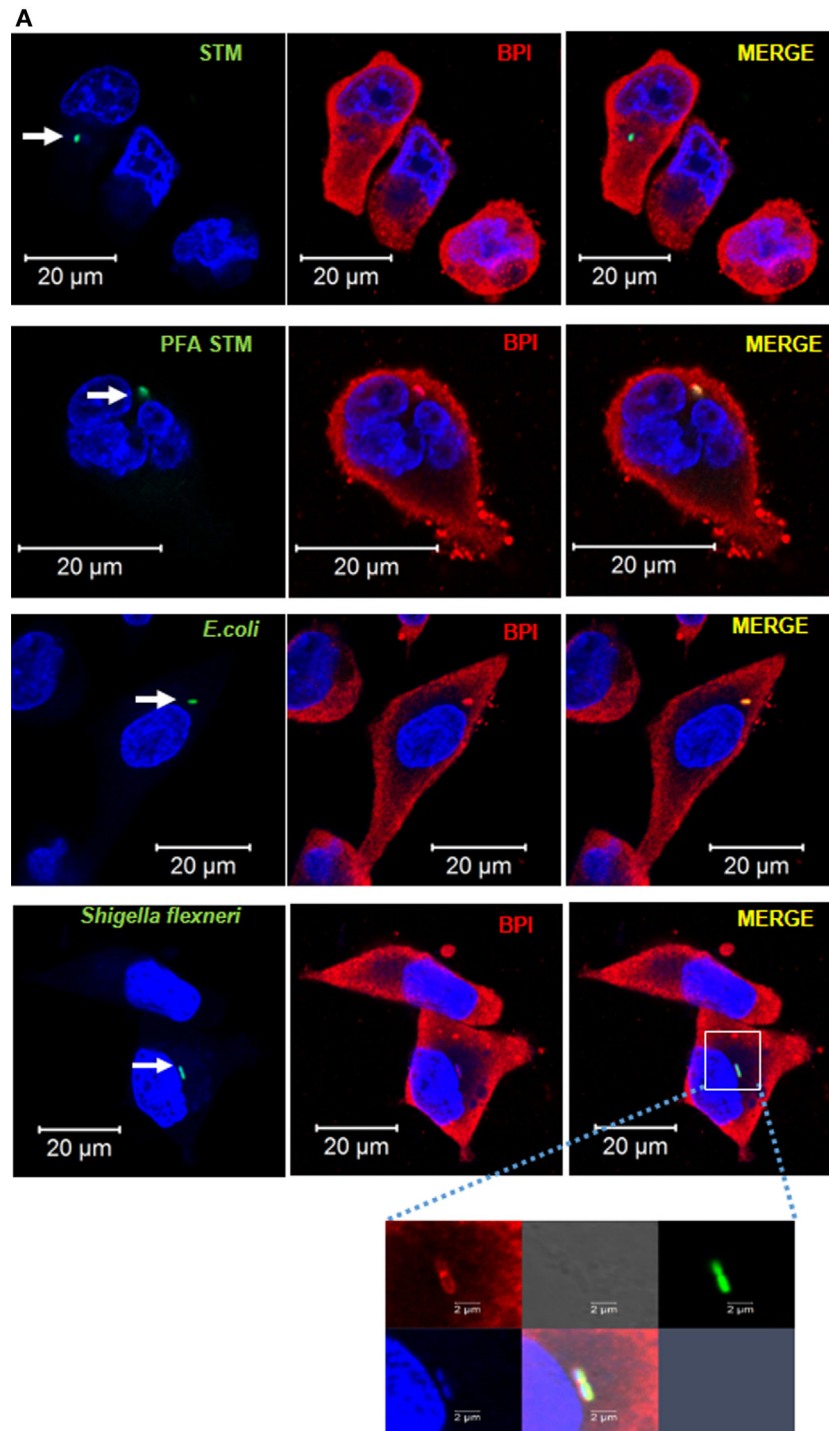


FIGURE 6 | BPI interaction with Gram-negative bacteria inside macrophages.

(Continued)

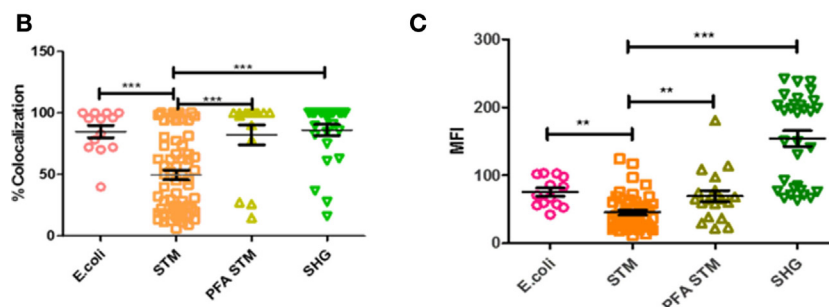


FIGURE 6 | Continued

(A) U937 macrophages were infected with GFP-tagged (green) Gram-negative bacteria [STM14028, PFA fixed STM, *E. coli* DH5 α , and *Shigella flexneri* (SHG)] at an MOI of 50. Two hours post-infection, cells were fixed with PFA and stained for BPI (red) and nuclei (blue). White arrows indicate GFP-positive bacteria. The boxed area in the SHG infected set is magnified to view BPI around the bacteria ($n = 3$). **(B)** Quantification of colocalization of BPI with the bacteria. **(C)** Quantification of MFI of BPI at ROI was done by the Zen Blue edition software provided by Zeiss [$n = 3$ (SD)]. Key: *** $p < 0.001$, ** $p < 0.005$, * $p < 0.05$.

the importance of the vacuolar life of *Salmonella* in maintaining a replicative niche devoid of BPI inside macrophages. STM Δ sifA showed an increased recruitment of BPI around the bacteria measured by evaluating MFI as well as the percentage colocalization at ROI (**Figures 7C,D**).

To confirm these results, we checked bacterial interaction with BPI under conditions which make bacteria quit the vesicle. To attain this, we expressed listeriolysin (LLO) in STM. LLO is known to make pores into vesicular membranes, which lead to the rupturing of the membrane (18). LLO expression was induced in STM LLO using IPTG. U937 macrophages were infected with STM EV (empty vector) or STM LLO. Two hours post-infection, cells were fixed and bacteria and cells were stained with DAPI (**Figures 7E,G**). An ROI was drawn around each bacteria marked upon DAPI staining. MFI of BPI, as well as % colocalization of BPI and bacteria, was evaluated at ROI. BPI and STM LLO colocalized significantly higher compared to STM EV as seen by increased % colocalization in STM LLO in comparison to STM EV (**Figure 7H**). Recruitment of BPI as measured by checking the MFI at ROI was also higher in STM LLO compared to STM EV (**Figure 7I**). Under these conditions, STM LLO showed a twofold decrease in bacterial proliferation compared to STM EV control (Figure S7 in Supplementary Material). These results suggest that STM avoids interaction with BPI by maintaining an actively replicating niche (SCV) inside macrophages.

DISCUSSION

Bactericidal/permeability-increasing protein is known to be expressed in human neutrophils, epithelial cells, eosinophils, and the genital tract (1). Previous studies by Buurman et al. (19) suggested the expression of BPI in human monocytes, but the experimental evidence clearly did not prove whether BPI is expressed in human monocytes or whether BPI is adsorbed on to the monocyte surface. In this current study, we clearly demonstrate the expression of BPI in human macrophages both at RNA and protein levels. Interestingly and in sharp contrast, under similar conditions, we could not detect BPI expression in murine macrophages. The reactive nitrogen species (RNS)-mediated

antibacterial activity of human macrophages is still a controversy (20). On the one hand, there are reports which suggest that unlike murine macrophages, human macrophages cannot produce RNS (5). On the other hand, RNS could be detected in human macrophages isolated from PBMCs of infected patients. Regardless these reports, in summary, suggest that NO production by human macrophages is either extremely low or requires complex signaling mechanisms for induction (4, 5). Our results on the expression of BPI only in human macrophages suggest that human macrophages during the course of evolution might have obtained a strong bias toward O_2 -independent mechanisms to kill pathogens as exemplified by BPI. This might give an advantage to the host to avoid free radical-mediated damage to the host cells associated with reactive nitrogen species produced during infection and in circumstances when oxygen tension is low.

Differentiation of monocytes to macrophages is known to increase the antibacterial activity of human macrophages (8). The precise mechanistic aspects of increased antibacterial activity of differentiated macrophages are not known. In our current study, we show that BPI expression is increased in differentiated macrophages upon bacterial infection, whereas in undifferentiated monocytes, there was no significant increase in BPI expression. Interestingly, knocking down of BPI in macrophages led to the proliferation of three different strains of Gram-negative bacteria tested, but did not affect proliferation of *S. aureus*. The signaling pathways that lead to BPI induction in human macrophages are not clear. In our study, we saw that all the PAMPs tested can induce BPI expression. These results show that BPI might be induced as a general antimicrobial protein during the infection by any pathogen in macrophages. Interestingly, LPS increase BPI protein level without changing the BPI mRNA level. This might be due to an increase in the stability of BPI mRNA mediated by MyD88 signaling pathway (21).

Previous results by Weiss and group demonstrated that BPI secreted by neutrophils can act as an opsonin and can induce opsonophagocytosis by macrophages (11). We could not detect BPI secreted by human macrophages upon infection. Interestingly, most of the BPI expressed in human macrophages is present on the cell surface. During the early stages of infection, BPI

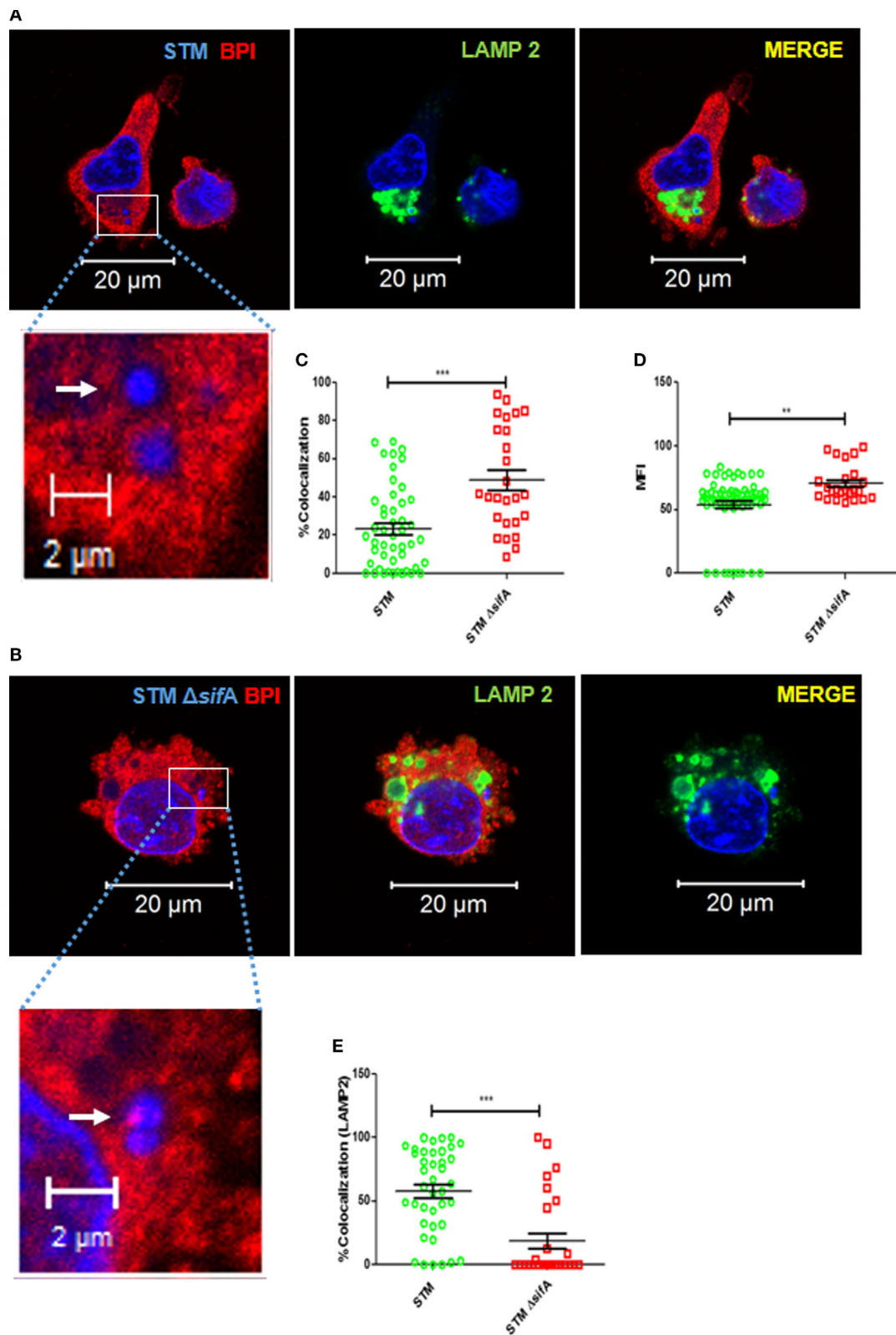


FIGURE 7 | STM avoids interaction with BPI inside macrophages by maintaining an actively replicating niche (SCV).

(Continued)

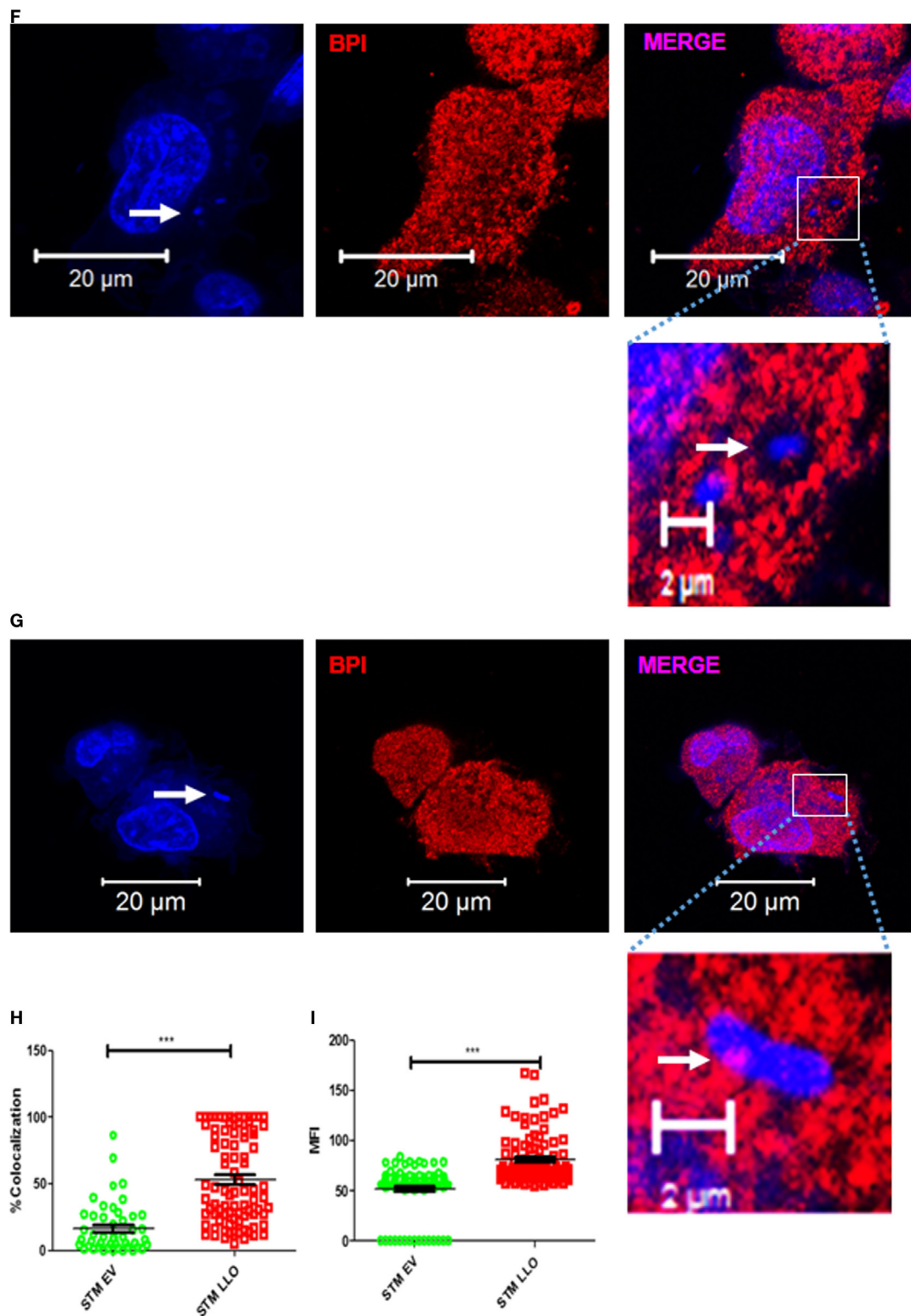


FIGURE 7 | Continued

U937 macrophages were infected with (A) STM, (B) STM Δ siA, (F) STM EV (empty vector), or (G) STM LLO at an MOI of 50. Two hours post-infection, the cells were fixed with PFA at the indicated time points and stained for BPI (red) and LAMP2 (green). Nuclei and bacteria were labeled with 4',6-diamidino-2-phenylindole (DAPI) (blue). White arrows indicate DAPI-positive bacteria. The boxed area in each set is magnified to view BPI around the bacteria. (C,H) Quantification of colocalization of BPI with bacteria. (D,I) Quantification of MFI of BPI at ROI. (E) Quantification of colocalization of LAMP2 with bacteria at ROI. All images were quantified by using the Zen Blue edition software provided by Zeiss [$n = 3$ (SD)]. Key: *** $p < 0.001$, ** $p < 0.005$.

significantly interacted with Gram-negative bacteria (**Figure 5B**). Knocking down of BPI in human macrophages led to a reduced entry of Gram-negative bacteria. These results may suggest that apart from acting as an opsonin, BPI expressed on the cell surface can itself act as a receptor for binding of Gram-negative bacteria. Whether this binding can itself induce phagocytosis or whether BPI helps the bacteria to adhere to macrophages and other phagocytic receptors distinct from the interaction with BPI can induce phagocytosis of bacteria is not clear. Recently, a report by Casanova et al. showed that GPCR brain-specific angiogenesis inhibitor 1 (BAI 1) can act as a receptor for the uptake of Gram-negative bacteria and can also induce bacterial killing by indirect means (22). BPI may also act as a phagocytic receptor, but unlike BAI 1, BPI itself can act as a direct bactericidal agent. Irrespective of the presence or absence of capsular material, BPI can induce phagocytosis of Gram-negative bacteria as seen by phagocytosis of Vi-positive *S. Typhi*. How antimicrobial peptides and proteins can interact with LPS of capsulated bacteria is still not clear. We believe that the extensions of LPS outside capsular polysaccharide might act as a docking site for BPI (11). Whether BPI can interact with the capsular polysaccharide itself is something that should be explored in the future. Since BPI-mediated phagocytosis did not need any opsonin, it will be interesting to check the level of BPI in macrophages which are present in opsonin poor environment (e.g., alveolar macrophages). BPI significantly interacted with STM during early time points of the infection. During later time points of the infection, STM avoids BPI interaction during later time points and thereby maintains an actively replicating niche inside macrophages. Bacteria which cannot maintain an actively replicating niche (SHG, *E. coli*) will be cleared easily by the macrophages. All these bacteria tested showed significant colocalization with BPI during later time points of infection as well. Interestingly, SHG which cannot multiply in human macrophages started multiplying in BPI knockdown macrophages. Confocal analysis showed the presence of BPI around SHG during later time points of the infection indicating the importance of BPI in controlling *Shigella*. Interestingly, once the actively dividing bacteria (STM) leaves the replicating niche (STM LLO, STM Δ sifA), they will interact with BPI. These results demonstrate an active role of BPI in eliminating Gram-negative bacterial pathogens inside macrophages. Whether BPI can cross talk with other signaling pathways and can induce additional bactericidal activity is not entirely clear. It will be interesting to analyze the contribution of macrophage-derived BPI in preventing various infectious diseases including parasitic and bacterial infections. Polymorphisms in BPI are associated with different inflammatory diseases, including Crohn's disease (CD) (23–25). Macrophages derived from patients with CD show impaired bacterial clearance (26, 27). Whether this impaired clearance of bacteria is due to polymorphisms of BPI in macrophages derived from CD patients is yet to be understood.

METHODS

Cell Culture

The human monocyte cell line U937 (NCCS, Pune) and murine macrophage cell line RAW 264.7 (kind gift from Prof. Anjali

Karandae, IISc) were maintained in RPMI (Sigma-Aldrich) containing 10% FBS (fetal bovine serum, Gibco). For induction of macrophage differentiation, cells were seeded and stimulated with 50-nM PMA (phorbol 12-myristate 13-acetate, Sigma-Aldrich) for 24 h. After PMA induction, non-attached cells were removed by gentle aspiration and attached cells were washed three times with RPMI containing 10% FBS.

Knockdown of BPI in Human Macrophages

In order to knock down BPI in human macrophages, BPI dsRNA was designed against three regions within the gene: (a) GGAGCTGAAGAGGATCAAGATTCCTGACTACTCAGACAGCTTTAAGATCAAGCATCTTGGGAAGGGGCA TTATAGCTTCTACAGCATGGACATCCGTGAATT CCAGCTTCCCAGTTCCCAGATAAGCATGGT GCCCAATGTGGGCCTTAAGTTCTCCATCAGC AACGCCAATATCAAGATC; (b) TGTCCACGTGC ACATCT CAAAGAG CAAAGTCGGGTGGCTGAT CCAACTCTTCCA CAAAAAATTGAGTCTGCGCTTCGAAAC AAGATGAACA GCCAGGTCTGCGAGAAAGTGACCAATTCTGTATC CTCCAAGCTGCAACCTTATTTCCAGACTCTGC; and (c) GGGTCTTGAAGATGACCCCTAGAGATGACATGA TTCCAAAGGAGTCCAAATTTTCGACTGACA ACCAAGTTCTTTGGAACCTTCTACCTGAGGTGGCC AAGAAGTTTCCCAACATGAAGATACAGATCCATGT CTCAGCCTCCACC. All dsRNA were obtained from chromous biotech. Transfection was done using oligofectamine as recommended by the manufacturer (Invitrogen, Life Technologies). Transfection was done for 24 h.

Cloning and Expression of Human BPI in Murine Macrophages

The complete BPI coding sequence was amplified from cDNA derived from human monocytes. PCR product was gel eluted and was cloned in pcDNA 3.1 expression vector (kind gift from Dr. G. Subbha Rao, IISc). The pcDNA 3.1 hBPI was transiently transfected to RAW macrophages using PEI transfection reagent (Sigma-Aldrich). Twenty-four hours post-transfection, cells were harvested and BPI expression was quantified by western blotting. Transiently transfected cells were used for infection assay. pcDNA 3.1 was used as empty vector control in all the experiments. The sequence for cloning primers is as follows: hBPI forward primer, 5' AAGGATCCA TGAGAGAGAACATGGCC3', and hBPI reverse primer, 5' GGCAAGCTTT CATTTATAGACAACGTC3'. The restriction sites within the primers are underlined.

Bacterial Strains and Growth Conditions

Salmonella Typhimurium (*Salmonella* enterica serovar Typhimurium ATCC14028s), STM Δ fliC (flagellin-deficient STM 14028), STY (STY ATCC CT18), SHG (SHG clinical isolate 1), SA (SA ATCC 25923), and *E. coli* (*E. coli* DH5 α ATCC) were grown in Luria-Bertani medium at 37°C. STM Δ sifA and STM Δ fliC was a gift from Michael Hensel, Universität Osnabrück, Germany. For immunostaining experiment STM, SHG and *E.*

coli were transformed with the pFPV25.1 plasmid containing the GFPmut3 gene (Addgene). In order to express listeriolysin O in STM, STM 14028 was transformed with pPROEX HT-b LLO (STM LLO). STM 14028 transformed with pPROEX HT-b (STM EV) was used as empty vector control. pPROEX HT-b LLO was a kind gift from Prof Sandhya S. Visweswariah, IISC. All the transformants were maintained in Luria-Bertani medium containing 100 µg/ml of ampicillin. For induction of LLO expression in STM LLO, log phase cells were treated with 500-nM IPTG (Sigma-Aldrich) for 6 h. STY Vi⁻ (STY lacking Vi polysaccharide) was a kind gift from Prof Ayub Qadri, NII.

Immunofluorescence Microscopy

U937 cells were seeded on glass coverslips overnight before infection or treatment. After treatment, cells were washed with PBS and fixed with 3.5% paraformaldehyde for 15 min. Cells were permeabilized using 1% saponin (Sigma-Aldrich) dissolved in PBS with 3% BSA. Immunostaining was done using anti-BPI antibody (Sigma-Aldrich) followed by anti-rabbit Alexa 647 antibody (DSHB, University of Iowa). To visualize macrophage population in human PBMC cultures, cells were stained with the anti-CD11b antibody (DSHB, University of Iowa) followed by anti-mouse Alexa 488 antibody (DSHB, University of Iowa). To label lysosomes, anti-LAMP2 antibody (DSHB, University of Iowa) was used followed by anti-mouse Alexa 488 antibody (DSHB, University of Iowa). Cells were counterstained using DAPI (Sigma-Aldrich) to label the nucleus. To visualize bacteria, either pFPV25.1 GFPmut3-transformed bacteria were used or bacteria were visualized using DAPI staining. Image acquisition was done with a Zeiss confocal microscope (LSM Meta 710). Quantitation of images was done as explained by Billings et al. (22). Briefly, for quantification of BPI recruitment and interaction with bacteria, an ROI was drawn around each bacterium based upon GFP signal (STM-GFP, SHG-GFP, and *E. coli* GFP) or DAPI staining (STM Δ sifA, STM LLO, and STM EV). Images were analyzed using Zen Blue edition software provided by Zeiss. The colocalization coefficient values at ROI were obtained using Zen Blue edition software and were multiplied by 100 to get the percent colocalization and plotted. The MFI of BPI at ROI was calculated Zen software and plotted.

Bacterial Phagocytic Uptake and Proliferation Assay

Bacteria were grown in Luria-Bertani medium, and overnight culture was used to infect U937 cells at a ratio of 10 bacteria per cell (MOI 10). Extracellular bacteria were removed 30 min post-infection, and cells were maintained in 100 µg/ml gentamycin for 1 h to kill any extracellular bacteria. Infected cells were maintained in DMEM containing 10 µg/ml of gentamycin. Phagocytosis of bacteria by macrophages was calculated by plating the macrophage cell lysates 30 min post-infection. For calculating the percentage phagocytosis, CFU was normalized with respect to untransfected control. The obtained value was multiplied by 100 to get percentage phagocytosis.

$$\% \text{Phagocytosis} = (\text{CFU of Test} / \text{CFU of Untransfected control}) \times 100$$

Bacterial replication inside macrophages was quantified by plating the cell lysates 2- and 18-h post-infection. Fold proliferation was calculated by normalizing the bacterial CFU at 18 h with respect to 2 h. For experiments using SHG, the bacterial survival was calculated instead of fold replication because SHG cannot proliferate inside macrophages.

$$\text{Fold proliferation} = \text{CFU of bacteria at 18h} / \text{CFU of bacteria at 2h}$$

$$\% \text{Survival (SHG)} = (\text{CFU of SHG at 2h} / \text{CFU of SHG at 18h}) \times 100$$

Infection and Stimulation of Human Macrophages

For stimulation with various inflammatory mediators, PMA-treated U937 cells were incubated with STM LPS 100 ng/mL (Sigma-Aldrich) and STM flagellin 500 ng/mL in DMEM containing 10% FBS for the indicated time periods. Flagellin was isolated from STM as previously described in details (28). U937 monocyte cell line was used as a control throughout the experiment.

Human PBMC Isolation

This study was approved by Institutional Biosafety guidelines (IBSC) at Indian Institute of Science, Bangalore, India (Ref No: IBSC/IISC/DC/04/2015), and written informed consent was obtained from all participants before participation. All the procedures were carried out by trained medical technician. Human PBMCs were isolated from healthy individuals using Himedia LSM as per instructors manual. Briefly, blood collected from healthy individuals were overlaid on LSM and separated into different layers using a low-speed centrifugation. The cell layer containing human PBMCs was collected after centrifugation and mixed with DMEM without serum. Cells were seeded into a six-well plate, and unattached cells were gently aspirated. Attached cells were washed three times with DMEM containing 10% FBS and maintained in the same for 24 h.

PCR and Real-time Analysis of BPI mRNA Expression

Total RNA was extracted from 1×10^6 cells using TRIzol reagent (Invitrogen) as per the manufacturer's protocol. After DNase treatment, 2 µg of RNA was used for cDNA synthesis using tetra-reverse transcriptase (Bioline). qRT-PCR was performed using the Kapa SYBR Green RT-PCR kit (Kapa Biosystems) as per the manufacturer's protocol in an Applied Biosystems® ViiA™ 7 Real-Time PCR instrument. The following primers were used for detecting BPI level by real-time PCR: hBPI forward primer, 5'ATGAACAGCCAGGTCT 3', and hBPI

reverse primer, 5'GGTCATTACTGGCAG 3'. Expression was normalized to the housekeeping gene beta-actin. Following primers were used for detecting actin level: actin forward primer, 5'GGTGGCTTTTAGGATGGCAAG3', and actin reverse primer, 5'ACTGGAACGGTGAAGGTGACAG3'. Expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method.

Western Blotting and FACS to Quantify Protein Levels

For quantifying BPI expression by western blot, 10^6 cells were grown on six-well plates and exposed to various conditions as mentioned. Cell lysates were prepared and proteins were resolved by 10% SDS-PAGE and transferred to PVDF membrane. The blots were incubated with an anti-BPI antibody (Sigma-Aldrich) followed by anti-rabbit HRP (DSHB, University of Iowa). Immunoblots were visualized by ECL reagent. Densitometric quantification of blots was done using the Multi Gauge software (FUJIFILM).

For quantification of protein expression by flow cytometry, cells were fixed using 3.5% paraformaldehyde for 15 min. Cells were permeabilized using 0.1% saponin dissolved in PBS with 3% BSA. Immunostaining was done using anti-BPI antibody (Sigma-Aldrich) followed by anti-rabbit Alexa 647 antibody (DSHB, University of Iowa). Cells were subjected to flow cytometric analysis (BD FACSCalibur™). Data were analyzed using BD FACSDIVA™ software.

REFERENCES

- Balakrishnan A, Marathe SA, Joglekar M, Chakravorty D. Bactericidal/permeability increasing protein: a multifaceted protein with functions beyond LPS neutralization. *Innate Immun* (2013) 19:339–47. doi:10.1177/1753425912465098
- Eckert M, Wittmann I, Rollinghoff M, Gessner A, Schnare M. Endotoxin-induced expression of murine bactericidal permeability/increasing protein is mediated exclusively by toll/IL-1 receptor domain-containing adaptor inducing IFN-beta-dependent pathways. *J Immunol* (2006) 176:522–8. doi:10.4049/jimmunol.176.1.522
- Gantt KR, Goldman TL, McCormick ML, Miller MA, Jeronimo SM, Nascimento ET, et al. Oxidative responses of human and murine macrophages during phagocytosis of *Leishmania chagasi*. *J Immunol* (2001) 167:893–901. doi:10.4049/jimmunol.167.2.893
- Fang FC. Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. *Nat Rev Microbiol* (2004) 2:820–32. doi:10.1038/nrmicro1004
- Schneemann M, Schoeden G. Macrophage biology and immunology: man is not a mouse. *J Leukoc Biol* (2007) 81:579; discussion 580. doi:10.1189/jlb.1106702
- Weiss J, Victor M, Stendhal O, Elsbach P. Killing of Gram-negative bacteria by polymorphonuclear leukocytes: role of an O_2 -independent bactericidal system. *J Clin Invest* (1982) 69:959–70. doi:10.1172/JCI110535
- Weinrauch Y, Foreman A, Shu C, Zarembek K, Levy O, Elsbach P, et al. Extracellular accumulation of potentially microbicidal bactericidal/permeability-increasing protein and p15s in an evolving sterile rabbit peritoneal inflammatory exudate. *J Clin Invest* (1995) 95:1916–24. doi:10.1172/JCI117873
- Takashiba S, Van Dyke TE, Amar S, Murayama Y, Soskolne AW, Shapira L. Differentiation of monocytes to macrophages primes cells for lipopolysaccharide stimulation via accumulation of cytoplasmic nuclear factor kappaB. *Infect Immun* (1999) 67:5573–8.

Statistical Analysis

The data were subjected to statistical analysis by applying Student's *t*-test by using Graph Pad prism 4 software.

AUTHOR CONTRIBUTIONS

AB and DC conceived the study; AB performed the experiments; and AB, DC, and MS analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

We thank the Confocal facility (Divisional and Departmental), IISc for the help. This work was supported by DAE SRC outstanding award (DAE0195) and DBT-IISc partnership program for advanced research in biological sciences and bioengineering to DC as well as Deutsche Forschungsgemeinschaft to MS (SCHN 635/4-1). Infrastructure support from ICMR (Center for Advanced Study in Molecular Medicine), DST (FIST), and UGC (special assistance) is acknowledged. We thank Lakshmi Menon for editing the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <http://journal.frontiersin.org/article/10.3389/fimmu.2016.00455>

- Mannion BA, Weiss J, Elsbach P. Separation of sublethal and lethal effects of the bactericidal/permeability increasing protein on *Escherichia coli*. *J Clin Invest* (1990) 85:853–60. doi:10.1172/JCI114512
- Marra MN, Wilde CG, Griffith JE, Snable JL, Scott RW. Bactericidal/permeability-increasing protein has endotoxin-neutralizing activity. *J Immunol* (1990) 144:662–6.
- Iovine NM, Elsbach P, Weiss J. An opsonic function of the neutrophil bactericidal/permeability-increasing protein depends on both its N- and C-terminal domains. *Proc Natl Acad Sci U S A* (1997) 94:10973–8. doi:10.1073/pnas.94.20.10973
- Wilson RP, Winter SE, Spees AM, Winter MG, Nishimori JH, Sanchez JF, et al. The Vi capsular polysaccharide prevents complement receptor 3-mediated clearance of *Salmonella enterica* serotype Typhi. *Infect Immun* (2011) 79:830–7. doi:10.1128/IAI.00961-10
- Wilson RP, Raffatellu M, Chessa D, Winter SE, Tukel C, Baumler AJ. The Vi-capsule prevents toll-like receptor 4 recognition of *Salmonella*. *Cell Microbiol* (2008) 10:876–90. doi:10.1111/j.1462-5822.2007.01090.x
- Tobias PS, Soldau K, Iovine NM, Elsbach P, Weiss J. Lipopolysaccharide (LPS)-binding proteins BPI and LBP form different types of complexes with LPS. *J Biol Chem* (1997) 272:18682–5. doi:10.1074/jbc.272.30.18682
- Merighi M, Ellermeier CD, Schlauch JM, Gunn JS. Resolvase-in vivo expression technology analysis of the *Salmonella enterica* serovar Typhimurium PhoP and PmrA regulons in BALB/c mice. *J Bacteriol* (2005) 187:7407–16. doi:10.1128/JB.187.21.7407-7416.2005
- Steele-Mortimer O. The *Salmonella*-containing vacuole: moving with the times. *Curr Opin Microbiol* (2008) 11:38–45. doi:10.1016/j.mib.2008.01.002
- Beuzon CR, Meresse S, Unsworth KE, Ruiz-Albert J, Garvis S, Waterman SR, et al. *Salmonella* maintains the integrity of its intracellular vacuole through the action of SifA. *EMBO J* (2000) 19:3235–49. doi:10.1093/emboj/19.13.3235
- Gentschev I, Sokolovic Z, Mollenkopf HJ, Hess J, Kaufmann SH, Kuhn M, et al. *Salmonella* strain secreting active listeriolysin changes its intracellular localization. *Infect Immun* (1995) 63:4202–5.

19. Dentener MA, Francot GJ, Buurman WA. Bactericidal/permeability-increasing protein, a lipopolysaccharide-specific protein on the surface of human peripheral blood monocytes. *J Infect Dis* (1996) 173:252–5. doi:10.1093/infdis/173.1.252
20. Mestas J, Hughes CC. Of mice and not men: differences between mouse and human immunology. *J Immunol* (2004) 172:2731–8. doi:10.4049/jimmunol.172.5.2731
21. Sun D, Ding A. MyD88-mediated stabilization of interferon-gamma-induced cytokine and chemokine mRNA. *Nat Immunol* (2006) 7:375–81. doi:10.1038/ni1308
22. Billings EA, Lee CS, Owen KA, D'Souza RS, Ravichandran KS, Casanova JE. The adhesion GPCR BAI1 mediates macrophage ROS production and microbicidal activity against Gram-negative bacteria. *Sci Signal* (2016) 9:ra14. doi:10.1126/scisignal.aac6250
23. Stoffel MP, Csernok E, Herzberg C, Johnson T, Carroll SF, Gross WL. Anti-neutrophil cytoplasmic antibodies (ANCA) directed against bactericidal/permeability increasing protein (BPI): a new seromarker for inflammatory bowel disease and associated disorders. *Clin Exp Immunol* (1996) 104:54–9. doi:10.1046/j.1365-2249.1996.d01-654.x
24. Walmsley RS, Zhao MH, Hamilton MI, Brownlee A, Chapman P, Pounder RE, et al. Antineutrophil cytoplasm autoantibodies against bactericidal/permeability-increasing protein in inflammatory bowel disease. *Gut* (1997) 40:105–9. doi:10.1136/gut.40.1.105
25. Akin H, Tahan G, Ture F, Eren F, Atug O, Tahan V, et al. Association between bactericidal/permeability increasing protein (BPI) gene polymorphism (Lys216Glu) and inflammatory bowel disease. *J Crohns Colitis* (2011) 5:14–8. doi:10.1016/j.crohns.2010.08.008
26. Casanova JL, Abel L. Revisiting Crohn's disease as a primary immunodeficiency of macrophages. *J Exp Med* (2009) 206:1839–43. doi:10.1084/jem.20091683
27. Smith AM, Rahman FZ, Hayee B, Graham SJ, Marks DJ, Sewell GW, et al. Disordered macrophage cytokine secretion underlies impaired acute inflammation and bacterial clearance in Crohn's disease. *J Exp Med* (2009) 206:1883–97. doi:10.1084/jem.20091233
28. Marathe SA, Balakrishnan A, Negi VD, Sakorey D, Chandra N, Chakravorty D. Curcumin reduces the motility of *Salmonella enterica* serovar Typhimurium by binding to the flagella thereby leading to flagellar fragility and shedding. *J Bacteriol* (2016) 198(13):1798–811. doi:10.1128/JB.00092-16

Conflict of Interest Statement: 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.

Copyright © 2016 Balakrishnan, Schnare and Chakravorty. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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: info@frontiersin.org | +41 21 510 17 00



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