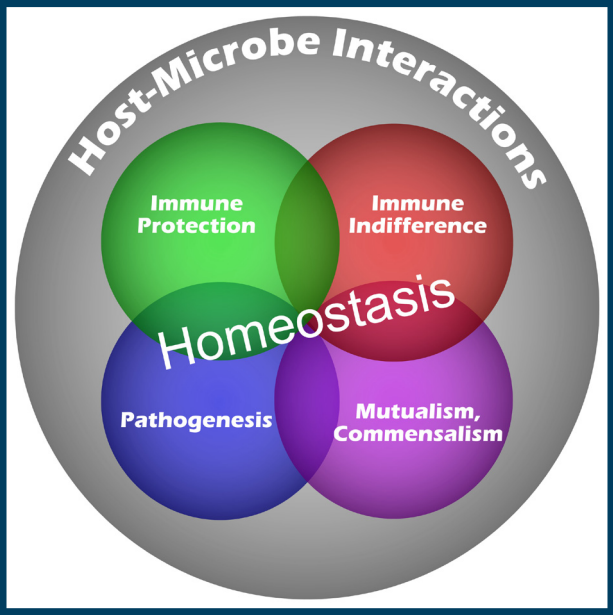
frontiers

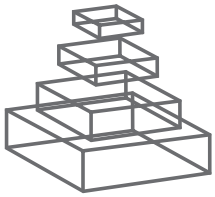
RESEARCH TOPICS



CHANGING VIEWS OF THE EVOLUTION OF IMMUNITY

Topic Editors
Gary W. Litman and Larry J. Dishaw

frontiers in IMMUNOLOGY



frontiers

ISSN 1664-8714
ISBN 978-2-88919-171-0
DOI 10.3389/978-2-88919-171-0

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

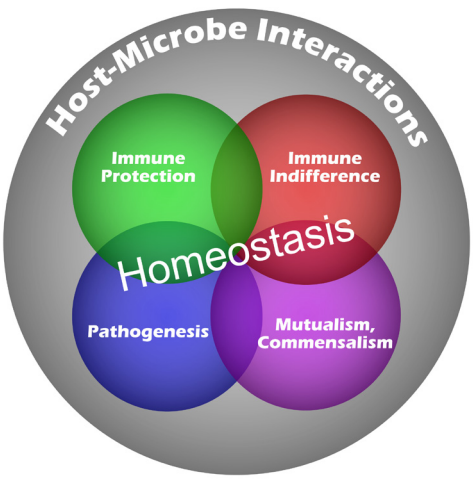
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - 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

CHANGING VIEWS OF THE EVOLUTION OF IMMUNITY

Topic Editors:
Gary W. Litman, University of South Florida, USA
Larry J. Dishaw, University of South Florida, USA



Evolving views of the evolution of immunity now incorporate a more diverse consideration of host interactions with microbiota. Animals coexist with a vast array of microbial species and the general view that the immune system evolved only to defend against pathogenic infection now appears to be limited. New evidence suggests that selective forces emerging from intimate associations with microbes, most of which are beneficial (mutualism) or benign (commensalism), and not pathogenic, may play a prominent role in the evolution of immune systems. The host must balance immune protection from and indifference to various symbiotic interactions to achieve homeostasis. These general forces have worked to shape the numerous, divergent immune systems that accommodate taxon-specific needs throughout animal phylogeny.
Image credit: Drs. J. P. Cannon, L. J. Dishaw, and G.W.Litman, USF

The multitude of cells, signaling pathways, receptors, novel genetic recombination mechanisms and interactive pathways of receptor function and cell differentiation that constitute the vertebrate adaptive immune system are integrally linked with the multicomponent innate immune system. At first glance, the levels of complexity seen in both systems at the phylogenetic level of mammals present what seem to be insurmountable hurdles in terms of achieving a systematic understanding of the evolution of immunity. New research directions and approaches suggest that resolution of many long-standing questions in this area is now possible.

Historically, immunologists considered lower vertebrates and invertebrates as "simpler" forms, i.e., they were expected to possess more basic (less layered) levels of immunological complexity and thus potentially would serve as important resources. By considering the systematic placement of

representative species in the context of phylogeny, characterizing their immune receptors, co-receptors as well as accessory molecules and evaluating responses to immunologic stimuli, it was thought that a clearer picture of immune evolution would emerge. There is no doubt that this approach has achieved some notable successes but for the most part it has fallen short in terms of achieving a broad understanding of the immunologic needs of many relevant models and how adaptive change in immune function is effected.

Even if a structurally relevant ortholog of an immune effector is identified in a model organism, there is no reason to assume that it functions in a corresponding manner in disparate phylogenetic taxa. For example, survival of a sessile marine invertebrate, whose anatomical form puts it in open and contiguous contact with a literal sea of microorganisms and viruses, would be thought to depend, at least in part, on a "capable" immune response; however, at present, we have no real understanding of how this is achieved in an integrated manner. Furthermore, questions arise as to whether or not phenomena that are considered integral components of vertebrate-type immunity such as memory, tolerance, somatic change and clonal selection exist in invertebrates and if their functions parallel those recognized in mammals. More often than not, our interpretations are guided by preconceived notions that are based on observations made in distant species that often do not apply to far- removed taxa. We anticipate that major advances in our understanding of this broad subject are now forthcoming as resources exist or are being developed for examining important model organisms in their natural environments instead of within the confines of in vitro systems of potentially remote physiological significance.

Taking a wide range of hypotheses, observations and interpretations into account, in this special topic, contributors have developed a comprehensive overview emphasizing new directions and interpretations for understanding basic aspects of immunity that consider unique features inherent to various model systems, their life histories and habitats. Approaches applied with key model organisms maintained and confronted with relevant challenges under natural conditions are emphasized. Current concepts of self and nonself are addressed not only in terms of immunity but also reproductive fitness. How genetic variation in immune effector molecules is achieved and maintained in natural populations is examined; particular attention is directed to response interfaces that factor in symbiotic interactions. Gene expansion and mechanisms of genetic diversification are explored. How diverse molecules and a variety of effector cells contribute to our broad understanding of the evolution of a remarkably complex, integrated system and how this work is facilitating our understanding of mammalian immunity is addressed.

Table of Contents

frontiers in
IMMUNOLOGY

EDITORIAL
published: 21 May 2013
doi: 10.3389/fimmu.2013.00122

Changing views of the evolution of immunity

Larry J. Dishaw * and Gary W. Litman

Division of Molecular Genetics, Department of Pediatrics, University of South Florida, Tampa, FL, USA
**Correspondence: ldishaw@health.usf.edu*

Edited by:
Claudia Kemper, King's College London, UK

Reviewed by:
Claudia Kemper, King's College London, UK

Views on the evolution of immunity have been redefined as studies in protostomes, invertebrate deuterostomes, and various vertebrates have elucidated molecular details of host defense (Litman and Cooper, 2007; Boehm, 2012). Diverse species possess vast repertoires of immune related defenses, which have evolved into sophisticated, integrated networks (Rast and Messier-Solek, 2008; Messier-Solek et al., 2010). Certain components of immune defense represent either homologous structures or reflect novel approaches to confronting pathogens and other environmental influences [e.g., VLRs in agnathans (Boehm et al., 2012)]. Disparate diversification mechanisms and exceptional sophistication in immune mediators in some species blur distinctions between innate and adaptive immunity, the latter of which is viewed traditionally as a vertebrate adaptation (Litman et al., 2005; Messier-Solek et al., 2010; Boehm, 2011) associated with extensive somatic diversity, antigen-specific affinity maturation, and memory (Flajnik and Kasahara, 2010; Boehm, 2012). Innate immunity, specifically the recognition of microbe-associated molecular patterns by pattern recognition receptors (PRRs), has taken center stage owing to its capacity to shape adaptive immunity (Schenten and Medzhitov, 2011). PRRs [e.g., toll-like receptors (TLRs)], also contribute significantly to immune homeostasis (Medzhitov, 2010; Carvalho et al., 2012). In this special issue we explore topics that are continuing to reshape our interpretations of immune evolution.

Historically, transplantation immunology framed our understanding of immunological recognition and the interplay between immunoglobulin domain-containing receptors, co-receptors, and the major histocompatibility complex (MHC) (Brent, 2003). These earlier concepts were extended to address graft rejection in jawless vertebrates as well as select invertebrates (Finstad and Good, 1964; Hildemann and Thoenes, 1969; Mayer et al., 2002; Little et al., 2005; Kvell et al., 2007). Today, various models of allorecognition are recognized, some of which are restricted to certain phyla (Buss, 1987), and can be traced to the ancestors of sessile invertebrates (Dishaw and Litman, 2009). Broad rules govern discrimination between conspecifics (Rosengarten and Nicotra, 2011). Nydam and De Tomaso (2011) update our understanding of the evolution of allorecognition, emphasizing commonality in the systems that generate polymorphisms, and discuss how genetic diversity is maintained.

Extensive variation in immune genes traditionally has been equated with the immunoglobulin and T cell receptor gene loci in B and T lymphocytes, respectively, as well as in some MHC loci (Hughes, 2002). Recent studies in some invertebrate deuterostomes provide evidence for expansion and germline diversification of immune receptor repertoires. Buckley and Rast (2012) demonstrate lineage-specific properties among expanded sea urchin TLRs. Their findings indicate that: (1) some antigen binding sites may be co-evolving with variable ligands, (2) TLR subfamilies are utilized differently between larval and adult coelomocytes, and (3) sea urchin TLRs most likely represent immune surveillance molecules. Satake and Sekiguchi (2012) review the evolution and functional diversification of TLRs among deuterostomes, highlighting a reduced repertoire in the tunicate, *Ciona intestinalis*. Only two TLRs can be detected in this species, with presumed hybrid functionality *in vitro* (Sasaki et al., 2009). Interestingly, neither TLR1 nor TLR2 recognizes bacterial lipopolysaccharide (LPS), suggesting that *Ciona* utilizes other mechanisms to detect LPS or that an accessory molecule(s) is involved.

Drosophila melanogaster (fruit fly) uses complex alternative RNA splicing to diversify the Down's syndrome cell adhesion molecule (DSCAM), a multiexonic receptor implicated in neuronal patterning (Shi and Lee, 2012). Some DSCAM isoforms serve as PRRs in peripheral hemocytes and exhibit increased specificity for distinct targets (Watson et al., 2005; Brites et al., 2008; Chou et al., 2009). These findings are reminiscent of the fibrinogen-related proteins (FREPs) (Adema et al., 1997; Zhang et al., 2004), which consist of fibrinogen and immunoglobulin superfamily-related domains that can undergo somatic mutation and gene conversion. Individual somatic lineages expressing FREPs respond to specific parasite burdens (Mone et al., 2010). Smith (2012) reviews Sp185/333 genes, a large family of innate receptors in sea urchin expressed in hemocytes. Variation in genes encoding Sp185/333 receptors arises via complex DNA rearrangements and may be influenced by persistent antigenic sources (Buckley et al., 2008; Dheilly et al., 2009).

Not all immune receptors are restricted to foreign determinants (Rabinovich and Croci, 2012). Some glycans can be found on both host and microbial surfaces (Davicino et al., 2011). Vasta et al. (2012) describe an apparent paradox among galectins, which until recently were considered essential in self-recognition (Rabinovich and Croci, 2012). Galectins now are considered PRRs that recognize related glycans on microbes (Sato et al., 2009). PRRs are thought to interact only with microbial products (Kawai and Akira, 2010); some, such as galectins, also may possess discriminatory properties (van Vliet et al., 2008). Galectin self-recognition may require interaction with accessory molecules on self-cells and warrants further investigation.

The role of PRRs in symbiotic relationships likely is ancient (Bosch, 2012), involving complex host-microbial interactions at the surface of mucosal tissues (Duerkop et al., 2009; Round

et al., 2011; Wells et al., 2011; Hill et al., 2012). Collins et al. (2012) describe a PRR that may govern such interactions between the bobtail squid and *Vibrio fisheri*, a bacterial symbiont of the light organ. Immune systems appear to have evolved mechanisms that discriminate among symbionts and pathogens, while promoting the former (Speckman et al., 2003; Lee and Mazmanian, 2010; Nyholm and Graf, 2012).

There has been a tendency to oversimplify or even ignore the broader roles of PRRs in host physiology. Arrieta and Finlay (2012) review the complex strategies that are used by gut bacteria to modulate immune homeostasis. The complex roles of adaptive immunity among vertebrates further complicates the roles of PRRs in homeostasis (Lee and Mazmanian, 2010; Hooper et al., 2012). Dishaw et al. (2012) argue that *Ciona intestinalis*, a protochordate, can help define host and microbe interactions at mucosal surfaces. Presumably, rules and relationships that govern homeostasis in this system may help reveal how perturbations can lead to a broad range of intestinal pathologies in higher vertebrates.

Specific molecules have been implicated in intestinal homeostasis and include alkaline phosphatase-intestinal (Alpi), a member of the alkaline phosphatase (Alp) family. One possible role for these molecules is the detoxification of LPS, which in turn minimizes innate responses to commensal or beneficial microbial communities (Beumer et al., 2003; Bates et al., 2007; Lalles, 2010). Yang et al. (2012) describe the complex evolutionary patterns of *Alpi* genes, which appear to be evolving independently in vertebrate, non-vertebrate and insect lineages. All four zebrafish *Alp* genes are shown to be expressed in the intestine, where *alp3* is expressed exclusively. The authors propose that intestinal expression of *Alp* may be an ancestral trait as alkaline-phosphatase-mediated LPS detoxification likely is central to the stability of gut microbe and host interactions.

Phylogenetic considerations, including the use of non-traditional models, have been instrumental in forging new thinking among immunologists (Loker et al., 2004). It is becoming increasingly clear that the immune system may have evolved, not only to recognize potential pathogens but also to help sustain and stabilize beneficial associations at the surface of mucosal tissues. Loker (2012) considers symbiosis as a driver of evolutionary novelty on both sides of the host-parasite struggle. In this broad, topical overview, host immunity is a pervasive requirement and the immune evolutionary process is seen to be influenced by conflict with parasites and/or the need to cooperate with symbionts. The work presented in this series already is proving critical in terms of broadening our view of immune complexity and the multifaceted role of the host-microbe dialog in maintaining homeostasis. Critical departures from our traditional views of immune defense are being revealed in detailed studies of alternative model systems and in turn are reshaping our understanding of immunity in conventional systems.

REFERENCES

Adema, C. M., Hertel, L. A., Miller, R. D., and Loker, E. S. (1997). A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8691–8696.

Arrieta, M. C., and Finlay, B. B. (2012). The commensal microbiota drives immune homeostasis. *Front. Immunol.* 3:33. doi: 10.3389/fimmu.2012.00033

Bates, J. M., Akerlund, J., Mittge, E., and Guillemin, K. (2007). Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota. *Cell Host Microbe* 2, 371–382.

Beumer, C., Wulferink, M., Raaben, W., Fiechter, D., Brands, R., and Seinen, W. (2003). Calf intestinal alkaline phosphatase, a novel therapeutic drug for lipopolysaccharide (LPS)-mediated diseases, attenuates LPS toxicity in mice and piglets. *J. Pharmacol. Exp. Ther.* 307, 737–744.

Boehm, T. (2011). Design principles of adaptive immune systems. *Nat. Rev. Immunol.* 11, 307–317.

Boehm, T. (2012). Evolution of vertebrate immunity. *Curr. Biol.* 22, R722–R732.

Boehm, T., McCurley, N., Sutoh, Y., Schorpp, M., Kasahara, M., and Cooper, M. D. (2012). VLR-based adaptive immunity. *Annu. Rev. Immunol.* 30, 203–220.

Bosch, T. C. (2012). Understanding complex host-microbe interactions in Hydra. *Gut Microbes* 3, 345–351.

Brent, L. (2003). The 50th anniversary of the discovery of immunologic tolerance. *N. Engl. J. Med.* 349, 1381–1383.

Brites, D., McTaggart, S., Morris, K., Anderson, J., Thomas, K., Colson, I., et al. (2008). The DSCAM homologue of the crustacean Daphnia is diversified by alternative splicing like in insects. *Mol. Biol. Evol.* 25, 1429–1439.

Buckley, K. M., Munshaw, S., Kepler, T. B., and Smith, L. C. (2008). The 185/333 gene family is a rapidly diversifying host-defense gene cluster in the purple sea urchin *Strongylocentrotus purpuratus*. *J. Mol. Biol.* 379, 912–928.

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

Buss, L. W. (1987). *The Evolution of Individuality*. Princeton: Princeton University Press.

Carvalho, F. A., Aitken, J. D., Vijay-Kumar, M., and Gewirtz, A. T. (2012). Toll-like receptor-gut microbiota interactions: perturb at your own risk! *Annu. Rev. Physiol.* 74, 177–198.

Chou, P. H., Chang, H. S., Chen, I. T., Lin, H. Y., Chen, Y. M., Yang, H. L., et al. (2009). 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.* 33, 1258–1267.

Collins, A. J., Schleicher, T. R., Rader, B. A., and Nyholm, S. V. (2012). Understanding the role of host hemocytes in a squid/vibrio symbiosis using transcriptomics and proteomics. *Front. Immunol.* 3:91. doi: 10.3389/fimmu.2012.00091

Davicino, R. C., Elicabe, R. J., Di Genaro, M. S., and Rabinovich, G. A. (2011). Coupling pathogen recognition to innate immunity through glycan-dependent mechanisms. *Int. Immunopharmacol.* 11, 1457–1463.

Dheilly, N. M., Nair, S. V., Smith, L. C., and Raftos, D. A. (2009). Highly variable immune-response proteins (185/333) from the sea urchin, *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. *J. Immunol.* 182, 2203–2212.

Dishaw, L. J., Flores-Torres, J. A., Mueller, M. G., Karrer, C. R., Skapura, D. P., Melillo, D., et al. (2012). A basal chordate model for studies of gut microbial immune interactions. *Front. Immunol.* 3:96. doi: 10.3389/fimmu.2012.00096

Dishaw, L. J., and Litman, G. W. (2009). Invertebrate allorecognition: the origins of histocompatibility. *Curr. Biol.* 19, R286–R288.

Duerkop, B. A., Vaishnava, S., and Hooper, L. V. (2009). Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* 31, 368–376.

Finstad, J., and Good, R. A. (1964). The evolution of the immune response. III. Immunologic responses in the lamprey. *J. Exp. Med.* 120, 1151–1168.

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

Hildemann, W. H., and Thoenes, G. H. (1969). Immunological responses of Pacific hagfish. I. Skin transplantation immunity. *Transplantation* 7, 506–521.

Hill, D. A., Siracusa, M. C., Abt, M. C., Kim, B. S., Kobuley, D., Kubo, M., et al. (2012). Commensal bacteria-derived signals regulate basophil hematopoiesis and allergic inflammation. *Nat. Med.* 18, 538–546.

Hooper, L. V., Littman, D. R., and MacPherson, A. J. (2012). Interactions between the microbiota and the immune system. *Science* 336, 1268–1273.

Hughes, A. L. (2002). Natural selection and the diversification of vertebrate immune effectors. *Immunol. Rev.* 190, 161–168.

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

Kvell, K., Cooper, E. L., Engelmann, P., Bovari, J., and Nemeth, P. (2007). Blurring borders: innate immunity with adaptive features. *Clin. Dev. Immunol.* 2007, 83671.

Lalles, J. P. (2010). Intestinal alkaline phosphatase: multiple biological roles in maintenance of intestinal homeostasis and modulation by diet. *Nutr. Rev.* 68, 323–332.

Lee, Y. K., and Mazmanian, S. K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* 330, 1768–1773.

Litman, G. W., Cannon, J. P., and Dishaw, L. J. (2005). Reconstructing immune phylogeny: new perspectives. *Nat. Rev. Immunol.* 5, 866–879.

Litman, G. W., and Cooper, M. D. (2007). Commentary: why study the evolution of immunity? *Nat. Immunol.* 8, 547–548.

Little, T. J., Hultmark, D., and Read, A. F. (2005). Invertebrate immunity and the limits of mechanistic immunology. *Nat. Immunol.* 6, 651–654.

Loker, E. S. (2012). Macroevolutionary Immunology: a role for immunity in the diversification of animal life. *Front. Immunol.* 3:25. doi: 10.3389/fimmu.2012.00025

Loker, E. S., Adema, C. M., Zhang, S.-M., and Kepler, T. B. (2004). Invertebrate immune systems – not homogeneous, not simple, not well understood. *Immunol. Rev.* 198, 10–24.

Mayer, W. E., Uinuk-Ool, T., Tichy, H., Gartland, L. A., Klein, J., and Cooper, M. D. (2002). Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14350–14355.

Medzhitov, R. (2010). Innate immunity: quo vadis? *Nat. Immunol.* 11, 551–553.

Messier-Solek, C., Buckley, K. M., and Rast, J. P. (2010). Highly diversified innate receptor systems and new forms of animal immunity. *Semin. Immunol.* 22, 39–47.

Mone, Y., Gourbal, B., Duval, D., Du, P. L., Kieffer-Jaquinod, S., and Mitta, G. (2010). 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.* 4:e813. doi: 10.1371/journal.pntd.0000813

Nydam, M. L., and De Tomaso, A. W. (2011). Creation and maintenance of variation in allorecognition Loci: molecular analysis in various model systems. *Front. Immunol.* 2:79. doi: 10.3389/fimmu.2011.00079

Nyholm, S. V., and Graf, J. (2012). Knowing your friends: invertebrate innate immunity fosters beneficial bacterial symbioses. *Nat. Rev. Microbiol.* 10, 815–827.

Rabinovich, G. A., and Croci, D. O. (2012). Regulatory circuits mediated by lectin-glycan interactions in autoimmunity and cancer. *Immunity* 36, 322–335.

Rast, J. P., and Messier-Solek, C. (2008). Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol. Bull.* 214, 274–283.

Rosengarten, R. D., and Nicotra, M. L. (2011). Model systems of invertebrate allorecognition. *Curr. Biol.* 21, R82–R92.

Round, J. L., Lee, S. M., Li, J., Tran, G., Jabri, B., Chatila, T. A., et al. (2011). The Toll-like receptor 2 pathway establishes colonization by a commensal of the human microbiota. *Science* 332, 974–977.

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

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

Sato, S., St-Pierre, C., Bhaumik, P., and Nieminen, J. (2009). Galectins in innate immunity: dual functions of host soluble beta-galactoside-binding lectins as damage-associated molecular patterns (DAMPs) and as receptors for pathogen-associated molecular patterns (PAMPs). *Immunol. Rev.* 230, 172–187.

Schenten, D., and Medzhitov, R. (2011). The control of adaptive immune responses by the innate immune system. *Adv. Immunol.* 109, 87–124.

Shi, L., and Lee, T. (2012). Molecular diversity of DSCAM and self-recognition. *Adv. Exp. Med. Biol.* 739, 262–275.

Smith, L. C. (2012). Innate immune complexity in the purple sea urchin: diversity of the sp185/333 system. *Front. Immunol.* 3:70. doi: 10.3389/fimmu.2012.00070

Speckman, R. A., Wright Daw, J. A., Helms, C., Duan, S., Cao, L., Taillon-Miller, P., et al. (2003). Novel immunoglobulin superfamily gene cluster, mapping to a region of human chromosome 17q25, linked to psoriasis susceptibility. *Hum. Genet.* 112, 34–41.

van Vliet, S. J., Garcia-Vallejo, J. J., and van Kooyk, Y. (2008). Dendritic cells and C-type lectin receptors: coupling innate to adaptive immune responses. *Immunol. Cell Biol.* 86, 580–587.

Vasta, G. R., Ahmed, H., Nita-Lazar, M., Banerjee, A., Pasek, M., Shridhar, S., et al. (2012). Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. *Front. Immunol.* 3:199. doi: 10.3389/fimmu.2012.00199

Watson, F. L., Püttmann-Holgado, R., Thomas, F., Lamar, D. L., Hughes, M., Kondo, M., et al. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 309, 1874–1878.

Wells, J. M., Rossi, O., Meijerink, M., and van, B. P. (2011). Epithelial crosstalk at the microbiota-mucosal interface. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4607–4614.

Yang, Y., Wandler, A. M., Postlethwait, J. H., and Guillemin, K. (2012). Dynamic evolution of the LPS-detoxifying enzyme intestinal alkaline phosphatase in zebrafish and other vertebrates. *Front. Immunol.* 3:314. doi: 10.3389/fimmu.2012.00314

Zhang, S.-M., Adema, C. M., Kepler, T. B., and Loker, E. S. (2004). Diversification of Ig superfamily genes in an invertebrate. *Science* 305, 251–254.

Received: 30 April 2013; accepted: 06 May 2013; published online: 21 May 2013.
*Citation: Dishaw LJ and Litman GW (2013) Changing views of the evolution of immunity. Front. Immunol. **4**:122. doi:10.3389/fimmu.2013.00122*
This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

REVIEW ARTICLE
published: 27 December 2011
doi: 10.3389/fimmu.2011.00079

Creation and maintenance of variation in allorecognition loci: molecular analysis in various model systems

Marie L. Nydam* and Anthony W. De Tomaso

Department of Molecular, Cellular and Developmental Biology, University of California Santa Barbara, Santa Barbara, CA, USA

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Min Wu, University of North Dakota, USA
Matthew Nicotra, University of Pittsburgh, USA

****Correspondence:***
Marie L. Nydam, Department of Molecular, Cellular and Developmental Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA.
e-mail: nydam@lifesci.ucsb.edu

Allorecognition is the ability of an organism to differentiate self or close relatives from unrelated conspecifics. Effective allorecognition systems are critical to the survival of organisms; they prevent inbreeding and facilitate fusions between close relatives. Where the loci governing allorecognition outcomes have been identified, the corresponding proteins often exhibit exceptional polymorphism. Two important questions about this polymorphism remain unresolved: how is it created, and how is it maintained. Because the genetic bases of several allorecognition systems have now been identified, including *alr1* and *alr2* in *Hydractinia*, fusion histocompatibility in *Botryllus*, the *het (vic)* loci in fungi, *tgrB1* and *tgrC1* in *Dictyostelium*, and self-incompatibility (SI) loci in several plant families, we are now poised to achieve a clearer understanding of how these loci evolve. In this review, we summarize what is currently known about the evolution of allorecognition loci, highlight open questions, and suggest future directions.

Keywords: allorecognition, mutation, recombination, variation, polymorphism, balancing selection

INTRODUCTION

Allorecognition, the ability of an organism to differentiate self or close relatives from unrelated conspecifics, occurs throughout the tree of life (Buss, 1982), in anemones (Mercier et al., 2011), angiosperms (Allen and Hiscock, 2008), ascidians (Raftos, 1994; Saito et al., 1994; Harada et al., 2008), bacteria (Gibbs et al., 2008), bryozoans (Hughes et al., 2004), cellular slime molds (Shaulsky and Kessin, 2007), corals (Hidaka et al., 1997), fungi (Glass et al., 2000), gymnosperms (Pandey, 1960; Runions and Owens, 1998), hydroids (Grosberg et al., 1996), plasmodial slime molds (Clark, 2003), and sponges (Fernandez-Busquets and Burger, 1999).

Effective allorecognition systems are critical to the survival of organisms: the SI loci prevent inbreeding depression in plants, and in many colonial organisms, fusing to a closely related individual can provide competitive and reproductive advantages where space is limited and reproductive output is based on the size of the organism (Buss, 1982).

Despite decades of research, the genetic basis of allorecognition remains hidden in many groups, including bryozoans, corals, and sponges. Many marine invertebrate species are difficult to culture and breed, which limits the crossing experiments necessary to pinpoint genomic locations involved in allorecognition (Grosberg and Plachetzki, 2010).

However, researchers have been studying the SI (self-incompatibility) loci in angiosperms for some time. And allorecognition genes have recently been identified in ascidians (Fusion Histocompatibility, FuHC in *Botryllus schlosseri* and s-themis/v-themis in *Ciona intestinalis*; De Tomaso et al., 2005; Harada et al., 2008, reviewed in Ben-Shlomo, 2008), bacteria (Gibbs et al., 2008), cellular slime molds [*tgrB1* and *tgrC1* in *Dictyostelium discoideum* (Shaulsky and Kessin, 2007)], hydroids [*alr1* and *alr2* in *Hydractinia symbiolongicarpus* (Nicotra et al., 2009)], and fungi [*het* or *vic* loci (Glass et al., 2000)]. In this review we will focus only on organisms where the genetic basis of allorecognition has been identified, and where the polymorphism in these loci has been studied. Although the loci governing self-incompatibility have recently been identified in the ascidian *C. intestinalis* (Harada et al., 2008) and the bacterium *Proteus mirabilis* (Gibbs et al., 2008), no evolutionary studies have yet been published.

In the systems where the loci governing allorecognition outcomes have been identified, the corresponding proteins have often exhibited exceptional polymorphism. In the clover *Trifolium pratense*, up to 193 S-alleles in the SI system were identified (Lawrence, 1996), and 13–16 S-alleles were identified from 20 *Arabidopsis lyrata* plants (Mable et al., 2003). In the colonial hydroid *H. symbiolongicarpus*, 35 alleles of the *alr2* allorecognition locus were sequenced from 18 colonies (Rosengarten et al., 2010).

Because allelic variation forms the basis of allorecognition, evolutionary studies of self/non-self recognition focus on this variation. If we can understand the evolutionary forces underlying the remarkable polymorphism in allorecognition loci, we gain valuable insights into the evolution and mechanisms of allorecognition systems. Two important questions about this polymorphism remain to be solved: how is it created and how is it maintained? We will address each question by summarizing and interpreting the available data.

CREATION OF POLYMORPHISM

Mutation and recombination are the two processes that commonly create variation in allorecognition loci. These forces have been examined in several systems: *het/vic* loci in fungi, FuHC in *B. schlosseri*, SI loci in the Brassicaceae (*Arabidopsis* and *Brassica*) and Solanaceae (*Lycium*, *Petunia*, *Physalis*, *Solanum*), and *alr2* in *Hydractinia*.

In the fungus *Podospora anserina*, the *het-d* and *het-e* loci belong to a 10-member gene family; diversity is created by sharing

of WD-repeats through recombinational processes within and between loci in this family. Because *het-d* and *het-e* have a large copy number of repeats, mutations arise frequently in the WD-domains, also creating diversity. So polymorphism is created by plentiful mutations, which are then exchanged within and between loci, creating even more polymorphism. An accelerated mutational process called repeat induced polymorphism (RIP) that targets repeat sequences in fungi is thought to further generate variation (Paoletti et al., 2007). Recombination does not often occur within the A-mating type locus of the basidiomycete fungus *Coprinus cinereus* (Day, 1963). But infrequent recombination events have created diversity in this locus; this is unusually in sex-determining loci (May and Matzke, 1995). Researchers found reduced recombination near the mating type locus (MAT) in the chestnut blight fungus, *Cryphonectria parasitica* (Kubisiak and Milgroom, 2006).

In *B. schlosseri,* FuHC experiences a substantial amount of intragenic recombination, based on three independent measures: R_m, the correlation between physical distance and three measures of linkage disequilibrium, and levels of recombination across the protein (Nydam, Taylor and De Tomaso unpublished data). Six populations were examined, 112 alleles and 77 individuals for Exons 1–14, 111 alleles and 76 individuals for Exons 18–31. This data set was used for all *Botryllus* analyses discussed in this paper. The relative contributions of mutation and recombination in generating polymorphism in FuHC were determined by calculating θ and R in DnaSP 5.10.01 (Librado and Rozas, 2009). $\theta = 4^*N^*\mu$, where N is the effective population size and μ is the mutation rate per DNA sequence per generation. $R = 4N^*r$ (Hudson, 1987), where N is the population size and r is the recombination rate per sequence. R is estimated from the variance of the average number of nucleotide differences between pairs of sequences (Hudson, 1987). All values were estimated using DnaSP. A ratio of $\theta/R = 1$ signifies an equal contribution of mutation and recombination, >1 a larger role for mutation, and <1 a larger role for recombination. θ/R was much less than one, so recombination clearly plays a larger role in the creation of FuHC than mutation.

When discussing the creation of variation in plant self-incompatibility (SI) loci, we must make a distinction between gametophytic and sporophytic SI systems. In gametophytic SI, the most common type of SI, the haploid self-incompatibility genotype of the pollen dictates its self-incompatibility phenotype. In sporophytic SI, the diploid self-incompatibility genotype of the plant dictates the self-incompatibility phenotype of the pollen produced by that plant (Newbigin et al., 1993).

Mutation plays a larger role than recombination in both gametophytic and sporophytic systems. Recombination would break up the association between pollen and pistil self-incompatibility loci, and thus is predicted to be suppressed around SI loci (Stein et al., 1991). Little evidence for recombination exists in gametophytic systems (Schierup et al., 2001), but this result may be due to the lack of power of recombination-detecting statistics, caused by the extraordinarily high polymorphism at these loci. Several successive mutations have occurred at the majority of segregating sites in these loci; this shows that mutation creates variation, but it also obscures the role of recombination. Recombination does play a substantial role in the creation of polymorphism in one species with gametophytic SI: *Petunia inflata* (Wang et al., 2001). The authors state that recombination events must be rare, and that recombinant alleles causing a reduction in fitness are removed by natural selection.

Numerous tests in *Arabidopsis* sporophytic SI systems have yielded scant evidence for recombination (Kamau and Charlesworth, 2005; Charlesworth et al., 2006; Hagenblad et al., 2006; Edh et al., 2009). Recombination has been detected in SI loci of *Brassica* species (Kusaba et al., 1997; Awadalla and Charlesworth, 1999; Takuno et al., 2007) but only in genes or gene domains that do not play a direct role in self-incompatibility specificity (Takuno et al., 2007; Edh et al., 2009). Mutation must create the majority of variation at SI loci; multiple mutations at variable sites are well documented (Edh et al., 2009).

Recombination likely contributes to *alr2* polymorphism in *Hydractinia*, based on the discovery of chimeric alleles having regions characteristic of two distinct types of structural polymorphism (Rosengarten et al., 2010), but the relative contribution of mutation and recombination to allelic diversity has not been assessed.

Except in the cases of sex-determining loci and SI loci in plants (where recombination is suppressed), mutation and recombination interact in allorecognition systems to create polymorphism.

MAINTENANCE OF POLYMORPHISM

DISTRIBUTION OF POLYMORPHISM WITHIN AND AMONG POPULATIONS

Using the *A*nalysis of *Mo*lecular *Va*riance (AMOVA), evolutionary biologists routinely partition the total molecular variation in a particular gene into three mutually exclusive groups: among geographical regions, among populations within geographical regions, and within populations. For example, geographical regions could be Europe and North America: how much of the variation is found when comparing these two regions? The "among populations within geographical regions" category asks, within Europe, how much of the variation is found when comparing the Valencia (Spain), Bergen (Norway), and Lucerne (Switzerland) populations? And finally, how much variation is found when populations are examined individually? If a large portion of the variation is found within populations, this means there is little genetic differentiation between populations (e.g., the same alleles would be found in Valencia, Bergen, and Lucerne). Fst is a related statistic; a statistically significant Fst signifies genetic differentiation between populations in the data set, pairwise Fst statistics are used to determine whether any pair of populations is significantly differentiated.

AMOVA and Fst calculations can inform us about the evolutionary forces operating on the allorecognition loci, allowing us to understand how polymorphism is maintained. Comparing allorecognition loci to neutral loci (usually microsatellites) with respect to Fst values and percentage of polymorphism within vs. among populations (AMOVA) allows one to test whether selection is occurring. Loci experiencing balancing selection (which maintains variation) should have larger amounts of polymorphism within populations and smaller amounts among populations than neutral loci (assuming selection pressures are similar between populations), whereas the opposite pattern is expected for loci experiencing directional selection (Schierup et al., 2000). Similar

genetic differentiation between allorecognition loci and neutral loci is taken as evidence for neutral evolution (e.g., genetic drift).

One might assume directional selection is acting when significant population structure is recovered, but allelic variation at allorecognition loci is likely older than current population structure. Limited gene flow between populations (conserving ancient variation), rather than selection, could lead to differentiation between contemporary populations (Richman et al., 2003). Because of this, strong inferences of selection on allorecognition loci should only be made from AMOVA and Fst values when these values are compared to other loci not presumed to be under balancing or directional selection.

AMOVA and Fst analyses have been completed in three allorecognition systems: *het/vic* loci in fungi, FuHC in *B. schlosseri*, and SI loci in the Asteraceae (*Guizotia abyssinica*), and Brassicaceae (*Arabidopsis* and *Brassica*). In both the chestnut blight fungus (*C. parasitica*) and the dry rot fungus (*Serpula lacrymans*), *het/vic* loci lack significant genetic differentiation among populations (Milgroom and Cortesi, 1999; Kauserud et al., 2006). In the ascidian *B. schlosseri*, >90% of the variation in the FuHC gene is found within populations, and Fst is not statistically significant (Nydam, Taylor, and De Tomaso unpublished data). This is in direct contrast to values obtained from two *B. schlosseri* genes not presumed to be under directional or balancing selection: mitochondrial cytochrome oxidase I and vasa. Both of these genes have less variation within populations than FuHC (81.2 and 27.16%, respectively), and both have highly significant Fst values ($p < 0.001$).

Patterns at SI loci are similar to those at *het/vic* and FuHC. In *G. abyssinica* (niger), 97% of the SI locus variation was found within populations, and Fst values were very low (although statistically significant; Geleta and Bryngelsson, 2010). Fst values are significantly lower when compared to neutral loci, in all cases (*A. lyrata*: Kamau et al., 2007; *A. halleri*: Ruggiero et al., 2008; *Brassica cretica*: Edh et al., 2009; *B. insularis*: Glemin et al., 2005). These results provide strong evidence for balancing selection driving the evolution of SI loci.

Het/vic, FuHC, and SI loci show similar patterns: a large percentage of the variation at these loci is found within populations, Fst values are not often statistically significant, and Fst values are significantly lower when compared with neutral loci. These patterns are consistent with a straightforward model of balancing selection, where selection pressures are similar in all environments.

TESTS OF SELECTION: POLYMORPHISM AND DIVERGENCE STATISTICS

Tests of selection using polymorphism (e.g., Tajima's *D*) and divergence (e.g., d_N/d_S) statistics commonly find evidence for selection at allorecognition loci. In fact, one of the earliest and most cited examples of $d_N/d_S > 1$ (non-synonymous substitution rate greater than synonymous substitution rate) comes from the peptide-binding region (PBR) in mouse and human MHC (Hughes and Nei, 1988). For polymorphism statistics such as Tajima's *D* and Fu and Li's D^* and F^*, values statistically greater than zero are evidence for balancing selection, and less than zero for directional selection (Tajima, 1989; Fu and Li, 1993). A pattern of $d_N/d_S > 1$ could indicate either directional or balancing selection (Garrigan and Hedrick, 2003); other data must be examined to determine which type of selection is occurring.

We will describe all the available data from the less well-studied loci (*tgrB1* and *tgrC1* in *D. discoideum*, *alr2* in *H. symbiolongicarpus,* mating-compatibility genes and *het/vic* loci in fungus, and FuHC in *B. schlosseri*). A complete description of all relevant studies in the SI literature is beyond the scope of this review; we instead highlight several recent studies from this allorecognition system.

In the cellular slime mold *D. discoideum,* the genes *tgrB1* and *tgrC1* are involved in kin recognition. Certain sections of these genes have d_N/d_S ratios > 1; the authors conclude that balancing selection is causing this pattern, given the extensive polymorphism at these loci (Benabentos et al., 2009). Nine codons in *alr2* of *H. symbiolongicarpus* have elevated d_N/d_S ratios; the majority are found in exon 2 (Rosengarten et al., 2010). The presence of 35 *alr2* alleles recovered from 36 individuals led the authors to conclude that negative frequency-dependent selection (a type of balancing selection where rare alleles are favored by selection) is occurring. At equilibrium, the alleles of a single locus subject to frequency-dependent selection are expected to be equally frequent (Grosberg, 1988).

Neither of two mating-compatibility genes examined in fungus species showed $d_N/d_S > 1$ (May et al., 1999; Rau et al., 2007), but the *b1* mating type gene in the mushroom fungus *C. cinereus* was shown to be experiencing balancing selection by comparing the topologies of gene genealogies under balancing selection and neutral scenarios (May et al., 1999). *Het-c* in *Neurospora crassa* was determined to be evolving under balancing selection; evidence included trans-species polymorphisms and an increase in non-synonymous substitutions in and around the specificity region of *het-c* (Wu et al., 1998). Four codon positions of the WD-40 repeats in *het-d* and *het-e* of *P. anserina* have $d_N/d_S > 1$. The authors conclude that balancing selection, rather than directional selection, is operating, because of the high number of amino acid combinations at the four codons of interest (Paoletti et al., 2007).

FuHC in *B. schlosseri* experiences selection, based on both polymorphism and d_N/d_S statistics (Nydam, Taylor, and De Tomaso unpublished data). Values of polymorphism statistics (Tajima's *D*, Fu and Li's D^* and F^*) were significantly negative in all East Coast populations, as well as Monterey, CA, USA on the West Coast, consistent with directional selection. But negative polymorphism statistics could be due to selective or demographic processes (e.g., recent population growth). In the case of FuHC, the pattern is likely due to selection rather than demography, given that none of the polymorphism statistics were significantly negative for a housekeeping gene (vasa). 11 additional housekeeping genes are currently being sequenced, to confirm that demographic processes are not causing this pattern. Omega statistics pinpointed 24 codons throughout the protein have a greater than 95% probability of $d_N/d_S > 1$. Four exon groups contained clusters of these positively selected sites: Exons 5, 6, 20, and 27. Exons 5, 6, and 20 had significantly higher omega values than the rest of the gene for a subset of populations; these exons will be targeted in future functional studies. Other tests are being conducted to determine whether this pattern is due to balancing or directional selection.

Inference of selection at SI loci in plants begins with Sewall Wright, who wrote, "It also fairly obvious that selection would tend

to increase the frequency of any additional alleles that may appear." (Wright, 1939). Because fertilization is aborted when pollen and pistil S-allele are identical, rare S-alleles have a selective advantage. As negative frequency-dependent selection is a type of balancing selection, researchers have spent considerable effort determining whether SI loci are evolving under balancing selection.

Data from numerous plant groups provide considerable support for balancing selection on SI loci. As in other allorecognition systems, the majority of these data are d_N/d_S ratios > 1. However, d_N/d_S ratios > 1 are consistent with directional and balancing selection, so additional data are needed to determine which of these scenarios is occurring. d_N/d_S ratios > 1 have been found in many plant families, both in gametophytic and sporophytic SI systems (Clark and Kao, 1991; Ishimizu et al., 1998; Sato et al., 2002; Takebayashi et al., 2003; Igic et al., 2007; Guo et al., 2011). d_N/d_S ratios > 1 were corroborated with additional data to infer the action of balancing selection: significantly positive Tajima's *D* values, little population structure compared to neutral markers, and low recombination for *SRK* and *SCR* in *B. cretica* (Edh et al., 2009), trans-species polymorphisms in *SRK* and *SCR* in several *Arabidopsis* species (Sato et al., 2002; Guo et al., 2011).

CONCLUSION

Unusually high polymorphism is a hallmark of allorecognition loci; how this polymorphism is created and maintained has interested biologists since Sewall Wright. From the studies presented in this review, we can conclude that polymorphism is created by an interaction between mutation and recombination, except in the cases of sex-determining loci and SI loci in plants (where recombination is suppressed).

AMOVA/Fst studies examining the distribution of polymorphism within and among populations in *Het/vic*, FuHC, and SI loci generally provide support for the role of balancing selection in maintaining polymorphism.

Divergence statistics often show patterns of d_N/d_S ratios > 1 for allorecognition loci; these values are consistent with both directional and balancing selection. In many cases, additional evidence such as significantly positive polymorphism statistics and/or identification of trans-species polymorphisms provide support for balancing over directional selection in the maintenance of variation in allorecognition loci.

FUTURE DIRECTIONS

We have only identified the loci involved in a handful of allorecognition systems, and an obvious future direction lies in identifying allorecognition loci in anemones, bryozoans, corals, plasmodial slime molds, and sponges. Such discoveries require painstaking work, and in some cases are hindered by the biology of the organisms. So while this may be a long term goal, it is nevertheless an important one if we are to achieve an understanding of the evolution of allorecognition across the tree of life.

A shorter term direction for evolutionary studies of allorecognition should involve the loci that have recently been discovered. We could not locate information on the distribution of polymorphism within and among populations for *tgrB1/tgrC1* in *D. discoideum*, or *alr2* in *H. symbiolongicarpus*, or on the creation of polymorphism for *tgrB1/tgrC1* in *D. discoideum*. No evolutionary studies have been conducted on the self-incompatibility genes in *C. intestinalis*. Additionally, recent studies concluding that balancing selection is occurring based on divergence statistics and the presence of polymorphism could provide more data to support these conclusions, including evidence of trans-species polymorphisms and AMOVA/Fst analyses showing larger amounts of polymorphism within populations and smaller amounts among populations than neutral loci. And finally, we have some information on the specific type of balancing selection operating in a well-studied allorecognition system (negative frequency-dependent selection in SI loci). Further analyses of the newly discovered loci should uncover more specifics about the types of balancing selection maintaining variation at these loci.

In well-studied allorecognition systems like the SI loci in plants, we have information on the evolution of allorecognition loci in many species. This allows us to examine the evolution of allorecognition across large phylogenetic groups, enabling broader conclusions than would be possible if only model organisms were used. Allorecognition loci can now be more easily identified and studied in species closely related to *B. schlosseri*, *C. intestinalis*, *D. discoideum*, and *H. symbiolongicarpus*. Using this approach, we could begin to understand the evolution of allorecognition not just in *B. schlosseri* and *C. intestinalis*, but in the urochordates as a whole, not just in *D. discoideum* but in the social amoebae, not just in *H. symbiolongicarpus* but in hydroids generally.

REFERENCES

Allen, A. M., and Hiscock, S. J. (2008). "Evolution and phylogeny of self-incompatibility systems in angiosperms," in *Self-Incompatibility in Flowering Plants – Evolution, Diversity and Mechanisms*, ed. V. E. Franklin-Tong (Berlin: Springer-Verlag), 73–101.

Awadalla, P., and Charlesworth, D. (1999). Recombination and selection at *Brassica* self-incompatibility loci. *Genetics* 152, 413–425.

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E., Strassmann, J., Queller, D., Zupan, B., Shaulsky, G., and Kuspa, A. (2009). Polymorphic members of the lag-gene family mediate kin-discrimination in *Dictyostelium*. *Curr. Biol.* 19, 567–572.

Ben-Shlomo, R. (2008). The molecular basis of allorecognition in ascidians. *Bioessays* 30, 1048–1051.

Buss, L. (1982). Somatic cell parasitism and the evolution of somatic tissue compatibility. *Proc. Natl. Acad. Sci. U.S.A.* 79, 5337–5341.

Charlesworth, D., Kamau, E., Hagenblad, J., and Tang, T. L. (2006). Trans-specificity at loci near the self-incompatibility locus in *Arabidopsis*. *Genetics* 172, 2699–2707.

Clark, A. G., and Kao, T.-H. (1991). Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9823–9827.

Clark, J. (2003). Plasmodial incompatibility in the myxomycete *Didymium squamulosum*. *Mycologia* 95, 24–26.

Day, P. R. (1963). The structure of the A mating-type factor in *Coprinus lagopus*: wild alleles. *Genet. Res.* 4, 323–325.

De Tomaso, A. W., Nyholm, S. V., Palmeri, K. J., Ishizuka K. J., Ludington, W. B., Mitchel, K., and Weissman I. L. (2005). Isolation and characterization of a protochordate histocompatibility locus. *Nature* 438, 454–459.

Edh, K., Widen, B., and Ceplitis, A. (2009). Molecular population genetics of the SRK and SCR Self-Incompatibility genes in the wild plant species *Brassica cretica* (Brassicaceae). *Genetics* 181, 985–995.

Fernandez-Busquets, X., and Burger, M. M. (1999). Cell adhesion and histocompatibility in sponges. *Microsc. Res. Tech.* 44, 204–218.

Fu, Y. X., and Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics* 133, 693–709.

Garrigan, D., and Hedrick, P. W. (2003). Detecting adaptive molecular polymorphism: lessons from the MHC. *Evolution* 57, 1707–1722.

Geleta, M., and Bryngelsson, T. (2010). Population genetics of self-incompatibility and developing self-compatible genotypes in niger (*Guizotia abyssinica*). *Euphytica* 176, 417–430.

Gibbs, K. A., Urbanowski, M. L., and Greenberg, E. P. (2008). Genetic determinants of self identity and social recognition in bacteria. *Science* 321, 256–259.

Glass, N. L., Jacobson, D. J., and Shiu, P. K. (2000). The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. *Annu. Rev. Genet.* 34, 165–186.

Glemin, S., Gaude, T., Guillemin, M.-L., Lourmas, M., Olivieri, E., and Mignot, A. (2005). Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*. *Genetics* 171, 279–289.

Grosberg, R. K. (1988). The evolution of allorecognition specificity in clonal invertebrates. *Q. Rev. Biol.* 63, 377–412.

Grosberg, R. K., Levitan, D. R., and Cameron, B. B. (1996). Evolutionary genetics of allorecognition in the colonial hydroid *Hydractinia symbiolongicarpus*. *Evolution* 50, 2221–2240.

Grosberg, R. K., and Plachetzki, D. (2010). "Marine invertebrates: genetics of colony recognition," in *Encyclopedia of Animal Behavior*, eds M. D. Breed and J. Moore (Oxford: Academic Press), 381–388.

Guo, Y.-L., Zhao, X., Lanz, C., and Weigel, D. (2011). Evolution of the S-locus region in *Arabidopsis thaliana* relatives. *Plant Physiol.* 157, 937–946.

Hagenblad, J., Bechsgaard, J., and Charlesworth, D. (2006). Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. *Genetics* 173, 1057–1073.

Harada, Y., Takagaki, Y., Sunagawa, M., Saito, H., Yamada, L., Taniguchi, H., Shoguchi, E., and Sawada, H. (2008). Mechanism of self sterility in a hermaphroditic chordate. *Science* 320, 548–550.

Hidaka, M., Yurugi, K., Sunagawa, S., and Kinzie, R. A. III. (1997). Contact reactions between young colonies of the coral *Pocillopora damicornis*. *Coral Reefs* 16, 13–20.

Hudson, R. R. (1987). Estimating the recombination parameter of a finite population model without selection. *Genet. Res.* 50, 245–250.

Hughes, A. L., and Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335, 167–170.

Hughes, R. N., Manriquez, P. H., Morley, S., Craig, S. F., and Bishop, J. D. D. (2004). Kin or self-recognition? Colonial fusibility of the bryozoan *Celleporella hyalina*. *Evol. Dev.* 6, 431–437.

Igic, B., Smith, W. A., Robertson, K. A., Schaal, B. A., and Kohn, J. R. (2007). Studies of self-incompatibility in wild tomatoes: I. S-allele diversity of *Solanum chilense* Dun. (Solanaceae). *Heredity* 99, 553–561.

Ishimizu, T., Endo, T., Yamaguchi-Kabata, Y., Nakamura, K. T., Sakiyama, F., and Norioka, S. (1998). Identification of regions in which positive selection may operate in S-RNase of Rosaceae: implication for S-allele-specific recognition sites in S-RNase. *FEBS Lett.* 440, 337–342.

Kamau, E., Charlesworth, B., and Charlesworth, D. (2007). Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. *Genetics* 176, 2357–2369.

Kamau, E., and Charlesworth, D. (2005). Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Curr. Biol.* 15, 1773–1778.

Kauserud, H., Saetre, G.-P., Schmidt, O., Decock, C., and Schumacher, T. (2006). Genetics of self/nonself recognition in *Serpula lacrymans*. *Fungal Genet. Biol.* 43, 503–510.

Kubisiak, T. L., and Milgroom, M. G. (2006). Markers linked to vegetative incompatibility (vic) genes and a region of high heterogeneity and reduced recombination near the mating type locus (MAT) in *Cryphonectria parasitica*. *Fungal Genet. Biol.* 43, 453–463.

Kusaba, M., Nishio, T., Satta, Y., Hinata, K., and Ockendon, D. (1997). Striking sequence similarity in inter-and intra-specific comparisons of class I SLG alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7673–7678.

Lawrence, M. J. (1996). Number of incompatibility alleles in clover and other species. *Heredity* 76, 610–615.

Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.

Mable, B. K., Schierup, M. H., and Charlesworth, D. (2003). Estimating the number, frequency, and dominance of S-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* 90, 422–431.

May, G., and Matzke, E. (1995). Recombination and variation at the A mating-type of *Coprinus cinereus*. *Mol. Biol. Evol.* 12, 794–802.

May, G., Shaw, F., Badrane, H., and Vekemans, H. (1999). The signature of balancing selection: fungal mating compatibility gene evolution. *Proc. Natl. Acad. Sci. U.S.A.* 96, 9172–9177.

Mercier, A., Sun, Z., and Hamel, J.-F. (2011). Internal brooding favours pre-metamorphic chimerism in a non-colonial cnidarian, the sea anemone *Urticina felina*. *Proc. Biol. Sci.* 278, 3517–3522.

Milgroom, M. G., and Cortesi, P. (1999). Analysis of population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10518–10523.

Newbigin, E., Anderson, M. A., and Clarke, A. E. (1993). Gametophytic self-incompatibility systems. *Plant Cell* 5, 1315–1324.

Nicotra, M. L., Powell, A. E., Rosengarten, R. D., Moreno, M., Grimwood, J., Lakkis, F. G., Dellaporta, S. L., and Buss, L. W. (2009). A hypervariable invertebrate allodeterminant. *Curr. Biol.* 19, 583–589.

Pandey, K. K. (1960). Incompatibility in *Abutilon* 'hybridum.' *Am. J. Bot.* 47, 877–883.

Paoletti, M., Saupe, S. J., and Clave, C. (2007). Genesis of a fungal non-self recognition repertoire. *PLoS ONE* 3, e283. doi:10.1371/journal.pone.0000283

Raftos, D. (1994). Allorecognition and humoral immunity and tunicates. *Ann. N. Y. Acad. Sci.* 712, 227–244.

Rau, D., Attene, G., Brown, A. H. D., Nanni, L., Maier, F. J., Balmas, V., Saba, E., Schafer, W., and Papa, R. (2007). Phylogeny and evolution of mating-type genes from *Pyrenophora teres,* the causal agent of barley "net blotch" disease. *Curr. Genet.* 51, 377–392.

Rosengarten, R. D., Moreno, M. A., Lakkis, F. G., Buss, L. W., and Dellaporta, S. L. (2010). Genetic diversity of the allodeterminant alr2 in *Hydractinia symbiolongicarpus*. *Mol. Biol. Evol.* 28, 933–947.

Richman, A. D., Gerardo Herrera, L., Nash, D., and Schierup, M. K. (2003). Relative roles of mutation and recombination in generating allelic polymorphism at an MHC class II locus in Peromyscus maniculatus. *Genet. Res.* 82, 89–99.

Ruggiero, M. V., *Jacquemin*, B., *Castric*, V., and *Vekemans*, X. (2008). Hitchhiking to a locus under balancing selection: high sequence diversity and low population subdivision at the S-locus genomic region in *Arabidopsis halleri*. *Genet. Res.* 90, 37–46.

Runions, C. J., and Owens, J. N. (1998). "Evidence of pre-zygotic self-incompatibility in a conifer," in *Reproductive Biology*, eds S. J. Owens and P. J. Rudall (Kew: Royal Botanic Gardens), 255–264.

Saito, Y., Hirose, E., and Watanabe, H. (1994). Allorecognition in compound ascidians. *Int. J. Dev. Biol.* 38, 237–247.

Sato, K., Nishio, T., Kimura, R., Kusaba, M., Suzuki, T., Hatakeyama, K., Ockendon, D. J., and Satta, Y. (2002). Coevolution of the S-Locus genes SRK, SLG, and SP11/SCR in *Brassica oleracea* and *B. rapa*. *Genetics* 162, 931–940.

Schierup, M. H., Mikkelsen, A. M., and Hein, J. (2001). Recombination, balancing selection, and phylogenies in MHC and self-incompatibility genes. *Genetics* 159, 1833–1844.

Schierup, M. H., Vekemans, X., and Charlesworth, D. (2000). The effect of subdivision on variation at multiallelic loci under balancing selection. *Genet. Res.* 76, 51–62.

Shaulsky, G., and Kessin, R. (2007). The cold war of the social amoebae. *Curr. Biol.* 17, 684–692.

Stein, J., Howlett, B., Boyes, D. C., Nasrallah, M. E., and Nasrallah, J. B. (1991). Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proc. Natl. Acad. Sci. U.S.A.* 88, 8816–8820.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123, 585–595.

Takebayashi, N., Brewer, P. B., Newbigin, E., and Uyenoyama, M. K. (2003). Patterns of variation within self-incompatibility loci. *Mol. Biol. Evol.* 20, 1778–1794.

Takuno, S., Fujimoto, R., Sugimura, T., Sato, K., Okamoto, S., Zhang, S.-L., and Nishio, T. (2007). Effects of recombination on hitchhiking

diversity in the *Brassica* self-incompatibility locus complex. *Genetics* 177, 949–958.

Wang, X., Hughes, A. L., Tsukamoto, T., Ando, T., and Kao, T.-H. (2001). Evidence that intragenic recombination contributes to allelic diversity of the S-RNase gene at the self-incompatibility (S) locus in *Petunia inflata*. *Plant Physiol.* 125, 1012–1022.

Wright, S. (1939). The distribution of self-sterility alleles in populations. *Genetics* 24, 538–552.

Wu, J., Saupe, S. J., and Glass, N. L. (1998). Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12398–12403.

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.

Received: 18 October 2011; paper pending published: 14 November 2011; accepted: 02 December 2011; published online: 27 December 2011.

*Citation: Nydam ML and De Tomaso AW. (2011) Creation and maintenance of variation in allorecognition loci: molecular analysis in various model systems. Front. Immun. **2**:79. doi: 10.3389/fimmu.2011.00079*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

ORIGINAL RESEARCH ARTICLE
published: 05 June 2012
doi: 10.3389/fimmu.2012.00136

Dynamic evolution of toll-like receptor multigene families in echinoderms

Katherine M. Buckley [1,2] ***and Jonathan P. Rast*** [1,2] ***

[1] *Department of Immunology, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada*
[2] *Department of Medical Biophysics, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada*

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Mike Criscitiello, Texas A&M University, USA
Zeev Pancer, University of Maryland, USA
Coenraad Adema, University of New Mexico, USA

***Correspondence:**
Jonathan P. Rast, Department of Medical Biophysics and Department of Immunology, Sunnybrook Research Institute, University of Toronto, 2075 Bayview Avenue, Room S126B, Toronto, ON, Canada M4N 3M5.
e-mail: jrast@sri.utoronto.ca

The genome sequence of the purple sea urchin, *Strongylocentrotus purpuratus*, a large and long-lived invertebrate, provides a new perspective on animal immunity. Analysis of this genome uncovered a highly complex immune system in which the gene families that encode homologs of the pattern recognition receptors that form the core of vertebrate innate immunity are encoded in large multigene families. The sea urchin genome contains 253 Toll-like receptor (TLR) sequences, more than 200 Nod-like receptors and 1095 scavenger receptor cysteine-rich domains, a 10-fold expansion relative to vertebrates. Given their stereotypic protein structure and simple intron-exon architecture, the TLRs are the most tractable of these families for more detailed analysis. A role for these receptors in immune defense is suggested by their similarity to TLRs in other organisms, sequence diversity, and expression in immunologically active tissues, including phagocytes. The complexity of the sea urchin TLR multigene families is largely derived from expansions independent of those in vertebrates and protostomes, although a small family of TLRs with structure similar to that of *Drosophila* Toll can be traced to an ancient eumetazoan ancestor. Several other echinoderm sequences are now available, including *Lytechinus variegatus*, as well as partial sequences from two other sea urchin species. Here, we present an analysis of the invertebrate deuterostome TLRs with emphasis on the echinoderms. Representatives of most of the *S. purpuratus* TLR subfamilies and homologs of the mccTLR sequences are found in *L. variegatus*, although the *L. variegatus* TLR gene family is notably smaller (68 TLR sequences). The phylogeny of these genes within sea urchins highlights lineage-specific expansions at higher resolution than is evident at the phylum level. These analyses identify quickly evolving TLR subfamilies that are likely to have novel immune recognition functions and other, more stable, subfamilies that may function more similarly to those of vertebrates.

Keywords: toll-like receptors, sea urchins, multigene family, evolution, innate immunity

INTRODUCTION

The discovery of an immune function for *Drosophila* Toll (Lemaitre et al., 1996) and the subsequent identification of immune recognition roles for mammalian Toll-like receptor (TLR) 4 (Medzhitov et al., 1997; Poltorak et al., 1998) catalyzed an intensely renewed interest in innate immunity and more generally an appreciation for the potential of invertebrate models in mainstream immunology. As genome sequences from an increasing number of animal phyla are resolved, it has become clear that TLRs are present in virtually all eumetazoans (Messier-Solek et al., 2010). In the genome of the purple sea urchin, *Strongylocentrotus purpuratus*, these receptors are encoded in a very large multigene family that contrasts sharply with the small families of insects and vertebrates (Hibino et al., 2006). Recently sequenced genomes from several animals, including another invertebrate deuterostome, amphioxus, and the annelid *Capitella capitata* suggest that large TLR repertoires may be widespread throughout Bilateria (e.g., Davidson et al., 2008; Huang et al., 2008). An understanding of the function of these TLRs may provide a new perspective on this important family of innate immune receptors.

It is far from settled whether or not these TLRs function in immunity (Leulier and Lemaitre, 2008). In insects and mammals, the two animals groups for which function is well-understood, the mechanisms by which TLRs recognize non-self and the systems in which they operate differ considerably. In mammals, where all TLRs function as immune receptors that interact directly with non-self factors, defense is the primary role. In contrast, *Drosophila* Toll signals far downstream of immune recognition and thus its role in immune recognition is indirect (Lemaitre and Hoffmann, 2007). The remaining eight *Drosophila* TLRs have not been associated with immunity and, where their function is defined, are more closely associated with development and other cellular processes. In *Drosophila*, Toll-9 is the single member of the Toll family that structurally resembles vertebrate TLRs. Although early work suggested that Toll-9 may be responsible for maintaining constitutive expression of antimicrobial peptides (Ooi et al., 2002), more recent studies analyzing Toll-9 mutants reveal that this protein is not required to mount an efficient antibacterial response (Narbonne-Reveau et al., 2011). The central role of *Drosophila* Toll signaling in mesoderm patterning (Huang et al., 1997) has

not been demonstrated outside of insects. While mammalian TLRs have relatively modest roles in modulating cell differentiation (e.g., in the gut), these are sequential to their function in immune recognition and are not counterparts to the developmental function of *Drosophila* Toll. Ancient homologs of TLRs are also present within the genomes of the cnidarians *Nematostella vectensis* and *Hydra magnipapillata* (Miller et al., 2007). The single TLR in *Nematostella* structurally resembles Toll, although its function has not been investigated. In contrast, the *Hydra* genome encodes four Toll-related proteins (HyTRR-1/HyLRR-1 and HyTRR2/Hy-LRR2) that interact to form two receptors that have been shown to play a role in epithelial immunity (Bosch et al., 2009). Thus it remains difficult to make definitive statements about function across animal phyla and inference of ancestral function remains elusive, although there is some indication of an ancient immune role.

Despite these difficulties, other characteristics of the genes that encode TLRs in the sea urchin and other animal genomes may shed light on their function and thus on TLR evolution. Here we present an analysis of TLR multiplicity, phylogeny, diversity, and expression in the purple sea urchin against the background of new sequence information from other sea urchin species. We find that the unique characteristics of TLRs in the purple sea urchin are present also in other sea urchin species. The multiplicity, apparently rapid gene turnover, and sequence diversity of the TLRs within this complex gene family, in addition to enriched expression in immunologically active tissues are consistent with a role in immunity. Most notably the evolutionary patterns of family member diversification suggest rapid changes in binding potential that are unlike those seen in the TLRs of vertebrates or *Drosophila*. Thus, TLRs in the sea urchin, and possibly other Bilateria, may have been co-opted for use in an immune recognition strategy that is more evolutionarily dynamic than the pathogen-associated molecular pattern (PAMP)-based systems of vertebrates and insects. In contrast to the paradigm of vertebrate TLRs, in which conserved receptors recognize static microbial elements, in sea urchins, closely related but rapidly diversifying variants of receptors may respond to quickly evolving pathogens.

MATERIALS AND METHODS

SEQUENCE ANALYSIS

The *S. purpuratus* genome sequence (v3.1; released July, 2011) was obtained from SpBase[1] (Cameron et al., 2009). Additional echinoderm genome sequences and unassembled genomic traces were obtained from the Sea Urchin Genome Project website of the Human Genome Sequencing Center at Baylor College of Medicine (HGSC-BCM[2]) and the National Center for Biotechnology Information (NCBI; *Lytechinus variegatus* GenBank Assembly ID: GCA_000239495.1; *L. variegatus* 454 sequence: SRX112894, SRX112895, SRX112896). The *Saccoglossus kowalevskii* genome (Skow_1.0) was obtained from the HGSC-BCM website[3].

Genome sequences were translated and open reading frames were identified using tools within the EMBOSS package.[4] All potential open reading frames greater than 75 amino acids in length, without requirement to start with a methionine were analyzed. Domain searches were performed with HMMER 3.0[5] and leucine-rich repeats (LRRs) were identified using LRRfinder (Offord et al., 2010). TLR sequences were classified into three categories: (1) *complete* genes that were uninterrupted by a stop codon in the translated sequence; (2) *pseudogenes* that were characterized by an in frame stop codon or frame shift leading to missense sequence; or (3) *partial* genes in which the sequences were truncated by either the end of a scaffold or indeterminate sequences (N's). Genomic coordinates and descriptions of the TLR sequences can be found in Tables S1 and S2 in Supplementary Material. Sequences were aligned using ClustalX (Larkin et al., 2007), and alignments were manually edited in Bioedit (Hall, 1999). Sequence entropy was calculated based on described methods (Durbin et al., 1998).

Phylogenetic analyses of the Toll/Interleukin-1 Receptor (TIR) domains were done in MEGA5.0 (Tamura et al., 2011). Neighbor-joining trees were constructed using evolutionary distances calculated with the Poisson correction method. Alignment positions containing gaps were removed from the entire analysis. Bootstrap support was calculated based on 1,000 replicates.

An analysis of evolutionary selection was performed for each TLR subfamily that contained eight or more complete, non-pseudogene sequences. Sequence alignments for that were used for these analyses are in Files S1–S6 in Supplementary Material. A maximum likelihood tree built in PHYLIP (Felsenstein, 2005) served as the working topology for the analyses. Selection within the sequences was analyzed in CODEML within PAML (Yang, 2007) under two models: the M7 model, which allows neutral or purifying selection, and the M8 model, which also includes a class of sites that evolve under positive selection. The two models were compared using a likelihood ratio test (Yang, 1998). Residues under positive selection were identified using the Bayes empirical Bayes approach under the M8 model (Yang, 2007).

To validate the multiplicity of the TLR gene families within the *S. purpuratus* and *L. variegatus* assembled genomes and also to estimate the gene family sizes in *Allocentrotus fragilis* and *Strongylocentrotus franciscanus*, we analyzed the unassembled genomic traces. The amino acid sequences of the TIR domains from the *S. purpuratus* and *L. variegatus* TLRs were used as queries in a tblastn search against the unassembled traces from *A. fragilis*, *S. franciscanus*, and *L. variegatus*. All traces that matched with an *e*-value of less than 0.01 were collected and used as queries in a blastx search against the TIR domains to classify the partial sequences by subfamily and to enumerate the sequences.

LARVAL CULTURE, INFECTION MODEL, AND COELOMOCYTES

S. purpuratus larvae were maintained at a concentration of four larvae per mL in artificial seawater (ASW; Instant Ocean) at 15°C and fed *Rhodomonas lens* (5,000/mL) starting at 5 days post-fertilization (dpf). For some of measurements of TLR transcript prevalence *S. purpuratus* larvae were exposed to *Vibrio diazotrophicus* (ATCC strain 33466). Samples were collected at 0, 6, 12, and

[1] www.spbase.org
[2] www.hgsc.bcm.tmc.edu/projects/seaurchin
[3] http://www.hgsc.bcm.tmc.edu/project-species-o-Acorn%20worm.hgsc
[4] emboss.sourceforge.net
[5] www.hmmer.org

24 h of exposure to bacteria and used in RNA-Seq analysis. The larvae in these four samples were derived a single fertilization of eggs from one female.

To induce an immune response, a single adult animal was injected intracoelomically with complex microbiota isolated from the gut of another adult animal (4.8×10^6 total bacteria). After 12 h, whole coelomocytes and gut tissue were collected for RNA-Seq experiments. Phagocytic coelomocytes were isolated using discontinuous gradient density centrifugation (Gross et al., 2000). Gut tissue was homogenized for RNA extraction and consisted of mixed samples from the entire length of the gut.

RNA-Seq

Total RNA was isolated using Trizol (Invitrogen), and mRNA was purified with the Poly(A)Purist kit (Ambion). cDNA sequencing was performed on an Applied Biosystems SOLiD4 and 5500 SOLiD machines at the Sunnybrook Genomics Facility. For the larval and coelomocyte samples, the paired-end reads were 50 and 35 nt long; from the gut, paired-end reads were 75 and 35 nt in length.

Sequences were mapped in color space to the *S. purpuratus* genome (v3.1) using Bowtie version 0.12.7 (Langmead et al., 2009) with the following parameters that differed from the default: up to 50 alignments reported for each read (-k); reads with greater than 50 alignments suppressed (-m); the maximum number of mismatches in the seed was set at 3 (-n); the maximum sum of the quality scores for mismatches was 900 (-e); five nucleotides were trimmed from the 3′ ends of the reads (-3); and the SNP fraction was set at 0.04 (–snpfrac), which is consistent with estimates of SNPs in the sea urchin genome (Sodergren et al., 2006). Only reads that mapped to the TIR domains were included in the expression analysis of TLR subfamilies and reads that mapped to TLR genes from more than one subfamily were excluded. TIR domain sequences for which >5% of the reads mapped in the incorrect direction with respect to the coding sequence were not included in the analysis.

RESULTS

THE EXPANDED TLR FAMILY IN THE PURPLE SEA URCHIN

Toll-like receptors are type-1 transmembrane proteins with a solenoid-like ectodomain structure composed of a series of LRRs that is responsible for ligand-binding (Jin and Lee, 2008). The hydrophobic core of this structure is capped on either end by specialized cysteine-rich LRR-NT and LRR-CT domains that are distinct in sequence and structure from the central LRRs. Hereafter, "LRRs" refers only to the central repeats. C-terminal to the single transmembrane region is a TIR domain that mediates interactions with downstream signaling factors (Gay and Keith, 1991; O'Neill and Bowie, 2007). Our previous analysis of the *S. purpuratus* genome (v2.1) identified 222 genes that encode TLR homologs (Hibino et al., 2006). This genome assembly contained 114,222 scaffolds with an N50 of 123.5 kb. Improvements to the assembly using additional BAC sequencing and high-throughput next-generation sequencing strategies have resulted in the most recent version (v3.1) that is composed of 32,008 scaffolds with an N50 of 401.9 kb (see text footnote 1). To incorporate these updates to the genome sequence into our analysis of the *S. purpuratus* TLR gene family, we reanalyzed the improved genome (v3.1) to identify open reading frames that contained TIR domains (Pfam domain PF01582.12). The majority of sea urchin TLRs are encoded in a single exon, which enables their identification directly from the translated genome, rather the predicted gene models. The sequence flanking the TIR domains was analyzed for the presence of other protein domains, including a transmembrane region, LRR-CT, central LRRs, LRR-NT, and signal peptides. In total, 284 TIR domains were identified in the genome that were part of authentic genes or pseudogenes. TIR domains are also present in several other molecules, including the TLR adaptors and IL1R family members, which were excluded from the analysis. The remaining TIR domains defined 253 TLR sequences within the sea urchin genome.

Most of the sea urchin TLR proteins (240) are structurally similar to those of vertebrates (**Figure 1**). The LRRs in the ectodomains of these proteins are flanked by LRR-NT and LRR-CT domains. TLRs with this type of extracellular domain are structurally distinct from *Drosophila* Toll (Rock et al., 1998) and are known as single cysteine cluster TLRs (sccTLRs; Leulier and Lemaitre, 2008). The sea urchin sccTLRs have between 21 and 25 LRRs. This is the structure of the vertebrate TLRs as well as *Drosophila* Toll-9 (**Table 1**). In addition, the sea urchin genome contains 13 TLRs that differ from the sccTLRs both in the structure of the ectodomain and also in the sequence of the TIR domain. Five of these divergent TLRs are characterized by shortened ectodomains that are composed of nine LRRs, rather than the typical 21–25. The LRRs within the ectodomains of these short TLRs are flanked by LRR-NT and LRR-CT domains (**Figure 1**). Four of the divergent TLRs, which comprise a supported clade, resemble the sccTLRs with respect to domain architecture, but the coding sequence is interrupted by a single intron. Finally, the ectodomains of four of the sea urchin TLRs resemble those of *Drosophila* Toll, in which LRR-CT and LRR-NT domains interrupt the typical LRRs. This domain organization has been termed multiple cysteine cluster TLRs (mccTLRs; **Figure 1**; Leulier and Lemaitre, 2008) and is the predominant structure of the *Drosophila* Toll proteins (**Table 1**).

The TIR domains of the 253 TLR sequences were used in phylogenetic analysis to further classify the genes (**Figure 2**). The 240 sccTLRs form a strongly supported clade that is distinct from the divergent short, intron-containing, and mccTLR sequences. Our previous analysis of these sequences identified seven groups of sccTLRs (I–VII; Hibino et al., 2006). Here, we describe the presence of an additional four groups (VIII–XI) based on conservation with other sea urchin species and by eliminating the previously named "orphan" sequences (**Figure 2**; File S7 in Supplementary Material). Some of the groups are also divided into smaller subfamilies. The group I genes fall into eight subfamilies (Ia–Ih) and the group II genes form the IIa and IIb subfamilies. Groups vary considerably in multiplicity and sequence variability. The largest subfamily (Ia) consists of 48 closely related genes. In contrast, the eight TLRs that belong to group VI are on longer branches that may reflect a more ancient evolutionary history (**Figure 2**). In contrast with our analysis of the TLRs from the previous genome assembly, an additional 31 TLR sequences were identified, the majority of which belong to the Ic subfamily [there were 13 Ic genes in the v2.1 assembly (Hibino et al., 2006), and 37 in v3.1 (**Table 3**)]. The genes within this subfamily are clustered in large tandem genomic

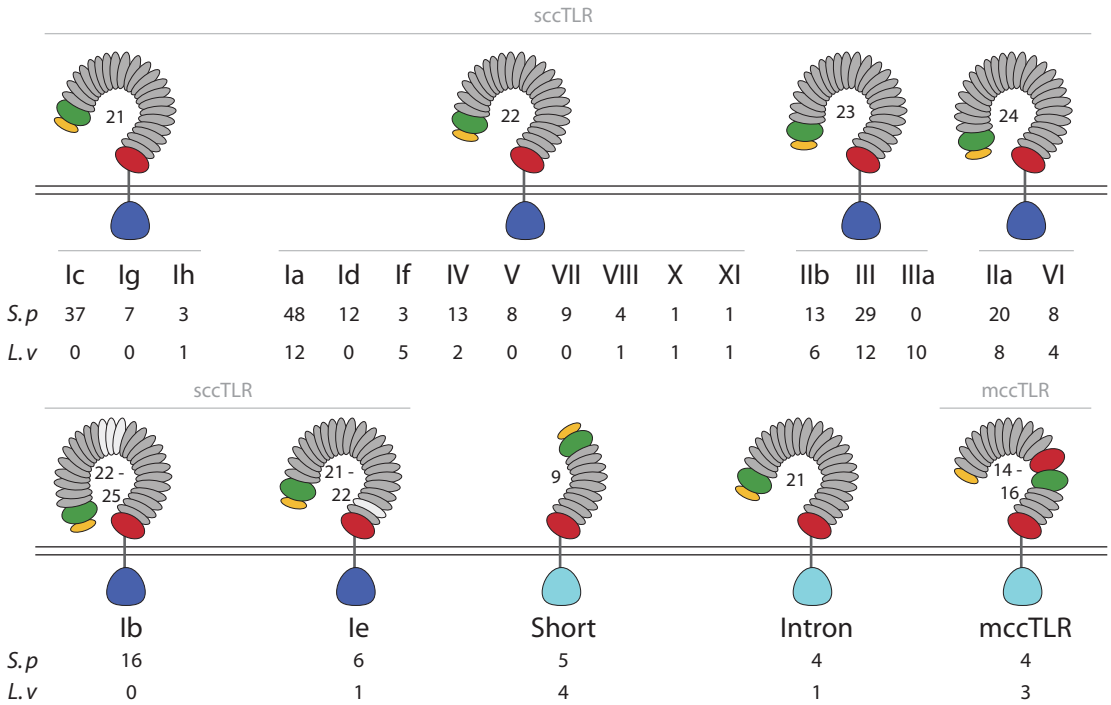


FIGURE 1 | The TLR subfamily in sea urchins. The protein domain structure for each of the TLR subfamilies is shown. Among the sccTLR sequences, the number of LRRs varies from 21 to 25. TLRs in subfamilies Ib and Ie vary in the total number of LRRs (LRRs that are not present in all sequences are shown in light gray). The divergent TIR sequences that are present in the mccTLRs, intron-containing, and short TLRs are shown in light blue (see **Figure 2**). The numbers of TIR domains from each group and subfamily that are present in the *S. purpuratus* and *L. variegatus* genomes are shown below the diagram structures.

Table 1 | Sizes of TLR gene family vary among species.

Phylogeny		Species	sccTLRs	mccTLRs
Deuterostome	Chordate	*Homo sapiens*	11	0
		Mus musculus	13	0
		Petromyzon marinus[1]	16	0
		Ciona intestinalis[2]	3	0
		Branchiostoma floridae[3]	60	12
	Echinoderm	*Strongylocentrotus purpuratus*	250	3
		Allocentrotus fragilis	276[4]	>1
		Strongylocentrotus franciscanus	228[4]	>1
		Lytechinus variegatus	64	3
	Hemichordate	*Saccoglossus kowalevskii*	7	1
Protostome	Ecdysozoa	*Drosophila melanogaster*	1	8
		Caenorhabditis elegans	0	1
	Lophotrochozoa	*Capitella capitata*[5]	104	1
		Helobdella robusta[5]	0	16
Cnidarian		*Nematostella vectensis*[6]	0	1
		Hydra magnipapillata[6]	2	

[1] *Kasamatsu et al. (2010); our independent analysis of this genome identified 19 TLRs.*

[2] *Sasaki et al. (2009).*

[3] *Holland et al. (2008).*

[4] *Estimates based on number of traces (see Table S3 in Supplementary Material).*

[5] *Davidson et al. (2008).*

[6] *Miller et al. (2007). The sequences in Hydra are divergent, TLR-related molecules consisting of two chains that cannot be assigned to either the sccTLRs or mccTLRs.*

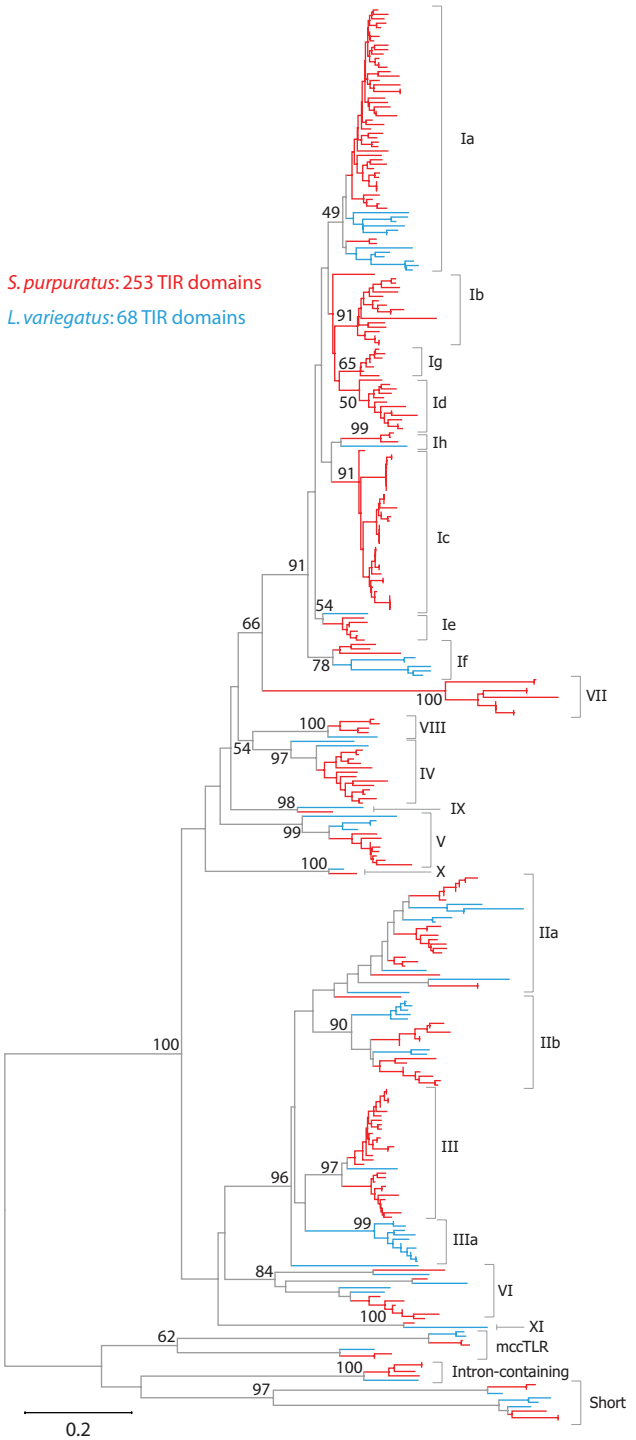


FIGURE 2 | Phylogeny of sea urchin TLR sequences. The TIR domains from the *S. purpuratus* (red branches) and *L. variegatus* (blue branches) TLRs were used to construct a neighbor-joining tree in MEGA 5.0 (Tamura et al., 2011) using Poisson corrected distances. Alignment positions containing gaps were eliminated completely from the analysis. Bootstrap values are shown and are based on 1000 replicates. Each of these clades is also reconstructed using other methods (maximum likelihood and maximum parsimony; data not shown). The subfamily for each clade is designated on the right. A more detailed version of this tree is shown in **Figure A1** in Appendix.

arrays. The larger scaffolds and higher quality sequence in the current assembly enable the identification of these genes.

Additionally, each of the TLR sequences was classified as a complete gene, pseudogene, or partial gene based on the presence of in frame stop codons, and the presence of ambiguous flanking sequence. Given the complexity of this gene family and the similarity among the sequences, it is not surprising that many of the TLR genes are partial due to difficulty in assembling very similar sequence. Overall, 23% of the 253 TLR TIR domains were from partial gene sequences (**Table 3**). Pseudogenes are identified as those with in frame stop codons or frame shifts that result in missense sequence. Most of the frame shifts and point mutations that were used to designated pseudogenes (80%) could be confirmed by analysis of the genomic trace sequences and chromatographs. However, a few genes that appeared to be pseudogenes in the assembly were shown to be intact genes when the traces were analyzed more carefully (this includes the single group XI gene). Some of the pseudogenes are very similar to complete genes, while others differ substantially in sequence. The proportion of pseudogenes varies among groups (**Table 3**; Hibino et al., 2006), which is likely a function of varying turnover rates across the subfamilies. In this analysis, we only included sequences that encode intact TIR domains. Thus, this assessment of pseudogenes is incomplete, and many other related sequences that appear to be pseudogenes are present in the genome, varying from almost intact genes to highly divergent sequence fragments.

Although the TIR domains from all the TLRs can be aligned, the LRR portions of these proteins are unalignable across subfamilies. The orthology of individual LRRs cannot be reliably established across groups due to the variation in the number of LRRs and the lack of sequence similarity. Despite the sequence diversity among groups, the ectodomains of TLRs within groups are similar, both with respect to sequence and also the number of LRRs (**Figure 1**). The exceptions to this are TLRs in subfamilies Ib and Ie, which, although they are similar in sequence, vary in the number of LRRs as a result of discrete deletions or insertions of one or more complete LRRs.

The overall evolution of these groups is difficult to determine. Although each of the subfamilies consistently forms a clade, there is little support for the deeper relationships between the groups. It is notable, however, that the sea urchin sccTLR sequences appear to be the result of an expansion specific to the echinoderm lineage. When TLRs from mammals, other invertebrate deuterostomes, including hemichordates, urochordates, or cephalochordates, or protostomes are included in the analysis, the sea urchin sccTLRs form a strongly supported clade, but support for inter-phyla relationships is not present (data not shown; Messier-Solek et al., 2010).

SEQUENCE DIVERSITY AND SIGNATURES OF SELECTION WITHIN THE TLR SUBFAMILIES

The sea urchin TLR sequences exhibit striking amino acid diversity. There is significant variability within the conserved leucine-rich repeat framework, both with respect to changes in the amino acid sequence and also short indels. To characterize this diversity we analyzed the sequence entropy of each alignment position for the subfamilies that contained eight or more complete sequences

excluding pseudogenes (Ia, Ib, Ic, IIa, III, and IV; **Figure 3**). Sequence entropy is a measure of diversity that is based on the frequency of each amino acid at each position (Durbin et al., 1998). Results indicate that within subfamilies, the TIR domains are much more conserved than the LRR-containing ectodomains. On average, the ectodomain diversity is three times higher than that of the intracellular TIR domain (**Figure 3**; **Table 2**). This is consistent with an association between LRR sequence diversity and ligand-binding function. Furthermore, the levels and patterns of diversity vary among the subfamilies. The average diversity of the Ia sequences was over three times that of the Ic sequences, although both groups are composed of a similar numbers of genes (**Figure 3**). The peak in LRR diversity also varied among subfamilies. In subfamily Ia, the most diverse region of the ectodomains is in LRR16-18, whereas in subfamily IIa, the highest diversity is observed in LRR3 and LRR14. This variation in sequence diversity may reflect differences in ligand-binding mechanisms among the TLR subfamilies.

We further analyzed the patterns of selection within the *S. purpuratus* TLRs. Sequence entropy measures the diversity of the amino acids sequences, whereas the selection analyses within PAML take into account the underlying relative frequencies of synonymous and non-synonymous nucleotide substitutions that result in the protein sequence variability. The evolution of the sea urchin TLR sequences was analyzed under two models implemented in PAML that were compared using a likelihood ratio test (Yang, 2007). The first model, M7, allows codons to evolve under only neutral and purifying selection, whereas the second model, M8, also includes a class for residues that evolve under positive selection. For each of the six subfamilies analyzed, the M8 model that incorporated positive selection was a significantly better fit to the data (**Table 2**), suggesting that at least some of the residues within the TLR genes are subject to positive selection (Yang, 1998).

Specific sites that are likely to be under positive selection were identified and mapped onto a generic structure for the LRR ectodomain that is based on a simple solenoid model (**Figure 4**). The subfamilies varied in the number and pattern of specific residues under positive selection. Of the 170 total residues likely to be under positive selection from the six TLR subfamilies analyzed, 156 fell within the typical LRRs of the ectodomain, which is a significant enrichment compared to the more conserved TIR domains (**Table 2**; **Figure 4A**). Only one of these residues included a conserved amino acids that form the LRR framework (subfamily Ic, LRR15). Two residues were located in the TIR domain (both in subfamily Ib) and the remaining 10 sites were within either the LRR-NT or LRR-CT domains. The TLRs of families Ia and Ib had the greatest number of sites under positive selection (51 and 53, respectively; **Table 2**). This is in contrast to subfamily IIa, in which no specific residues were identified as significantly likely to be under positive selection.

Notably, the sites under diversifying selection are highly clustered on the three dimensional interpretation of the ectodomain structure (**Figure 4**). In subfamily Ia, the vast majority of the sites were located within the β-strands that form the concave face of the solenoid ectodomain (red dots; **Figure 4B**). In contrast, the positively selected residues of subfamily Ic are more scattered throughout the ectodomain (**Figure 4A**). Subfamilies Ia and IV

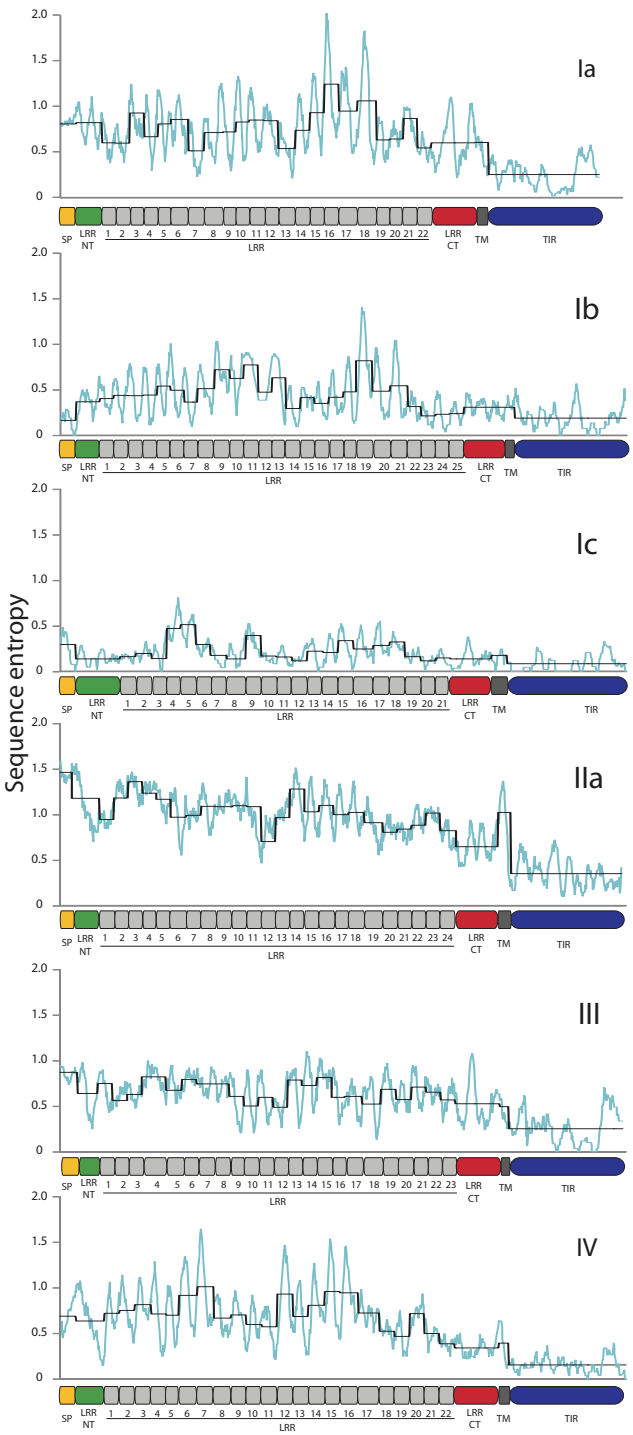


FIGURE 3 | TLR TIR domains are more conserved than the LRR regions. The diversity of the amino acid sequences for each of the subfamilies that contain more than eight complete sequences was analyzed as a measure of sequence entropy (Durbin et al., 1998). In the graphs shown, the light blue line indicates the average diversity over a sliding window of 10 amino acids, and the black line shows the average diversity of each of the protein domains marked on the *x*-axis. In each of the subfamilies, the TIR domains exhibit greater conservation than the ectodomains, and there is significant sequence variation within the LRR domains.

Table 2 | Diversity and evolution of the TLR sequences.

Group	No. of Seq[1]	lnL		−2lnΔL	Sites likely under positive selection		Entropy[2]	
		M7	M8		Total	Codon positions[3]	ECD	TIR
Ia	19	−21206.2	−20866.9	678.7*	51	55 64 **81 105 126 129 131 153 154 158 175 177 178 180 196 205 206 207 229 230 254 279 283 284 310 311 313 335 414 416 417 419 438 441 443 463 464 465 468 492 494 495 496 502 522 524 532 556 558** 604 606	0.70	0.23
Ib	9	−10933.8	−10749.8	368.0*	53	18 53 **74 75 98 99 118 120 124 148 149 150 172 173 189 191 193 217 243 244 256 267 269 270 271 280 296 392 416 418 440 445 464 496 498 516 517 518 520 524 538 540 542 543 544 574 575 576 579 624** 699 *781 848*	0.41	0.17
Ic	22	−8888.6	−8855.4	66.4*	24	5 11 37 90 **102 148 151 165 167 194 208 210 212 213 267 274 291 303 314 321 421 425 426 442**	0.21	0.08
IIa	8	−16706.3	−16695.2	22.2*	0	n/a	0.91	0.32
III	10	−13414.9	−13377.4	74.9*	13	**118 286 287 310 332 335 405 410 434 457 535 585 611**	0.59	0.23
IV	9	−12938.6	−12834.3	208.5*	29	42 **82 84 107 154 157 181 183 204 205 231 232 283 387 388 390 409 412 413 436 440 441 442 462 463 466 467 503 588**	0.63	0.14

*$p < 0.005$.

[1] *Includes only complete, non-pseudogenes.*

[2] *Average sequence entropy of all residues within the ectodomain (ECD) or TIR domain.*

[3] *Codon positions refer to those in Files S1–S6 in Supplementary Material. The domain structure of the TLRs is a signal peptide, LRR-NT, LRRs, LRR-CT, transmembrane region, and the TIR domain (see **Figure 1**). Residues shown in bold are located within the LRRs. Underlined residues are located in either the LRR-NT or LRR-CT. Residues shown in italics are located within the TIR domain.*

contain two distinct clusters of residues under positive selection (LRR1-11/LRR14-21 in subfamily Ia and LRR1-9/LRR13-17 for subfamily IV; **Figure 4A**). In general, the LRRs with greater positive selection also correspond to the more diverse LRRs shown in **Figure 3** although these analyses measure different elements of sequence diversity.

TLR EXPRESSION

The expression levels of the TLR subfamilies were analyzed in sea urchin larvae and adult immune cells and gut tissue using an RNA-Seq approach (**Figure 5**). A single batch of sea urchin larvae (9 dpf) was exposed to the marine bacterium *V. diazotrophicus*, and samples were collected at 0, 6, 12, and 24 h. For each time point, ∼75 million paired-end SOLiD sequencing reads were obtained. Additionally, an adult sea urchin was challenged using bacteria isolated from the digestive tract of another animal to mimic a perforation in the gut and systemic infection. This is intended as a physiologically relevant immune challenge that may be expected to induce a coordinated and complex immune response. Adult phagocytic coelomocytes and gut were isolated 12 h after challenge and used in RNA-Seq experiments, with approximately 130 million and 70 million paired-end reads obtained for each tissue, respectively. From this animal, ∼40 million phagocytes were collected, from which 1.5 μg of polyadenylated mRNA was isolated and used to generate cDNA for sequencing.

Using RNA-Seq data to analyze the expression of genes from multigene families is not trivial. Standard protocols from RNA-Seq analysis require sequence reads to map uniquely to a reference genome. However, this prevents reads from mapping to closely related paralogs and may artificially lower the expression values for these types of gene. Furthermore, most high-throughput sequence mapping programs are designed for use with the genomes of inbred organisms. Given the similarity of the TLR genes within subfamilies and the relatively high polymorphism among sea urchins (estimated genome heterozygosity is 4–5%; Britten et al., 1978; Pespeni et al., 2011), we have relaxed the stringency of the mapping parameters to analyze the expression of the TLRs. Reads were allowed to map to the genome up to 50 times to accommodate a single read mapping to multiple TLR paralogs, which is slightly larger than the biggest subfamily (Ia, which has 48 sequences). Including only uniquely mapping reads in the analysis disproportionately reduces the expression of the larger subfamilies with closely related genes. Therefore, while we are unable to assess the transcript prevalence of any particular gene relative to its subfamily counterparts, we are able to quantify collective subfamily expression. To clearly assign reads to specific subfamilies, reads that mapped to TLRs from multiple groups, were removed from the analysis (this represented 771 of 20,332 total reads; 3.7%). To account for the high heterozygosity of the sea urchin genome and the expected genetic differences between the experimental animal and that used for the reference genome, we also increased the number of mismatches. These relaxed parameters, however, did not result in a high background of spurious read mapping. Reads that mapped to TLR sequences were directionally specific (<2% of reads mapped in the incorrect orientation for genes that stood above background), which lends additional confidence to our measurements.

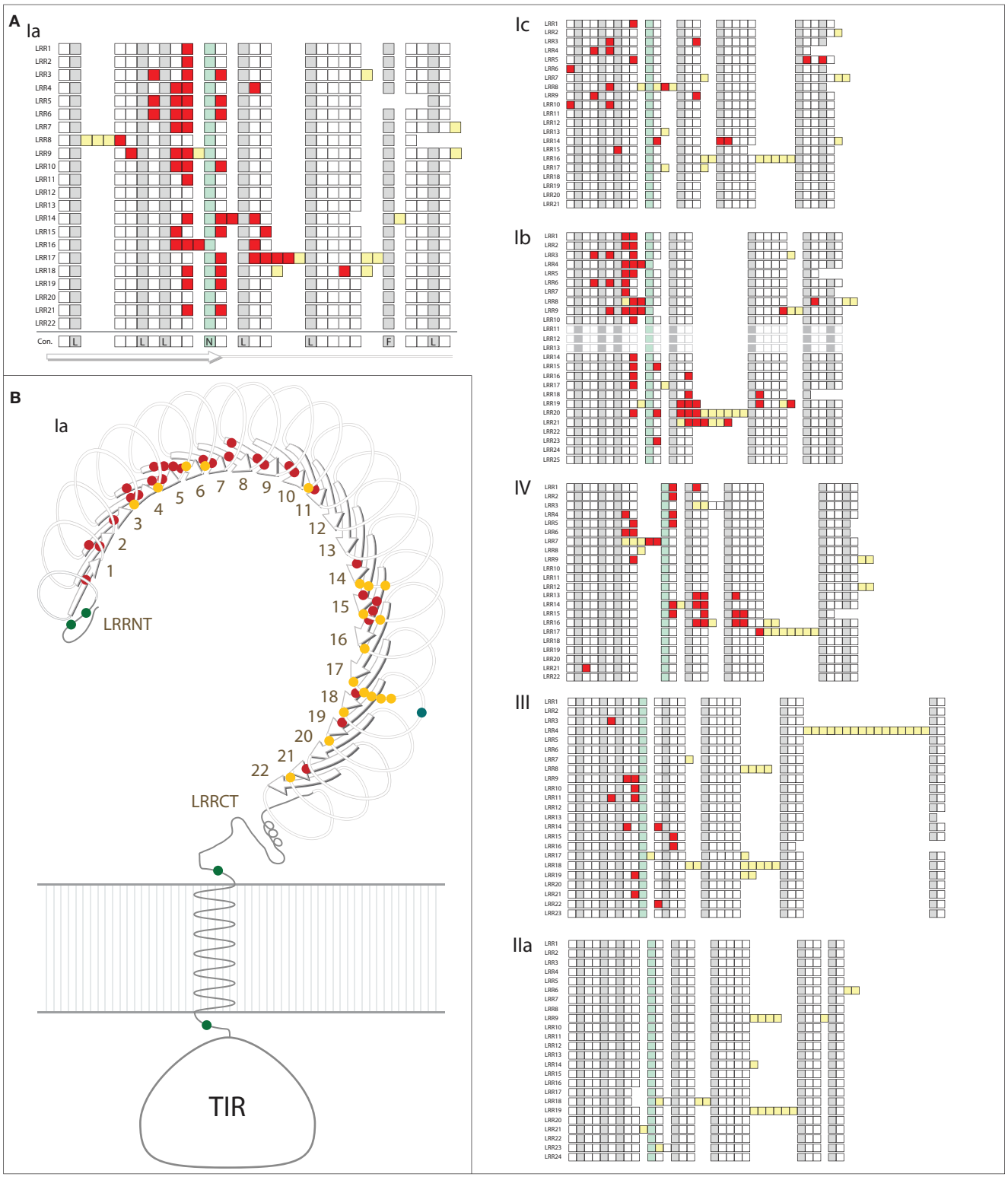


FIGURE 4 | Analysis of positive selection within TLR subfamilies. (A) The majority of residues subject to positive selection are located in the putative convex face of the ectodomain. Residues under positive selection were identified in each subfamily containing more than eight complete, non-pseudogene sequences using PAML (Ia, Ib, Ic, IIa, III, and IV; Yang, 2007). Sequence alignments can be found in Files S1-S6 in Supplementary Material. Each of the LRRs is shown as a single row with individual amino acids indicated as squares. Consensus hydrophobic LRR residues are shown in gray and the conserved asparagines residues are indicated in

(Continued)

FIGURE 4 | Continued
green (Bell et al., 2003; Kang and Lee, 2011). Amino acids under positive selection are shown in red; insertions into the LRR framework are indicated by yellow squares. The predicted consensus structure is indicated below the squares: the first 10 amino acids form the β-strands on the convex surface of the ectodomains, while the rest of the LRR forms the loop structures. In subfamily IIa, no individual residues were found to be likely under positive selection. **(B)** Positive selection within the Ia subfamily. A generalized TLR structure is shown based on known solenoid structures of LRR-containing proteins. The majority of positively selected positions are located within the β-strands (red dots), and a few residues are also located on the front face of the TLR just beyond the β-strands (yellow dots). One positively selected residue is found on the outside of the LRRs (blue dot), and there are four residues under positive selection that are not located within the LRRs (two in the LRR-NT domain, and one on either side of the transmembrane domain; green dots).

Gene expression levels are measured as the number of reads that mapped to the gene per kilobase per million reads mapped to the genome (RPKM; Mortazavi et al., 2008). RPKM is a standard measure of gene expression used in high-throughput sequence analysis that takes into account the length of the gene (longer transcripts produce more sequence fragments), and the total size and quality of the library (poor quality libraries produce reads that do not map to the genome and there is always variation in the number of fragments that are sequenced). These values are comparable across samples and time points. Given the similarity of the TLRs, and the possibility for a read to map to multiple subfamily members, we present the data as the average RPKM for each subfamily (**Figure 5**).

In sea urchin larvae, the RPKM values of the TLR subfamilies are generally low relative to the expression values of the adult tissues (**Figure 5A**). However, because the whole animal was used in the sequencing, it may be expected that the TLRs are expressed at relatively higher levels in a small subset of cells within the animal. The highest expression levels are observed for TLR subfamily Id, group VI and the single sequence of group X. This is consistent with qPCR measurements of transcript prevalence (data not shown). Furthermore, while the larvae are able to mount a robust and complex antibacterial response, only modest change is observed in TLR expression in response to bacterial challenge. Many of the TLR subfamilies that are prevalent in adult tissues were not evident in this ontogenetic stage.

We also analyzed RNA-Seq data from phagocytic coelomocytes collected from an immune-challenged animal. There are four primary classes of adult sea urchin coelomocytes: phagocytes, vibratile, colorless spherule cells, and red spherule cells (Smith et al., 2011). Preliminary qPCR data suggests that TLR expression is minimal in the vibratile, colorless spherule, and red spherule cell fractions relative to the phagocytes (data not shown). Analysis of RNA-Seq data from immune-challenged phagocytes indicates that several immune-related genes are expressed in these cells, suggesting that the animal is responding to immune challenge. In contrast to the larval stage, adult coelomocytes express a different suite of TLR subfamilies (**Figure 5B**). Genes from subfamilies Ib and IIa are expressed highly, as is the TLR gene that comprises the X group, and, notably, the mccTLR genes. Compared to the RPKM values from the larval stage, the average expression levels for the TLR subfamilies in coelomocytes are 20 times higher. Although there may be specific cells within the larvae that express the TLRs at higher levels, this enrichment of RNA-Seq reads that map to the TLR genes in the coelomocyte sample further suggests a role of these proteins in immunity.

Gut tissue was also collected from the immune-challenged animal described above and analyzed by RNA-Seq. The adult gut expresses a suite of TLR genes that is distinct from those expressed in adult coelomocytes and the larval stage (**Figure 5C**). Here, TLRs from subfamilies III and Ia, as well as the divergent, intron-containing TLRs exhibit the highest levels of expression. Expression of the group X and mccTLRs, which are highly expressed by coelomocytes, was not observed in the gut. These varied expression patterns may point to different roles in the TLR subfamilies in different tissues and at different life stages.

TLRs IN OTHER SEA URCHIN SPECIES

As part of the Sea Urchin Genome Project, genome and transcriptome sequencing is underway in several other echinoderm species. A genome sequence is available from *L. variegatus*, which last shared a common ancestor with *S. purpuratus* about 50 million years ago (Smith et al., 2006). We analyzed the *L. variegatus* genome using the same methods as the *S. purpuratus* genome and identified 68 TLR genes, either as ORFs that contained a TIR domain in addition to the presence of a transmembrane region, LRR-CT domain, and LRRs or by sequence similarity to *S. purpuratus* TLRs. The TIR domain sequences from these genes were used in phylogenetic analysis with the *S. purpuratus* TLR TIR domains to classify the *L. variegatus* genes into subfamilies (**Figure 2**). The majority of the *L. variegatus* TLR genes group with the *S. purpuratus* sccTLR subfamilies (60 of 68). Notably, the *L. variegatus* genome also encodes orthologs of the intron-containing, mccTLR and short TLR sequences. Although homologous representatives of most of the *S. purpuratus* sccTLR subfamilies were present in the *L. variegatus* genome, no sequences were identified that were orthologous to subfamilies Ib, Ic, Id, Ie, Ig or group VII genes (**Table 3**). In addition to the 10 *S. purpuratus* sccTLR subfamilies, an additional subfamily, IIIa, is found only within *L. variegatus* (**Figure 2**). This strongly supported clade consists of eight sequences and is sister to the *S. purpuratus* group III sequences. Notably, there are three pairs of orthologous sequences with a single representative in each sea urchin species, the phylogenetic stability of which may suggest a more conserved ligand-binding function relative to the other subfamilies of higher diversity and multiplicity (see groups IX, X, and XI; **Figure 2**). One of these, subfamily X, which has a single representative in both sea urchin species, is highly expressed in both the *S. purpuratus* larva and coelomocytes (**Figure 5**).

To estimate gene copy number of the TLRs in *L. variegatus* in an assembly independent manner, we analyzed the unassembled genomic trace sequences. A conserved region of each of the TIR domains of the 253 *S. purpuratus* and the 68 *L. variegatus* TLR sequences was used as a query in a BLAST search against the sequences that were used to assemble the *L. variegatus* genome. In total, 1054 unique sequences were recovered with similarity to the

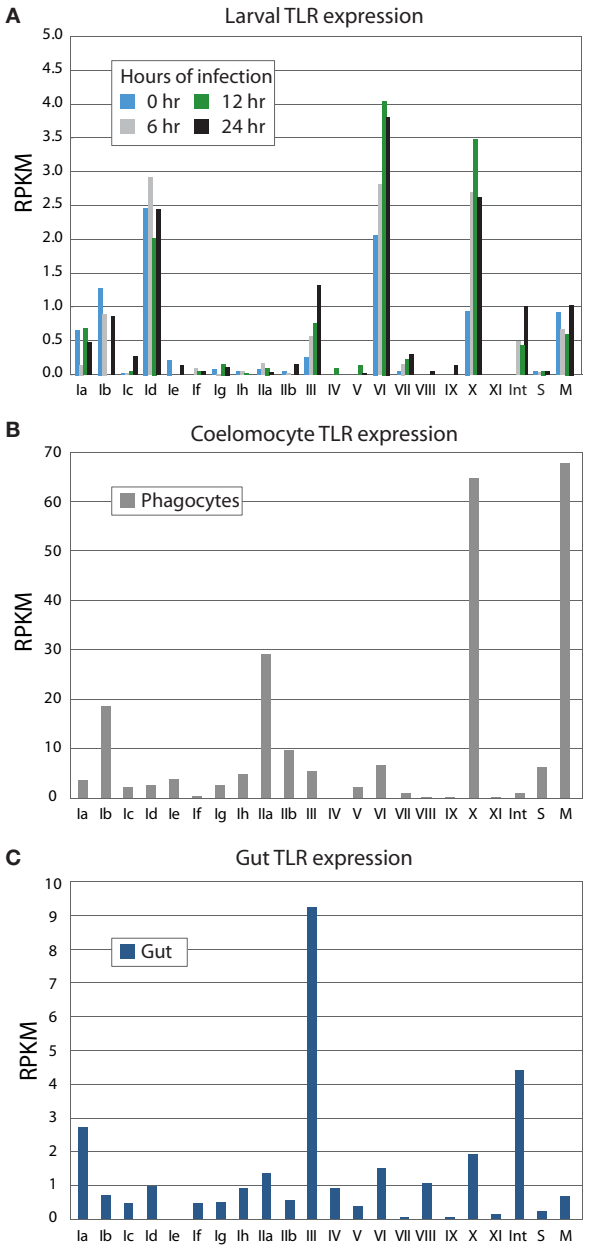


FIGURE 5 | TLR subfamilies are differentially expressed. RNA-Seq was used to analyze gene expression from larvae **(A)**, immune cells **(B)**, and gut **(C)**. Phagocytic coelomocytes and gut were isolated from an animal 12 h after intracoelomic injection of a gut bacteria preparation. Larvae were exposed to *V. diazotrophicus* for 0, 6, 12, and 24 h, and collected for transcriptome sequencing. RPKM values were calculated for each of the TIR domains; the average RPKM for each group is shown. The sccTLR subfamilies correspond to those shown in **Figure 2**. Note that the scales are different for each graph. The divergent TLRs are indicated as follows: M, mccTLRs; S, short; Int, intron-containing (**Figure 1**). In phagocytic coelomocytes, TLRs from the Ib, IIa, and X subfamilies, as well as the mccTLR genes are most highly expressed. In contrast, the III TLRs are primarily expressed in the gut tissue. The TLRs are expressed at lower levels in the larvae, and little change in TLR expression in the larvae is observed in response to bacterial challenge. The predominantly expressed families at the larval stage are Id and VI, are different than those expressed in coelomocytes.

Table 3 | Sizes of sea urchin TLR subfamilies.

Group	*S. purpuratus*				*L. variegatus*
	Total	Complete	Partial	Pseudo[1] (%)	
Ia	48	19	13	16 (33)	12
Ib	16	9	4	3 (19)	0
Ic	37	22	3	12 (32)	0
Id	12	3	4	5 (42)	0
Ie	6	3	1	2 (33)	1
If	3	2	0	1 (33)	5
Ig	7	3	3	1 (14)	0
Ih	3	2	1	0 (0)	1
I orphan	1	1	0	0 (0)	0
IIa	20	8	6	6 (30)	8
IIb	13	6	3	4 (31)	7
III	29	11	11	7 (24)	2
IIIa	0	0	0	0 (0)	10
IV	13	9	1	3 (23)	2
V	8	2	3	3 (38)	0
VI	8	5	2	1 (13)	8
VII	9	6	1	2 (22)	0
VIII	4	4	0	0 (0)	1
IX	1	1	0	0 (0)	1
X	1	1	0	0 (0)	1
XI	1	1	0	0 (0)	1
Intron	4	2	2	0 (0)	1
Short	5	5	0	0 (0)	4
mccTLR	4	3	1	0	3
Total	253	127	59	67 (26)	68

[1] *Includes only pseudogenes that encode intact TIR domains. Actual numbers are higher.*

sea urchin TLRs. Given 19.5× coverage (the 48,120,406 reads had an average length of 340 nt, and the *L. variegatus* genome is estimated to be 840 MB; Hinegardner, 1974), this indicates that there are 54 TLR TIR domains (Table S3 in Supplementary Material). This is slightly lower than the number of TLR sequences within the assembled genome, which may reflect the presence of allelic copies retained in the assembly. This data is consistent with a *L. variegatus* TLR gene family that is smaller than that of *S. purpuratus*, but still expanded relative to vertebrate TLR families.

Low coverage 454 reads are also available for the genomes of two sea urchin species closely related to *S. purpuratus*: *A. fragilis*, and *S. franciscanus*. Despite the nomenclature, *A. fragilis* is most closely related to *S. purpuratus*, with an estimated divergence time of 5–7 million years ago. These sister species shared a common ancestor with *S. franciscanus* about 20 million years ago (Biermann et al., 2003; Lee, 2003). These low coverage reads (∼2×) are insufficient to assemble a complete genome sequence, but allow us to estimate the multiplicity of the TLR gene family. The sequences from both species are an average of ∼235 nt in length, which is shorter than most TIR domains (the average size of the TLR TIR domains from *S. purpuratus* and *L. variegatus* is 354 nt). To simplify the analysis and to avoid sequences matching to partial TIR domains, we extracted a conserved region of the *S. purpuratus* and

L. variegatus TLR TIR domains (50 amino acids) to use as queries in a BLAST search against the unassembled traces from *A. fragilis* and *S. franciscanus*. Positive reads were isolated and used as queries in a BLAST search against the whole TIR domains from *S. purpuratus* and *L. variegatus* to classify the reads by subfamily (Table S3 in Supplementary Material).

In total, 580 unique reads from *A. fragilis* and 524 reads from *S. franciscanus* were identified that exhibited similarity to the sea urchin TLR domains, which indicates that the TLR gene families in these species consist of 276 and 228 sequences, respectively (Table S3 in Supplementary Material). The distribution of TLRs among subfamilies is consistent with that in *S. purpuratus*. There is a reduced number of Ic TLRs in *A. fragilis* and *S. franciscanus* (4 and 16, respectively, compared to 37 in *S. purpuratus*), suggesting that these highly similar and genomically clustered genes may be the product of a very recent expansion in *S. purpuratus*. Similarly, there is an enrichment in the number of group III TLRs within the *A. fragilis* genome, which is estimated to have 61 TLR sequences, as compared to 29 in *S. purpuratus* and 26 in *S. franciscanus*. There are also homologs of each of the groups that contain a single representative in *S. purpuratus* and *L. variegatus* (groups IX, X, and XI), which may point to a conserved function for these receptors. Homologs of each of the divergent *S. purpuratus* subfamilies are present in both *A. fragilis* and *S. franciscanus*, including the mccTLRs. None of the sequences showed similarity to the group IIIa sequences, which appear to be unique to the *L. variegatus* lineage. As genome sequences become available for additional sea urchin species, as well as other echinoderms, our understanding of the evolution of this complex gene family will be further resolved.

DISCUSSION

The sizes of the gene families that encode TLRs vary substantially among metazoan species (**Table 1**). Although sea urchin genomes encode the largest of these families, significant expansions have also occurred in the genomes of amphioxus, which has 72 TLR genes (Huang et al., 2008), and the annelid *C. capitata*, which encodes 105 (Davidson et al., 2008) as well as other invertebrate species that are now being sequenced. Each of these expansions generates a unique suite of TLRs that are not generally orthologous to TLRs in other species. This is not surprising, however, given the apparent rapid turnover of these genes, as suggested by the proportion of pseudogenes and high similarity of some family members. This pattern of species-restricted paralogy is consistent with that seen for other immune multigene families. In all cases where these genes are present as highly expanded multigene families, both in protostomes and deuterostomes, it is the vertebrate-like sccTLRs that are amplified (**Table 1**).

Although not present in vertebrates, the prototypic Toll-like mccTLR type can be identified in all eumetazoan phyla for which representative genome sequences are available, including the lower chordates. Usually this TLR type is present in single-copy or as very small gene families although moderate expansion is evident in a few species (**Table 1**). The presence of the mccTLR type as the only TLR gene in a basal eumetazoan, the cnidarian *N. vectensis* (Miller et al., 2007), as well as in all protostomes and invertebrate deuterostomes suggests that the mccTLR was a primitive component of eumetazoan genomes and that this receptor was lost in the vertebrate lineage. It is notable, however, that this type of receptor is always present in low numbers even in the species with expanded sccTLR gene families. In the sea urchin, mccTLRs are expressed at high levels in activated coelomocytes (**Figure 5**), which is consistent with an immune-related function.

The members of TLR multigene families in the sea urchin are characterized by apparently rapid sequence divergence within the ectodomain and conservation within the TIR region (**Figure 3**). This could be explained either by a lack of constraint in the diverging LRRs or by a more active process of diversifying selection. Our analysis suggests that positive selection plays a role in the diversification process and that it does so in spatially restricted regions of the TLR structure. Nearly all residues that are likely under positive selection are located in the LRRs, mainly in the concave region that is formed by the LRR β-strands. Almost no selection is indicated for residues within the TIR domain. This is consistent with observations in *Drosophila* immune genes, in which proteins involved in immunity, particularly those involved in pathogen recognition, were shown to have a higher proportion of residues under positive selection as compared to non-immune proteins (Sackton et al., 2007). The pathogen-interacting domains of phagocytosis receptors and two peptidoglycan recognition proteins were particularly enriched in codons likely to be under positive selection with respect to the remainder of the proteins (Sackton et al., 2007). This is also the case for many TLRs in analyses of positive selection carried out on the vertebrate sequences (Wlasiuk and Nachman, 2010; Alcaide and Edwards, 2011; Areal et al., 2011; Tschirren et al., 2011). When signatures of positive selection are detected in the vertebrate TLRs with known ligand-binding structure, it tends to be in regions that are known to interact with non-self and in regions that mediate dimerization. Thus the residues likely to be under positive selection in the sea urchin may also correspond to regions that interact with non-self. Notably, not all groups analyzed showed evidence of specific residues under positive selection and, in those that did, there was variation in the pattern of these residues. This may reflect different mechanisms of function within the subfamilies.

Multiplicity and patterns of incremental diversification among members of the major sea urchin sccTLR subfamilies in the ectodomain imply a direct form of ligand recognition, although some of the smaller, more conserved TLR gene families may be specialized to function differently. The sea urchin TLR genes may operate by recognizing non-self molecules that are similar to those recognized by vertebrate TLRs but with greater specialization. Alternatively, they may have evolved to recognize entirely different classes of molecules. The latter possibility is suggested by the spatial distribution and extent of diversity, which is unlike that seen among vertebrate TLRs. Given their multiplicity, the increased variation in LRR regions, the signature of positive selection in the portion of the genes encoding the ectodomain and the range of variation from near identity to high divergence, the sea urchin TLR genes appear to be evolving in response to a changing array of binding requirements.

One problem in analyzing the large families of sea urchin TLRs in the past has been the inability to find any level of orthology among subfamilies in inter-phyla comparisons (Roach et al., 2005; Hibino et al., 2006). The exception to this is the mccTLR-sccTLR

division which shows a weak signal of orthology even between sea urchin and *Drosophila* genes (Hibino et al., 2006; Messier-Solek et al., 2010). The introduction of a second sea urchin genome into this analysis lends considerable insight into this issue. Phylogenetic analysis of the combined sea urchin TLR genes reveals cases of relative conservation in terms of gene number and cases of species-specific expansion or reduction. This can be used as an indicator of which genes may have unique and necessary functions and which genes may have interrelated, evolutionarily labile functions. At the extremes, some groups are encoded in single-copy in *L. variegatus* and greatly expanded in *S. purpuratus* (for example the group III genes) while others, such as the group X and mccTLR genes, appear to be more phylogenetically stable with single copies present in the genomes of both species. It is not clear whether the difference in the sizes of the gene families in these species is the result of an expansion in the TLR gene family within the strongylocentrotid lineage or gene loss *L. variegatus*. As more sea urchin genome sequences from outgroups to this clade are completed, this will become better resolved.

The question of whether or not the sea urchin TLRs are non-self receptors remains open but circumstantial evidence is consistent with an immune function for many of the subfamilies. This includes the following observations. (1) While expression of TLRs is generally low, for some of the largest subfamilies, transcription is greatly enhanced in phagocytic coelomocytes, many-fold over other tissues. (2) Expression of TLRs is not detectable in the embryo when primary developmental processes are unfolding but is initiated in the feeding larva coincident with the transcriptional activation of a suite of immune genes. (3) Multiplicity, variability, and sequence signatures of positive selection are common features of immune multigene families. (4) Finally, while the majority of *Drosophila* TLRs have not been associated with immunity but are associated with other biological processes (Narbonne-Reveau et al., 2011), all of the mammalian TLRs function as direct immune recognition receptors. The sea urchin is more closely related to vertebrates and the sea urchin TLRs resemble the vertebrate TLRs more closely than they do the *Drosophila* TLRs with known non-immune functions.

Of course the identification of the ligands for the sea urchin TLRs would answer this question definitively but this is a difficult technical challenge especially if, as may well be the case, the ligands are non-self and diverse. A more tractable path to understand the function of these receptors may be to focus on some of the smaller families which can be experimentally targeted but are nonetheless closely enough related to the expanded subfamilies to imply a similar function. Phylogenetic analysis of TLRs among sea urchins reveals some small TLR subfamilies that fit this pattern and comparative work in species like *L. variegatus* with relatively smaller TLR families will also be useful.

Whatever the exact biological roles of the large TLR gene families, it is probable that the sea urchin has co-opted this well known receptor to a new variation of function that is more evolutionarily labile than what has been well described in the vertebrates. Some of this reassignment may have taken place within the sea urchins as suggested by species-specific expansions. Nonetheless recent and emerging genome sequences from across the bilaterians indicate that large TLR repertoires may be widespread. It remains to be seen whether these expansions share a common functional purpose or whether they are each the result of a unique reaction to specific evolutionary pressures. While much of the justification for turning to as yet unstudied animal phyla is focused on aspects of host defense that are shared with mammals, in the long run comparative approaches will make a much richer contribution by revealing what is novel across animal immunity.

ACKNOWLEDGMENTS

We thank Eric Ho for developing the infection strategy used in the larval RNA-Seq measurements and Yutaka Amemiya and Arun Seth of the Sunnybrook Research Institute Genomics Core Facility. We also thank the reviewers for many helpful comments. This work is supported by grants from the Canadian Institutes for Health Research (MOP74667) and the Natural Sciences and Engineering Research Council of Canada (NSERC 312221) to Jonathan P. Rast.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/Molecular_Innate_Immunity/10.3389/fimmu.2012.00136/abstract

REFERENCES

Alcaide, M., and Edwards, S. V. (2011). Molecular evolution of the toll-like receptor multigene family in birds. *Mol. Biol. Evol.* 28, 1703–1715.

Areal, H., Abrantes, J., and Esteves, P. J. (2011). Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol. Biol.* 11, 368. doi:10.1186/1471-2148-11-368

Bell, J. K., Mullen, G. E., Leifer, C. A., Mazzoni, A., Davies, D. R., and Segal, D. M. (2003). Leucine-rich repeats and pathogen recognition in Toll-like receptors. *Trends Immunol.* 24, 528–533.

Biermann, C. H., Kessing, B. D., and Palumbi, S. R. (2003). Phylogeny and development of marine model species: strongylocentrotid sea urchins. *Evol. Dev.* 5, 360–371.

Bosch, T. C., Augustin, R., Anton-Erxleben, F., Fraune, S., Hemmrich, G., Zill, H., Rosenstiel, P., Jacobs, G., Schreiber, S., Leippe, M., Stanisak, M., Grötzinger, J., Jung, S., Podschun, R., Bartels, J., Harder, J., and Schröder, J. M. (2009). Uncovering the evolutionary history of innate immunity: the simple metazoan Hydra uses epithelial cells for host defence. *Dev. Comp. Immunol.* 33, 559–569.

Britten, R. J., Cetta, A., and Davidson, E. H. (1978). The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. *Cell* 15, 1175–1186.

Cameron, R. A., Samanta, M., Yuan, A., He, D., and Davidson, E. (2009). SpBase: the sea urchin genome database and web site. *Nucleic Acids Res.* 37, D750–D754.

Davidson, C. R., Best, N. M., Francis, J. W., Cooper, E. L., and Wood, T. C. (2008). Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). *Dev. Comp. Immunol.* 32, 608–612.

Durbin, R., Eddy, S., Krogh, A., and Mitchison, G. (1998). *Biological Sequence Analysis, Probability Models of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press.

Felsenstein, J. (2005). *PHYLIP (Phylogeny Inference Package) Version 3.6*. Seattle: University of Washington.

Gay, N. J., and Keith, F. J. (1991). Drosophila Toll and IL-1 receptor. *Nature* 351, 355–356.

Gross, P. S., Clow, L. A., and Smith, L. C. (2000). SpC3, the complement homologue from the purple sea urchin, *Strongylocentrotus purpuratus*, is expressed in two subpopulations of the phagocytic coelomocytes. *Immunogenetics* 51, 1034–1044.

Hall, T. A. (1999). BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser. (Oxf.)* 41, 95–98.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P., Buckley, K. M., Brockton, V., Nair, S. V., Berney, K., Fugmann, S. D., Anderson, M. K., Pancer, Z., Cameron, R. A., Smith, L. C., and Rast, J. P. (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* 300, 349–365.

Hinegardner, R. (1974). Cellular DNA content of the echinodermata. *Comp. Biochem. Physiol. B* 49, 219–226.

Holland, L. Z., Albalat, R., Azumi, K., Benito-Gutiérrez, E., Blow, M. J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L. J., Ferrier, D. E., Garcia-Fernàndez, J., Gibson-Brown, J. J., Gissi, C., Godzik, A., Hallböök, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G. W., McHardy, A. C., Meulemans, D., Nonaka, M., Olinski, R. P., Pancer, Z., Pennacchio, L. A., Pestarino, M., Rast, J. P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J. K., Zhang, Q., Zmasek, C. M., de Jong, P. J., Osoegawa, K., Putnam, N. H., Rokhsar, D. S., Satoh, N., and Holland, P. W. (2008). The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* 18, 1100–1111.

Huang, A. M., Rusch, J., and Levine, M. (1997). An anteroposterior dorsal gradient in the *Drosophila* embryo. *Genes Dev.* 11, 1963–1973.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., Ge, J., Yu, Y., Huang, H., Dong, M., Yu, C., Chen, S., and Xu, A. (2008). Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* 18, 1112–1126.

Jin, M. S., and Lee, J. O. (2008). Structures of the toll-like receptor family and its ligand complexes. *Immunity* 29, 182–191.

Kang, J. Y., and Lee, J. O. (2011). Structural biology of the Toll-like receptor family. *Annu. Rev. Biochem.* 80, 917–941.

Kasamatsu, J., Oshiumi, H., Matsumoto, M., Kasahara, M., and Seya, T. (2010). Phylogenetic and expression analysis of lamprey toll-like receptors. *Dev. Comp. Immunol.* 34, 855–865.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.

Lee, Y. H. (2003). Molecular phylogenies and divergence times of sea urchin species of Strongylocentrotidae, Echinoida. *Mol. Biol. Evol.* 20, 1211–1221.

Lemaitre, B., and Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annu. Rev. Immunol.* 25, 697–743.

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

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

Medzhitov, R., Preston-Hurlburt, P., and Janeway, C. A. Jr. (1997). A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* 388, 394–397.

Messier-Solek, C., Buckley, K. M., and Rast, J. P. (2010). Highly diversified innate receptor systems and new forms of animal immunity. *Semin. Immunol.* 22, 39–47.

Miller, D. J., Hemmrich, G., Ball, E. E., Hayward, D. C., Khalturin, K., Funayama, N., Agata, K., and Bosch, T. C. (2007). The innate immune repertoire in cnidaria – ancestral complexity and stochastic gene loss. *Genome Biol.* 8, R59.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.

Narbonne-Reveau, K., Charroux, B., and Royet, J. (2011). Lack of an Antibacterial response defect in *Drosophila* Toll-9 mutant. *PLoS ONE* 6, 17470. doi:10.1371/journal.pone.0017470

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

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

Ooi, J. Y., Yagi, Y., Hu, X., and Ip, Y. T. (2002). The *Drosophila* Toll-9 activates a constitutive antimicrobial defense. *EMBO Rep.* 3, 82–87.

Pespeni, M. H., Garfield, D. A., Manier, M. K., and Palumbi, S. R. (2011). Genome-wide polymorphisms show unexpected targets of natural selection. *Proc. Biol. Sci.* 279, 1412–1420.

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B. (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 282, 2085–2088.

Roach, J. C., Glusman, G., Rowen, L., Kaur, A., Purcell, M. K., Smith, K. D., Hood, L. E., and Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9577–9582.

Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A., and Bazan, J. F. (1998). A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl. Acad. Sci. U.S.A.* 95, 588–593.

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., and Clark, A. G. (2007). Dynamic evolution of the innate immune system in *Drosophila*. *Nat. Genet.* 39, 1461–1468.

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

Smith, A. B., Pisani, D., Mackenzie-Dodds, J. A., Stockley, B., Webster, B. L., and Littlewood, D. T. (2006). Testing the molecular clock: molecular and paleontological estimates of divergence times in the Echinoidea (Echinodermata). *Mol. Biol. Evol.* 23, 1832–1851.

Smith, L. C., Ghosh, J., Buckley, K. M., Clow, L. A., Dheilly, N. M., Haug, T., Henson, J. H., Li, C., Lun, C. M., Majeske, A. J., Matranga, V., Nair, S. V., Rast, J. P., Raftos, D. A., Roth, M., Sacchi, S., Schrankel, C. S., and Stensvåg, K. (2011). Echinoderm immunity. *Adv. Exp. Med. Biol.* 708, 260–301.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Ettensohn, C. A., Foltz, K. R., Hamdoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., McClay, D. R., Morris, R. L., Mushegian, A., Rast, J. P., Smith, L. C., Thorndyke, M. C., Vacquier, V. D., Wessel, G. M., Wray, G., Zhang, L., Elsik, C. G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M. J., Mackey, A. J., Maglott, D., Panopoulou, G., Poustka, A. J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C. A., Worley, K., Durbin, K. J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M. L., Jackson, A. R., Milosavljevic, A., Tong, M., Killian, C. E., Livingston, B. T., Wilt, F. H., Adams, N., Bellé, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Genevière, A. M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A. J., Goldstone, J. V., Cole, B., Epel, D., Gold, B., Hahn, M. E., Howard-Ashby, M., Scally, M., Stegeman, J. J., Allgood, E. L., Cool, J., Judkins, K. M., McCafferty, S. S., Musante, A. M., Obar, R. A., Rawson, A. P., Rossetti, B. J., Gibbons, I. R., Hoffman, M. P., Leone, A., Istrail, S., Materna, S. C., Samanta, M. P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K. F., Brandhorst, B. P., Whittle, J., Berney, K., Bottjer, D. J., Calestani, C., Peterson, K., Chow, E., Yuan, Q. A., Elhaik, E., Graur, D., Reese, J. T., Bosdet, I., Heesun, S., Marra, M. A., Schein, J., Anderson, M. K., Brockton, V., Buckley, K. M., Cohen, A. H., Fugmann, S. D., Hibino, T., Loza-Coll, M., Majeske, A. J., Messier, C., Nair, S. V., Pancer, Z., Terwilliger, D. P., Agca, C., Arboleda, E., Chen, N., Churcher, A. M., Hallböök, F., Humphrey, G. W., Idris, M. M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R. P., Raible, F., Rowe, M., Taylor, J. S., Tessmar-Raible, K., Wang, D., Wilson, K. H., Yaguchi, S., Gaasterland, T., Galindo, B. E., Gunaratne, H. J., Juliano, C., Kinukawa, M., Moy, G. W., Neill, A. T., Nomura, M., Raisch, M., Reade, A., Roux, M. M., Song, J. L., Su, Y. H., Townley, I. K., Voronina, E., Wong, J. L., Amore, G., Branno, M., Brown, E. R., Cavalieri,

V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Röttinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S. Y., Xu, R., Brown, C. T., Chen, L., Gray, R. F., Lee, P. Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L. V., Okwuonu, G., Parker, D., Pu, L. L., Thorn, R., and Wright, R. (2006). The genome of the sea urchin *Strongylocentrotus purpuratus. Science* 314, 941–952.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.

Tschirren, B., Raberg, L., and Westerdahl, H. (2011). Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. *J. Evol. Biol.* 24, 1232–1240.

Wlasiuk, G., and Nachman, M. W. (2010). Adaptation and constraint at Toll-like receptors in primates. *Mol. Biol. Evol.* 27, 2172–2186.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591.

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.

Received: 21 February 2012; accepted: 12 May 2012; published online: 05 June 2012.

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

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

APPENDIX

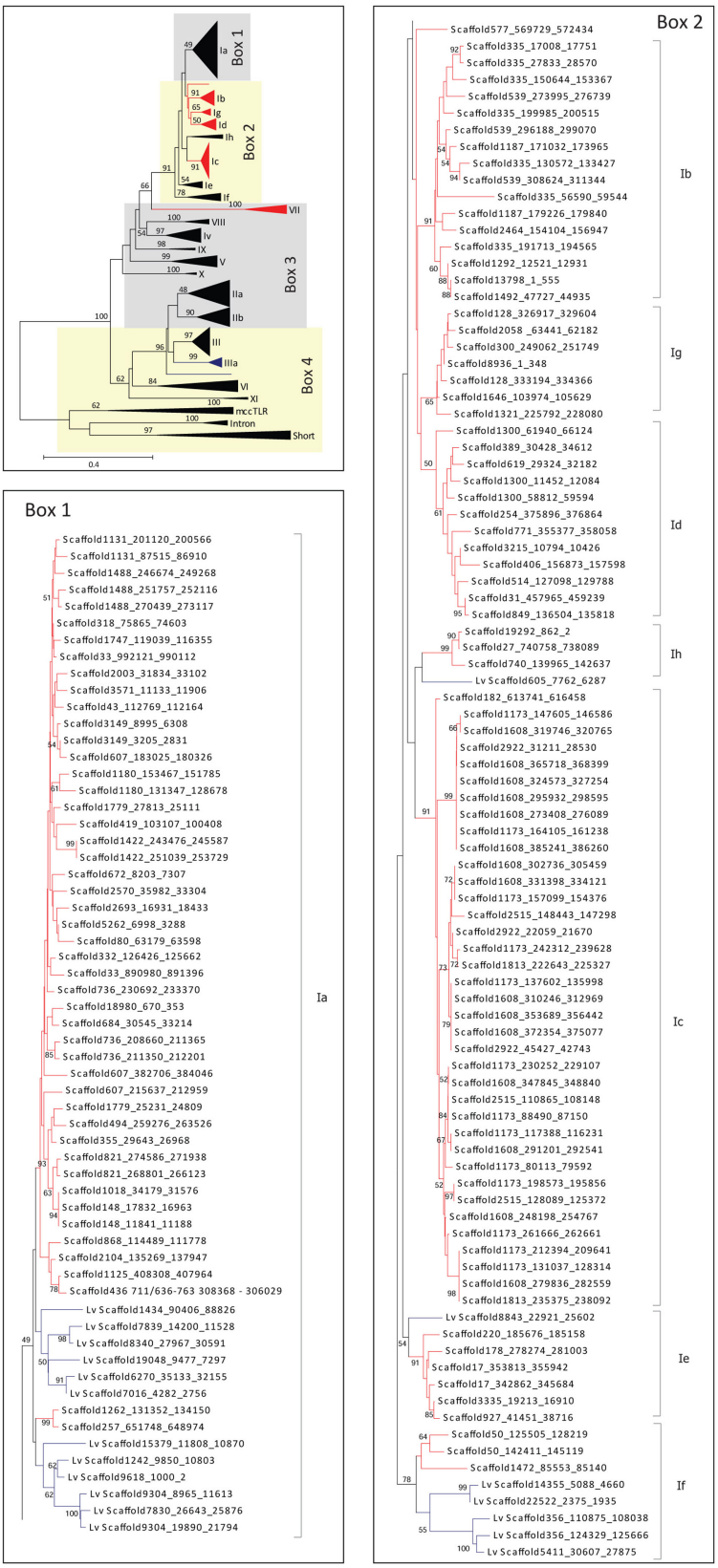


FIGURE A1 | Continued

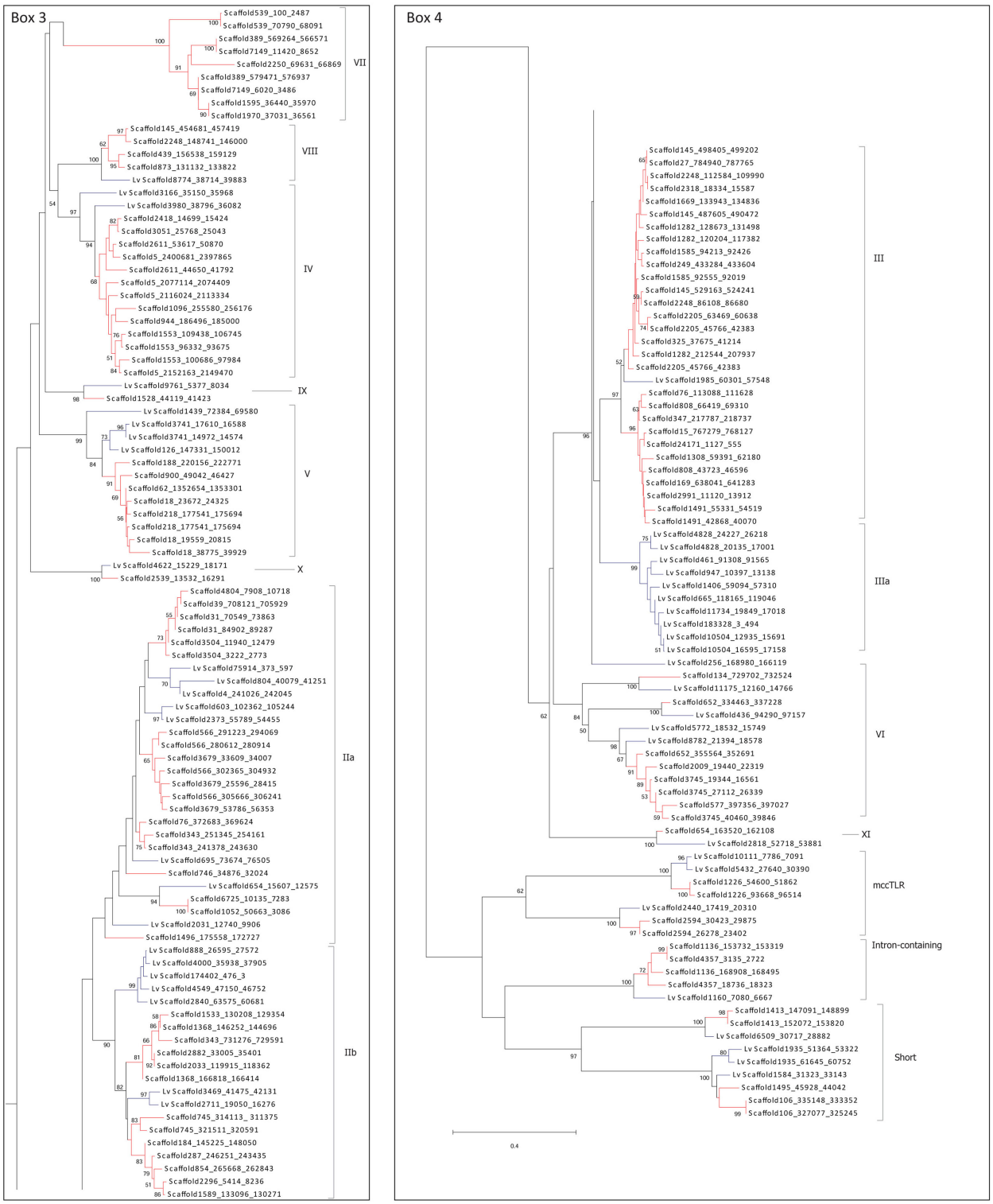


FIGURE A1 | Complete phylogeny of the sea urchin TLR sequences. The TIR domains of the TLR sequences from *S. purpuratus* and *L. variegatus* were used to construct the tree shown, which is a more detailed version of the tree in **Figure 2**. Bootstrap values greater than 50 are shown. Red indicates clades that are specific to *S. purpuratus*, blue clades contain only sequences from *L. variegatus*, and black clades contain sequences from both species. Each of the boxes (1–4) is shown in greater detail as indicated. The sequences are labeled by scaffold number and the position of the open reading frame (scaffold_start_stop). More information about the sequences can be found in Tables S1 and S2 in Supplementary Material. Subfamily designations are indicated on the right of each tree and correspond to those shown in **Figure 2**.

frontiers in
IMMUNOLOGY

REVIEW ARTICLE
published: 29 February 2012
doi: 10.3389/fimmu.2012.00034

Toll-like receptors of deuterostome invertebrates

Honoo Satake* and Toshio Sekiguchi

Bioorganic Research Institute, Suntory Foundation for Life Sciences, Osaka, Japan

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Dmitri V. Krysko, Ghent University, Belgium
Shun-Ichiro Kawabata, Kyushu University, Japan
Jared C. Roach, Institute for Systems Biology, USA

****Correspondence:***
Honoo Satake, Bioorganic Research Institute, Suntory Foundation for Life Sciences, 1-1-1 Wakayamadai, Shimamoto, Mishima, Osaka 618-8503, Japan.
e-mail: satake@sunbor.or.jp

Defensive systems against pathogens are responsible not only for survival or lifetime of an individual but also for the evolution of a species. Innate immunity is expected to be more important for invertebrates than mammals, given that adaptive immunity has not been acquired in the former. Toll-like receptors (TLRs) have been shown to play a crucial role in host defense of pathogenic microbes in innate immunity of mammals. Recent genome-wide analyses have suggested that TLR or their related genes are conserved in invertebrates. In particular, numerous TLR-related gene candidates were detected in deuterostome invertebrates, including a sea urchin (222 TLR-related gene candidates) and amphioxus (72 TLR-related gene candidates). Molecular phylogenetic analysis verified that most of sea urchin or amphioxus TLR candidates are paralogous, suggesting that these organisms expanded TLR-related genes in a species-specific manner. In contrast, another deuterostome invertebrate, the ascidian *Ciona intestinalis*, was found to possess only two TLR genes. Moreover, *Ciona* TLRs, Ci-TLR1 and Ci-TLR2, were shown to possess "hybrid" functionality of mammalian TLRs. Such functionality of Ci-TLRs could not be predicted by sequence comparison with vertebrate TLRs, indicating confounding evolutionary lineages of deuterostome invertebrate TLRs or their candidates. In this review article, we present recent advances in studies of TLRs or their candidates among deuterostome invertebrates, and provide insight into an evolutionary process of TLRs.

Keywords: Toll-like receptor, innate immunity, evolution, deuterostome invertebrate, diversity

INTRODUCTION

Innate immunity is a primary defense system against invading pathogens in invertebrates, in which adaptive immune systems have not been fully developed. Invertebrate innate immunity consists of diverse pathogen recognition systems including: hemolymph coagulation, lectin-mediated complement activation, antimicrobial peptides, and variable region-containing chitin-binding proteins (Nonaka, 2001; Fujita, 2002; Khalturin et al., 2004; Iwanaga and Lee, 2005; Miller et al., 2007; Rast and Messier-Solek, 2008; Bonura et al., 2009; Nonaka and Satake, 2010; Satake and Sasaki, 2010; Dishaw et al., 2011). Most invertebrate immune systems diverged among species, whereas recent genome surveys of various organisms revealed that Toll-like receptors (TLRs) or their related proteins are essentially conserved in an extensive range of organisms from cnidarians to mammals. In particular, deuterostome invertebrates and vertebrates conserve TLR-directed innate immunity (Akira and Takeda, 2004; Khalturin et al., 2004; Dunne and O'Neill, 2005; Iwanaga and Lee, 2005; Roach et al., 2005; Takeda and Akira, 2005; Ishii et al., 2007; Takano et al., 2007; Davidson et al., 2008; Fedders et al., 2008; Matsuo et al., 2008; Oshiumi et al., 2008; Rast and Messier-Solek, 2008; Bosch et al., 2009; Schikorski et al., 2009; Yu et al., 2009; Macagno et al., 2010; Satake and Sasaki, 2010; Cuvillier-Hot et al., 2011; Park et al., 2011; Wu et al., 2011). These studies not only suggest that the TLR system of immune receptors is the most widely distributed throughout the deuterostomes but also reveal several characteristics of TLRs specific to deuterostome invertebrates.

Deuterostomia consists of six major phyla: Xenoturbellida, Echinodermata, Hemichordata, Cephalochordata, Urochordata, and Vertebrata. Each phylum shares the basal developmental process in which the blastopore is converted into the mouth, although their forms, lifestyles, and life spans are dramatically different within deuterostome invertebrates, making them attractive targets for the molecular and functional evolution of various biological phenotypes and systems, including innate immunity. To date, the genomes of three deuterostome invertebrates, *Ciona intestinalis* (ascidian, Urochordata), *Branchiostoma floridae* (amphioxus, Cephalochordata), and *Strongylocentrotus purpuratus* (sea urchin, Echinodermata) have been determined. Genomic analyses have verified that sea urchin, amphioxus, and ascidian possess 222, 72, and 2 ascidian TLR candidates (Hibino et al., 2006; Rast et al., 2006; Huang et al., 2008; Sasaki et al., 2009), indicating high diversity in the number of TLR candidates among deuterostome invertebrates. In contrast, the number of TLRs in vertebrates does not dramatically differ among species; 10 for human and 12 for Takifugu (Roach et al., 2005; Matsuo et al., 2008). These findings suggest that molecular and functional studies of deuterostome invertebrate TLRs have led to their evolution and diversity as well as to the elucidation of molecular mechanisms underlying their innate immunity. In this review, we provide an overview and possible evolutionary scenarios of deuterostome invertebrate TLRs.

TLRs: CHARACTERIZATION, FAMILY MEMBERS, PATHOGEN RECOGNITION, AND SIGNAL TRANSDUCTION

Toll-like receptors were characterized originally as a mammalian ortholog of the *Drosophila melanogaster* (fruitfly) transmembrane protein Toll, which is responsible for antifungal protection and

Table 1 | Comparison of human and *Ciona* TLRs.

TLR	Number of LRR	Ligands	Intracellular localization
hTLR1 (with TLR2)	5	Triacylated lipoprotein	Plasma membrane
hTLR2	9	Zymosan (yeast cell wall) 1,3-β-glucan lipoarabinomannan Heat-killed *Legionella pneumophila* (HKLP, Gram-negative) Heat-killed *Staphylococcu aureus* (HKSA, Gram-positive) Glycosylphosphatidylinositol (GPI)-anchored glycoprotein	Plasma membrane
hTLR3	17	Poly(I:C) (double-stranded RNA)	Endosome
hTLR4 (with MD2)	11	LPS (lipopolysaccharide from Gram-negative bacteria) Lipid A (lipid component of LPS)	Plasma membrane
hTLR5	10	Flagellin (bacterial flagellar filament)	Plasma membrane
hTLR6	6	MALP-2 (mycoplasma-derived macrophage-activating lipopeptide) FSL1 (mycoplasma-derived lipoprotein)	Plasma membrane
hTLR7	14	Imidazoquimod (imidazoquinolone amino acid analog)	Endosome
hTLR8	16	Single-stranded RNA	Endosome
hTLR9	19	Unmethylated CpG DNA	Endosome
Ci-TLR1	7	Zymosan (yeast cell wall), heat-killed *Legionella pneumophila* (HKLP, Gram-negative), poly(I:C) (double-stranded RNA), flagellin (bacterial flagellar filament)	Plasma membrane and endosome
Ci-TLR2	13		

determination of the dorsal/ventral pattern (Lemaitre et al., 1996; Hoffmann, 2003). TLRs are type I transmembrane proteins, featured by an intracellular Toll/Interleukin-1 receptor (TIR) domain and leucine-rich repeat (LRR) motifs in the extracellular domain (Akira and Takeda, 2004; Dunne and O'Neill, 2005; Takeda and Akira, 2005), and have been found to play pivotal roles in mammalian host defenses via the innate immune system. LRRs exhibit specific pathogenic ligand recognition, and TIRs participate in activation of signaling pathways. TLR4 is the first mammalian TLR that was described and was characterized by functional analyses of a mouse deficient in host defense against a typical endotoxin, lipopolysaccharide (LPS; Poltorak et al., 1998). Other TLRs were structurally and functionally characterized in human and mice. To date, 10 functional human TLRs (hTLRs) have been identified (Akira and Takeda, 2004; Dunne and O'Neill, 2005; Roach et al., 2005; Takeda and Akira, 2005; Park et al., 2011). As summarized in **Table 1**, each hTLR differs in the number and organization of LRRs, and recognizes their specific ligands, namely pathogen-associated molecular patterns (PAMPs). The diversity in the numbers and organization of LRR domains are believed to enable the specific and sensitive recognition of PAMPs by respective TLRs (Akira and Takeda, 2004; Dunne and O'Neill, 2005; Takeda and Akira, 2005). TLRs are expressed in the lung, kidney, small intestine, stomach, ovary, and testis as well as in immune cells such as lymphocytes, macrophages, and dendritic cells (Akira and Takeda, 2004; Dunne and O'Neill, 2005; Takeda and Akira, 2005). TLR1, TLR2, TLR4, TLR5, and TLR6 are responsible for recognition of extracellular microbial pathogenic components on plasma membranes, whereas TLR3, TLR7, TLR8, and TLR9, exclusively present on endosomes, participate in defensive responses to viral DNA or RNA incorporated into the cytoplasm (Akira and Takeda, 2004; Dunne and O'Neill, 2005; Takeda and Akira, 2005). Thus, TLRs are categorized into plasma membrane TLR groups (TLR1, -2, -4, -5, and -6) or endosome TLR groups (TLR3, -7, -8, and -9) in terms of their intracellular localization. In addition, TLR4 requires an extracellular associated protein, MD2, to recognize LPS (Shimazu et al., 1999; Akashi et al., 2000; Dunne and O'Neill, 2005; Takeda and Akira, 2005). Moreover, the tertiary structures of several TLR–ligand complexes have been resolved (Jin and Lee, 2008; Liu et al., 2008; Park et al., 2009). Interaction of TLRs with their specific PAMPs triggers signal transduction pathways via adaptor proteins (MyD88, TIRAP, TRIF, and TRAM) followed by activation of a wide range of inducible transcriptional factors such as: NF-κB, AP-1, and IRF, leading to production of inflammatory cytokines, chemokines, and/or type I interferon (Akira and Takeda, 2004; Dunne and O'Neill, 2005; Takeda and Akira, 2005). As stated above, molecular forms and biological roles of TLR were investigated extensively in mammals. However, over the past decades, deuterostome invertebrates also have been shown to share, at least in part, the TLR-triggered innate immune system.

TLR IN *CIONA INTESTINALIS*: THE FIRST FUNCTIONALLY CHARACTERIZED DEUTEROSTOME INVERTEBRATE TLRs

A *Ciona* genome survey and molecular cloning revealed the presence of two TLRs in *C. intestinalis*, Ci-TLR1 and Ci-TRL2 (Azumi et al., 2003; Sasaki et al., 2009), both of which are composed of a TIR, a transmembrane, and a LRR domain. Moreover, 7 and 13 LRRs are found in Ci-TLR1 and Ci-TLR2, respectively (**Figure 1**). The *ci-tlr1* and *ci-tlr2* genes are expressed intensively in the stomach, intestine, and numerous hemocytes and, to a lesser degree, the central nervous system (Sasaki et al., 2009). These findings indicate that Ci-TLRs function mainly in the alimentary tracts and hemocytes. Interestingly, both Ci-TLRs, unlike any vertebrate TLRs, are present on the plasma membrane and a number of late

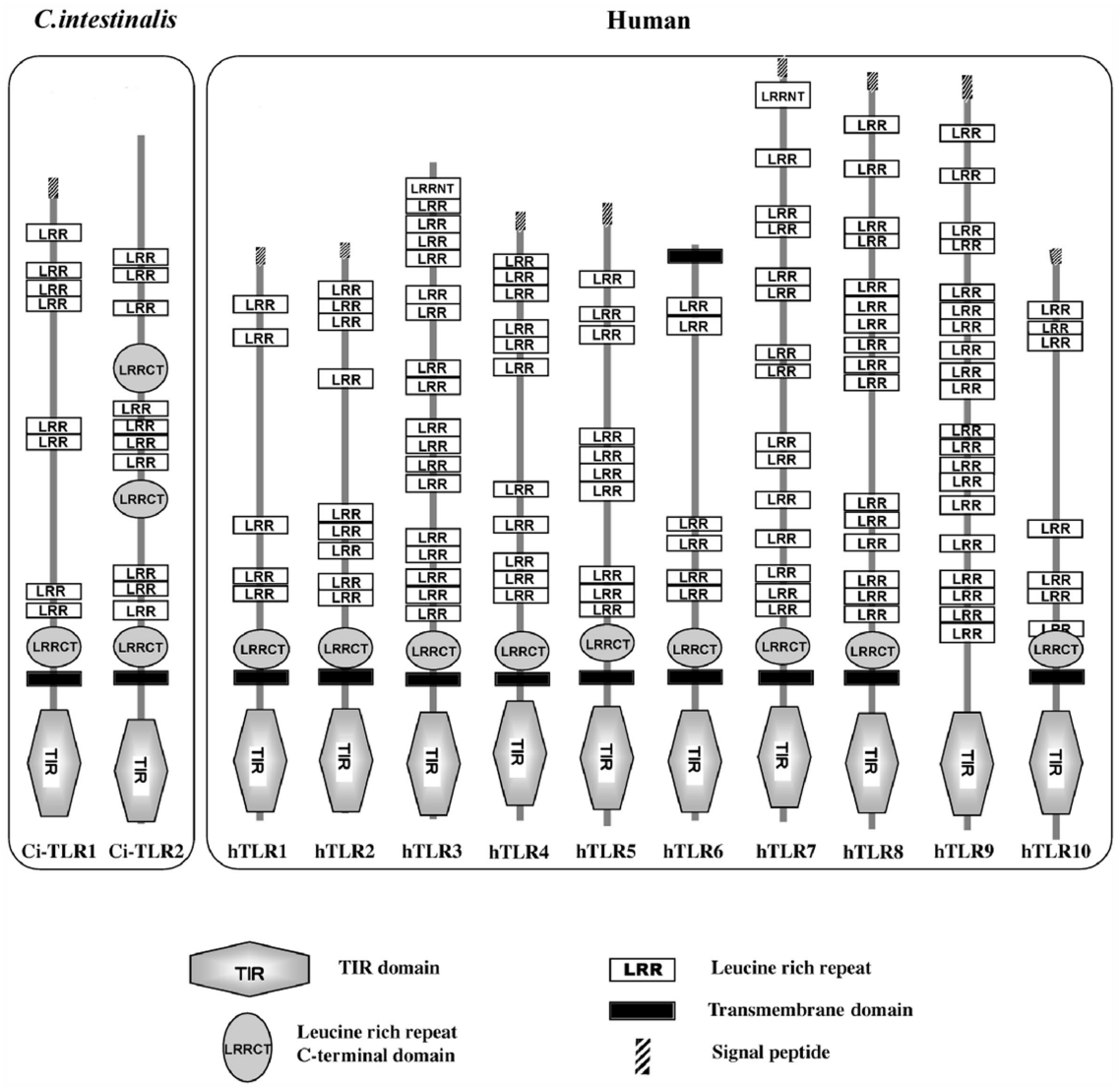


FIGURE 1 | Structural organization of *Ciona* TLRs (Ci-TLR*)* and human TLRs (hTLR).

endosomes (Sasaki et al., 2009). A striking feature is that Ci-TLR1 and Ci-TLR2 expressed in HEK293MSR cells activated NF-κB in response to multiple TLR ligands (**Table 1**), which are recognized by different mammalian TLRs: zymosan (*Saccharomyces cerevisiae* cell wall) for hTLR2, heat-killed *Legionella pneumophila* (HKLP, a Gram-negative bacterium) for hTLR2, double-stranded RNA poly(I:C) for hTLR3, and *Salmonella typhimurium* Flagellin (the major component of the bacterial flagellar filament) for hTLR5, which elicited a dose-dependent transactivation of NF-κB in the *ci-tlr1*- or *ci-tlr2*-expressing cells (Sasaki et al., 2009). Poly(I:C) and flagellin also elicited approximately 4- and 10-fold Ci-TNFα expression in the stomach, anterior and/or middle intestine where *ci-tlr1* or *ci-tlr2* are abundantly expressed, but not in the posterior intestine where expression of *ci-tlr*s was not detected (Sasaki et al., 2009). These data lead to two important conclusions. First, Ci-TLRs, like vertebrate TLRs, directly recognize PAMPs and trigger the transactivation of NF-κB. This is in contrast with *Drosophila* Toll, which requires insect-specific accessory proteins for pathogen recognition and the following signal transduction (Hoffmann, 2003). Consequently, Ci-TLRs share the essential ligand recognition and signaling mechanisms with the vertebrate TLR family, which is compatible with a previous genomic survey clarifying the presence of *Ciona* orthologs of TLR-signaling factors such as MyD88, NF-κB, and TNFα (Azumi et al., 2003). Second, Ci-TLRs recognize more extensive PAMPs than vertebrate TLRs (**Table 1**). This is the first report on TLRs that respond to multiple PAMPs of respective hTLRs. The PAMP recognition by Ci-TLRs is in good agreement with intracellular localization to both the plasma membrane and endosome, given that poly(I:C) is recognized by hTLR3 present on endosome, and that hTLR2 and hTLR5 on the plasma membrane interact with zymosan, HKLP, and flagellin. Collectively these

multiple PAMP recognition specificities and intracellular localization differences provide evidence that Ci-TLRs are functionally hybrid TLRs of vertebrate cell-surface TLRs (hTLR1, -2, -4, -5, -6) and endosome TLRs (hTLR3, -7, -8, -9). Such functions of Ci-TLRs are not predicted on the basis of sequence homology to hTLRs, leading to the unique location in the molecular phylogenetic trees of chordate TLRs and their candidates. It is noteworthy that both Ci-TLRs elicited equipotent NF-κB activation in response to the same ligands, whereas Ci-TLRs display different structural organization of LRRs (**Table 1**). These findings suggest that recognition of PAMPs and that the resulting signaling by deuterostome invertebrate TLRs (including Ci-TLRs) cannot be elucidated or even predicted by sequence homology-based analyses.

Ci-TLRs are devoid of any responses to LPS (Sasaki et al., 2009). In *C. intestinalis*, *in vivo* challenge of LPS resulted in the upregulation of a TLR-triggering cytokine, TNFα (Parrinello et al., 2008), and a mannose-binding lectin-like collectin (Bonura et al., 2009), indicating that *C. intestinalis* possesses a LPS-responsive system. In mammals, LPS is recognized by a complex of hTLR4 and MD2, not by hTLR4 alone (Shimazu et al., 1999; Akashi et al., 2000; Dunne and O'Neill, 2005; Takeda and Akira, 2005). This is consistent with the data showing that Ci-TLRs fail to recognize LPS. Moreover, no MD2 homolog was detected in the *C. intestinalis* genome (Azumi et al., 2003). Taken together, these findings lead to two hypotheses: (1) Ci-TLRs possess no ability to recognize LPS, and (2) other LPS-responsive TLR systems are functional in *C. intestinalis*. Interestingly, recognition of LPS by non-TLR receptors in fish is also suggested (Iliev et al., 2005), i.e., recognition of LPS by TLR4 with MD2 may have been acquired during the evolution of vertebrates. Alternatively, Ci-TLRs may respond to LPS through formation of a complex with other associated proteins than MD2. If this is true, the question is raised as to whether such an MD2-independent LPS-responsive system is specific to *C. intestinalis* or is common among deuterostome invertebrates. Both molecular and functional studies will address this issue.

Ci-TLRs display moderate differences in their expression despite almost identical ligand recognition; Ci-TLR1 is expressed much more in the anterior intestine, compared to Ci-TLR2 (Sasaki et al., 2009). On the other hand, Ci-TLR1 and Ci-TLR2 share their PAMPs (**Table 1**) and exhibit equipotent NF-κB activation. These findings indicate the possibility that Ci-TLR1 and Ci-TLR2 have different immunological and/or biological roles. For instance, they may differentially recognize other PAMPs that have not been investigated or may have another signaling cascade that differs from NF-κB activation. Alternatively, Ci-TLR1 and Ci-TLR2 are anticipated to play a differential role in other biological functions other than innate immunity. Recently, TLRs have also been shown to recognize several high-concentration endogenous ligands, including heme and endogenous double-stranded RNA resulting from apoptosis (Figueiredo et al., 2007; Wu et al., 2008). These findings suggest biological roles for TLRs as a sensor of unusual generation of endogenous components, and that the expression profiles for Ci-TLRs reflect different functions in recognition of endogenous ligands. Elucidation of their principal functional roles is underway in our laboratory.

AMPHIOXUS TLRs

Seventy-two TLR or TLR-related genes were detected in the genome of *B. floridae* (Huang et al., 2008). Recently, an amphioxus TLR (bbtTLR) was identified in another amphioxus species, *Branchiostoma belcheri tsingtauense* (Yuan et al., 2009). The expression of *bbttlr* was initiated at the late blastula/early gastrula stage, and observed from the embryonic to larval stages (Yuan et al., 2009), suggesting a developmental role. In adults, the *bbttlr* expression was detected predominantly in the gill, skin, and gut, and, to a lesser degree in the ovaries, muscles, hemocytes, and notochord (Yuan et al., 2009). Such *bbttlr* expression also suggests an immunological function in the innate immunity, given that the gill, skin, and gut are exposed directly to pathogens. Interestingly, LPS, lipoteichoic acids, and Gram-negative bacteria (*Vibrio vulnificus*) upregulate the bbtTLR transcript, whereas poly(I:C) and glucan downregulate the transcript (Yuan et al., 2009). These data support the hypothesis that bbtTLR participates in defense responses to certain pathogens. However, neither intact nor chimeric bbtTLR recognize LPS, lipoteichoic acids, peptidoglycan, poly(I:C), or glucan, although bbtTLR was localized to the plasma membrane (Yuan et al., 2009). This failure is due to the heterologous expression of bbtTLR in human culture cells. Otherwise, other TLR ligands are likely to activate bbtTLR, which may require a certain associated protein for recognition of ligands, as MD2 is a prerequisite for recognition of LPS by hTLR4 (Shimazu et al., 1999; Akashi et al., 2000; Dunne and O'Neill, 2005; Takeda and Akira, 2005). Elucidation of function of bbtTLR awaits further analysis.

SEA URCHIN TLRs: THE LARGEST REPERTOIRE

The genome-wide analysis of the purple sea urchin, *Strongylocentrotus purpuratus*, detected 222 TLR-like genes (Hibino et al., 2006; Rast et al., 2006). A molecular phylogenetic tree analysis led to classification of *S. purpuratus* TLR genes to seven major groups (Hibino et al., 2006). The vast majority of sea urchin TLR genes display higher similarity to each other than to TLRs of other species TLRs and are encoded in tandem arrays, suggesting an enormous gene expansion that is specific to *Strongylocentrotus* TLRs (Hibino et al., 2006). Although tissue distribution of sea urchin TLRs has yet to be investigated, TLR cDNAs were identified from coelomocyte. Major classes of sea urchin TLR cDNAs were not expressed in embryos prior to the end of gastrulation. These data support the hypothesis that sea urchin TLRs have primarily an immunological role, unlike insect Toll. Function and cellular localization of sea urchin TLRs await further investigation. Random investigations of all 222 sea urchin TLRs is difficult. Nevertheless, TLR-signaling-associated protein homologs such as: MyD88, NF-κB, and TNFα, were detected in the genome of *S. purpuratus* (Hibino et al., 2006), implying that the essential TLR-triggered innate immune system is conserved in sea urchin. Thus, a promising target is several sea urchin TLRs that are abundantly expressed in immune tissues and cells, such as alimentary tract or hemocytes.

EVOLUTIONARY SCENARIO OF TLR IN DEUTEROSTOME INVERTEBRATES

Amphioxus and sea urchin possess a great number of TLRs or related genes (Hibino et al., 2006; Rast et al., 2006; Huang et al., 2008). Furthermore, molecular phylogenetic analyses demonstrate

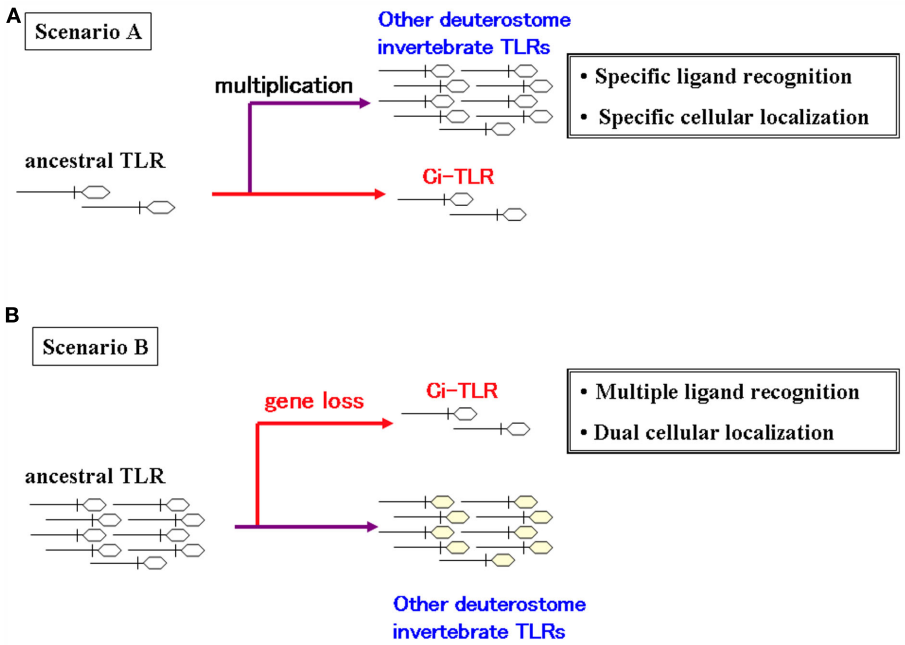


FIGURE 2 | Possible evolutionary scenarios of deuterostome invertebrate TLRs. (A) only a few TLRs or their related genes might have existed in a common deuterostome ancestor or **(B)** a common deuterostome ancestor might have numerous TLR family genes.

that most of these genes are generated via species-specific gene duplication, suggesting that these deuterostome invertebrates expand TLRs or their related genes in unique lineages of innate immunity, if most of the genes are functional (Hibino et al., 2006; Huang et al., 2008; Rast and Messier-Solek, 2008). In contrast, *C. intestinalis* possesses only two authentic TLRs (Sasaki et al., 2009), leading to two evolutionary scenarios of TLRs or their related genes (**Figure 2**). First, only a few TLR or their related genes might have existed in a common deuterostome ancestor. In this case, *C. intestinalis* conserves the ancestral characteristics, whereas sea urchins and amphioxus expanded their TLR or related gene paralogs during their divergence in concert with the variation of their lifetimes, life cycles, or environments. This presumption is also compatible with the aforementioned molecular phylogenetic analyses of sea urchin and amphioxus TLRs and their related genes. Moreover, the ancestral TLRs, as seen in Ci-TLRs, might have responded to multiple PAMPs, which are differentially recognized by respective currently existing vertebrate TLRs and have been present both on cell surface and on endosomes. In keeping with this, vertebrate and probably most of sea urchins and amphioxus TLRs might have acquired their specific PAMP recognition and intracellular localization during their evolution. Alternatively, a common deuterostome antecedent might have numerous TLR family genes, as seen in sea urchin and amphioxus. If true, *C. intestinalis* should have lost a large part of ancestral TLR family genes, instead, Ci-TLRs are highly likely to have acquired multiple PAMP recognition and intracellular localization. Therefore, characterization of PAMPs and intracellular localization of sea urchin and amphioxus TLRs is expected to contribute not only to elucidation of their biological roles but also to investigation of molecular and functional divergence of the deuterostome TLR family.

It is noteworthy that all deuterostome invertebrates are marine organisms that should establish an immune system specific to ocean pathogens, given that these pathogens do not completely coincide with those for land mammals. This view may be supported by difference in PAMP recognition between Ci-TLRs and hTLR2; two ligands for TLR2, zymosan and HKLP exhibit prominent activation of Ci-TLRs, whereas other hTLR2 ligands, including HKLM, 1,3-β-glucan, and HKSA, are devoid of any activity at Ci-TLRs (Sasaki et al., 2009). Collectively, these findings indicate the immunological significance of the considerably high diversity in sequences and organization of LRRs, which is consistent with the recognized PAMPs. Innate immunity against pathogens is expected to be closely related to the lifetimes and life cycles of organisms. For instance, the sea urchin has an extremely long lifetime (30–40 years), indicating that this organism is endowed with excellent defense against pathogens. In contrast, the lifetime of *C. intestinalis* is only 3 months. Such differences imply that the numbers of TLRs or TLR-like candidate genes may be correlate with lifespans of organisms. In keeping with this, the molecular and functional characterization of TLRs of another ascidian, *Halocynthia roretzi,* should lead to incisive findings due to its much longer lifespan (up to 8 years) and its distribution in relatively different habitats. Feeding strategies may also participate in the diversity in the number and functions of TLRs. For instance, *Ciona* feeds exclusively on plant planktons and sea urchin feeds on both dead fish and plant planktons. These feeding differences suggest that sea urchin may encounter more various pathogens than *Ciona* (Davidson et al., 2008). Intriguingly, a land annelid, the blood-sucking leech *Helobdella robusta*, was found to possess 16 TLR-like genes, whereas the marine non-blood-sucking annelid *Capitella capitata* was found to possess 105 TLR-like genes. The

authors hypothesized that such low number of *H. robusta* TLR-like genes compared with that of *C. capitata* may be due to the possibility that blood-sucking leeches obtain functional immune components from their host animals.

CONCLUSIONS AND PERSPECTIVES

Deuterostome TLRs or their candidate genes have been identified with growing efficiency over the past few years. However, Ci-TLRs are at present the only functionally characterized TLRs of deuterostome invertebrates. This can be attributed to the high numbers of sea urchin and amphioxus TLRs, high molecular divergence in the numbers and organization of the LRRs among deuterostome invertebrate TLR (or TLR-like) genes, and severe limitations of homology-based sequence analysis for characterizing crucial functions of PAMPs, cellular localization, and signaling pathways, which are of interest in amphioxus and sea urchin. Apart from the high number of sea urchin and amphioxus TLR candidate genes, one of the difficulties in functional analyses of deuterostome invertebrate TLRs results from failure of functional expression of the TLRs in certain heterologous host cultured cells due to the lack of availability of homologous expression systems including cognate cultured cell lines, as seen in studies of amphioxus and cyclostomes (Ishii et al., 2007; Yuan et al., 2009). Therefore, establishment of cultured cells derived from each organism will strongly enhance various functional studies of deuterostome invertebrate TLRs. In this context, Ci-TLRs possess distinct advantages as they can be expressed efficiently in HEK293MSR cells, thus facilitating the characterization of their PAMPs, intracellular localization, and signaling.

Also of biological and immunological significance is the characterization of functional relationship TLRs and other innate immune systems, including the complement activation system, antibacterial peptides synthesis and hemolymph coagulation. For instance, it has been speculated that TLR-triggered signaling in *C. intestinalis* can lead to complement activation. Furthermore, variable region-containing chitin-binding proteins may serve cooperatively as mediators of immunity along with TLRs. Integration of functional roles is expected to pave the way to the characterization of innate immunity at the molecular level in deuterostome invertebrates.

ACKNOWLEDGMENTS

We thank Ms. Kazuko Hirayama and all members of Maizuru Fisheries Research Station for cultivation of the ascidians. All ascidians were provided by Kyoto University through the National Bio-Resource Project of the MEXT, Japan.

REFERENCES

Akashi, S., Shimazu, R., Ogata, H., Nagai, Y., Takeda, K., Kimoto, M., and Miyake, K. (2000). Cutting edge: cell surface expression and lipopolysaccharide signaling via the toll-like receptor 4-MD-2 complex on mouse peritoneal macrophages. *J. Immunol.* 164, 3471–3475.

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

Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., Arai, M., Inoue, Y., Shimizu, T., Satoh, N., Rokhsar, D. S., Du Pasquier, L., Kasahara, M., Satake, M., and Nonaka, M. (2003). Genomic analysis of immunity in a urochordate and the emergence of the vertebrate immune system: "waiting for Godot." *Immunogenetics* 55, 570–581.

Bonura, A., Vizzini, A., Salerno, G., Parrinello, N., Longo, V., and Colombo, P. (2009). 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.* 46, 2389–2394.

Bosch, T. C. G., Augustin, R., Anton-Erxleben, F., Fraune, S., Hemmrich, G., Zill, H., Rosenstiel, P., Jacobs, G., Schreiber, S., Leippe, M., Stanisak, M., Grötzinger, J., Jung, S., Podschun, R., Bartels, J., Harder, J., and Schröder, J. M. (2009). Uncovering the evolutionary history of innate immunity: the simple metazoan *Hydra* uses epithelial cells for host defense. *Dev. Comp. Immunol.* 33, 559–569.

Cuvillier-Hot, V., Boidin-Wichlacz, C., Slomianny, C., Salzet, M., and Tasiemski, A. (2011). Characterization and immune function of two intracellular sensors, HmTLR1 and HmNLR, in the injured CNS of an invertebrate. *Dev. Comp. Immunol.* 35, 214–226.

Davidson, C. R., Best, N. M., Francis, J. W., Cooper, E. L., and Wood, T. C. (2008). Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). *Dev. Comp. Immunol.* 32, 608–612.

Dishaw, L. J., Giacomelli, S., Melillo, D., Zucchetti, I., Haire, R. N., Natale, L., Russo, N. A., De Santis, R., Litman, G. W., and Pinto, M. R. (2011). A role for variable region-containing chitin-binding proteins (VCBPs) in host gut-bacteria interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16747–16752.

Dunne, A., and O'Neill, L. A. (2005). Adaptor usage and Toll-like receptor signaling specificity. *FEBS Lett.* 579, 3330–3335.

Fedders, H., Michalek, M., Grötzinger, J., and Leippe, M. (2008). An exceptional salt-tolerant antimicrobial peptide derived from a novel gene family of haemocytes of the marine invertebrate *Ciona intestinalis*. *Biochem. J.* 416, 65–75.

Figueiredo, R. T., Fernandez, P. L., Mourao-Sa, D. S., Porto, B. N., Dutra, F. F., Alves, L. S., Oliveira, M. F., Oliveira, P. L., Graça-Souza, A. V., and Bozza, M. T. (2007). Characterization of heme as activator of Toll-like receptor 4. *J. Biol. Chem.* 282, 20221–20229.

Fujita, T. (2002). Evolution of the lectin-complement pathway and its role in innate immunity. *Nat. Rev. Immunol.* 2, 346–353.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P., Buckley, K. M., Brockton, V., Nair, S. V., Berney, K., Fugmann, S. D., Anderson, M. K., Pancer, Z., Cameron, R. A., Smith, L. C., and Rast, J. P. (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* 300, 349–365.

Hoffmann, J. A. (2003). The immune response of *Drosophila*. *Nature* 426, 33–38.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., Ge, J., Yu, Y., Huang, H., Dong, M., Yu, C., Chen, S., and Xu, A. (2008). Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* 18, 1112–1126.

Iliev, D. B., Roach, J. C., Mackenzie, S., Planas, J. V., and Goetz, F. W. (2005). Endotoxin recognition: in fish or not in fish? *FEBS Lett.* 579, 6519–6528.

Ishii, A., Matsuo, A., Sawa, H., Tsujita, T., Shida, K., Matsumoto, M., and Seya, T. (2007). Lamprey TLRs with properties distinct from those of the variable lymphocyte receptors. *J. Immunol.* 178, 397–406.

Iwanaga, S., and Lee, B. L. (2005). Recent advances in the innate immunity of invertebrate animals. *J. Biochem. Mol. Biol.* 38, 128–150.

Jin, M. S., and Lee, J. O. (2008). Structures of the toll-like receptor family and its ligand complexes. *Immunity* 29, 182–191.

Khalturin, K., Panzer, Z., Cooper, M. D., and Bosch, T. C. G. (2004). Recognition strategies in the innate immune system of ancestral chordates. *Mol. Immunol.* 41, 1077–1187.

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

Liu, L., Botos, I., Wang, Y., Leonard, J. N., Shiloach, J., Segal, D. M., and Davies, D. R. (2008). Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* 320, 379–381.

Macagno, E. R., Gaasterland, T., Edsall, L., Bafna, V., Soares, M. B., Scheetz, T., Casavant, T., Da Silva, C., Wincker, P., Tasiemski, A., and Salzet, M. (2010). Construction of a medicinal leech transcriptome database and its application to the identification of leech homologs of neural and innate immune genes. *BMC Genomics* 11, 407. doi:10.1186/1471-2164-11-407

Matsuo, A., Oshiumi, H., Tsujita, T., Mitani, H., Kasai, H., Yoshimizu, M., Matsumoto, M., and Seya, T. (2008). Teleost TLR22 recognizes RNA duplex to induce IFN and protect cells from birnaviruses. *J. Immunol.* 18, 3474–3485.

Miller, D. J., Hemmrich, G., Ball, E. E., Hayward, D. C., Khalturin, K., Funayama, N., Agata, K., and Bosch, T. C. G. (2007). The innate immune repertoire in cnidaria – ancestral complexity and stochastic gene loss. *Genome Biol.* 8, R59.

Nonaka, M. (2001). Evolution of the complement system. *Curr. Opin. Immunol.* 13, 69–73.

Nonaka, M., and Satake, H. (2010). Urochordate immunity. *Adv. Exp. Med. Biol.* 708, 302–310.

Oshiumi, H., Matsuo, A., Matsumoto, M., and Seya, T. (2008). Panvertebrate Toll-like receptors during evolution. *Curr. Genomics* 9, 488–493.

Park, B. S., Song, D. H., Kim, H. M., Choi, B. S., Lee, H., and Lee, J. O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 458, 1191–1195.

Park, H. J., Hahn, W. H., Suh, J. S., Kim, M. J., Kang, S. W., Lee, J. S., Kim, J. W., Chung, J. H., and Cho, B. S. (2011). Association between toll-like receptor 10 (TLR10) gene polymorphisms and childhood IgA nephropathy. *Eur. J. Pediatr.* 170, 503–509.

Parrinello, N., Vizzini, A., Arizza, V., Salerno, G., Parrinello, D., Cammarata, M., Giaramita, F. T., and Vazzana, M. (2008). Enhanced expression of a cloned and sequenced *Ciona intestinalis* TNFα-like (CiTNFα) gene during the LPS-induced inflammatory response. *Cell Tissue Res.* 334, 305–317.

Poltorak, A., He, X., Smirnova, I., Liu, M. Y., Van Huffel, C., Du, X., Birdwell, D., Alejos, E., Silva, M., Galanos, C., Freudenberg, M., Ricciardi-Castagnoli, P., Layton, B., and Beutler, B. (1998). Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 282, 2085–2088.

Rast, J. P., and Messier-Solek, C. (2008). Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol. Bull.* 214, 274–283.

Rast, J. P., Smith, L. C., Loza-Coll, M., Hibino, T., and Litman, G. W. (2006). Genomic insights into the immune system of the sea urchin. *Science* 314, 952–956.

Roach, J. C., Glusman, G., Rowen, L., Kaur, A., Purcell, M. K., Smith, K. D., Hood, L. E., and Aderem, A. (2005). The evolution of vertebrate Toll-like receptors. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9577–9582.

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

Satake, H., and Sasaki, N. (2010). Comparative view on Toll-like receptors of lower animals. *Zool. Sci.* 27, 154–161.

Schikorski, D., Cuvillier-Hot, V., Boidin-Wichlacz, C., Slomianny, C., Salzet, M., and Tasiemski, A. (2009). Deciphering the immune function and regulation by a TLR of the cytokine EMAPII in the lesioned central nervous system using a leech model. *J. Immunol.* 183, 7119–7128.

Shimazu, R., Akashi, S., Ogata, H., Nagai, Y., Fukudome, K., Miyake, K., and Kimoto, M. (1999). MD-2, a molecule that confers lipopolysaccharide responsiveness on Toll-like receptor 4. *J. Exp. Med.* 189, 1777–1782.

Takano, T., Kondo, H., Hirono, I., Endo, M., Saito-Taki, T., and Aoki, T. (2007). Molecular cloning and characterization of Toll-like receptor 9 in Japanese flounder, *Paralichthys olivaceus*. *Mol. Immunol.* 44, 1845–1853.

Takeda, K., and Akira, S. (2005). Toll-like receptors in innate immunity. *Int. Immunol.* 17, 1–14.

Wu, B., Xin, B., Jin, M., Wei, T., and Bai, Z. (2011). Comparative and phylogenetic analyses of three TIR domain-containing adaptors in metazoans: implications for evolution of TLR signaling pathways. *Dev. Comp. Immunol.* 35, 764–773.

Wu, H., Wang, H., Xiong, W., Chen, S., Tang, H., and Han, D. (2008). Expression patterns and functions of toll-like receptors in mouse sertoli cells. *Endocrinology* 149, 4402–4412.

Yu, Y., Zhong, Q., Li, C., Jiang, L., Yan, F., Wang, Z., and Zhang, Q. (2009). Isolation and characterization of Toll-like receptor 9 in half-smooth tongue sole *Cynoglossus semilaevis*. *Fish Shellfish Immunol.* 26, 492–499.

Yuan, S., Huang, S., Zhang, W., Wu, T., Dong, M., Yu, Y., Liu, T., Wu, K., Liu, H., Yang, M., Zhang, H., and Xu, A. (2009). An amphioxus TLR with dynamic embryonic expression pattern responses to pathogens and activates NF-κB pathway via MyD88. *Mol. Immunol.* 46, 2348–2356.

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.

Received: 28 November 2011; accepted: 15 February 2012; published online: 29 February 2012.

Citation: Satake H and Sekiguchi T (2012) Toll-like receptors of deuterostome invertebrates. Front. Immun. ***3****:34. doi: 10.3389/fimmu.2012.00034*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

REVIEW ARTICLE
published: 12 April 2012
doi: 10.3389/fimmu.2012.00070

Innate immune complexity in the purple sea urchin: diversity of the *Sp185/333* system

L. Courtney Smith*

Department of Biological Sciences, George Washington University, Washington, DC, USA

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Yuko Ota, University of Maryland, USA
Philip Rosenstiel, University Kiel, Germany
Tatsuya Ota, The Graduate University for Advanced Studies, Japan

****Correspondence:***
L. Courtney Smith, Department of Biological Sciences, George Washington University, 340 Lisner Hall, 2023 G Street NorthWest, Washington, DC 20052, USA.
e-mail: csmith@gwu.edu

The California purple sea urchin, *Strongylocentrotus purpuratus*, is a long-lived echinoderm with a complex and sophisticated innate immune system. There are several large gene families that function in immunity in this species including the *Sp185/333* gene family that has ∼50 (±10) members. The family shows intriguing sequence diversity and encodes a broad array of diverse yet similar proteins. The genes have two exons of which the second encodes the mature protein and has repeats and blocks of sequence called *elements*. Mosaics of *element patterns* plus single nucleotide polymorphisms-based variants of the elements result in significant sequence diversity among the genes yet maintains similar structure among the members of the family. Sequence of a bacterial artificial chromosome insert shows a cluster of six, tightly linked *Sp185/333* genes that are flanked by GA microsatellites. The sequences between the GA microsatellites in which the *Sp185/333* genes and flanking regions are located, are much more similar to each other than are the sequences outside the microsatellites suggesting processes such as gene conversion, recombination, or duplication. However, close linkage does not correspond with greater sequence similarity compared to randomly cloned and sequenced genes that are unlikely to be linked. There are three segmental duplications that are bounded by GAT microsatellites and include three almost identical genes plus flanking regions. RNA editing is detectible throughout the mRNAs based on comparisons to the genes, which, in combination with putative post-translational modifications to the proteins, results in broad arrays of Sp185/333 proteins that differ among individuals. The mature proteins have an N-terminal glycine-rich region, a central RGD motif, and a C-terminal histidine-rich region. The Sp185/333 proteins are localized to the cell surface and are found within vesicles in subsets of polygonal and small phagocytes. The coelomocyte proteome shows full-length and truncated proteins, including some with missense sequence. Current results suggest that both native Sp185/333 proteins and a recombinant protein bind bacteria and are likely important in sea urchin innate immunity.

Keywords: echinoid, invertebrate, evolution, innate immunity, coelomocyte, microsatellites, RNA editing, gene family

INTRODUCTION

SEA URCHINS AND ECHINODERMS

The phylum Echinodermata is phylogenetically positioned within the deuterostome lineage of animals that includes the Chordata and a few minor invertebrate phyla. There are five extant classes of echinoderms, of which the echinoid class includes sea urchins and sand dollars (**Figure 1**). Immunologists initially assumed that all animals had adaptive immune functions like mammals and worked toward demonstrating these characteristics in many species of animals including invertebrates. Initial investigations to determine the immune capabilities in echinoderms employed allograft rejection assays (Hildemann and Dix, 1972; Karp and Hildemann, 1976; Coffaro and Hinegardner, 1977; Coffaro, 1980; Varadarajan and Karp, 1983). However, the rejection kinetics of first set, second set, and third party allografts in the sea urchin, *Lytechinus pictus* (Coffaro and Hinegardner, 1977; Coffaro, 1979), showed that the response did not demonstrate specific immune recognition and immune memory and was therefore solely innate (Smith and Davidson, 1992). The initial assumptions that immunologists held about universal functions of adaptive immunity were very wrong.

The cellular mediators of immunity in echinoderms are coelomocytes that are present in the fluid that fills the coelomic cavity (Boolootian and Geise, 1958; Johnson, 1969a; Smith, 1981; reviewed in Smith et al., 2010). Coelomocyte immune function was first recognized by Metchnikoff (1893) when he inserted rose prickles, glass rods, and bacteria into the blastocoel of larval sea stars and observed either encapsulation or phagocytosis by the blastocoelar cells (larval equivalents of coelomocytes). More recent analysis of the clearance of foreign cells and particles from adult and larval echinoderms shows that it is swift, efficient, and is mediated by the phagocytic cells in the coelomic and blastocoelar cavities (Reinisch and Bank, 1971; Yui and Bayne, 1983; Plytycz and Seljelid, 1993; Silva, 2000; Furukawa et al., 2009) In the purple

FIGURE 1 | An adult California purple sea urchin, *Strongylocentrotus purpuratus*. Image kindly provided by Hung-Yen Chou and Yen-Lin Kuo.

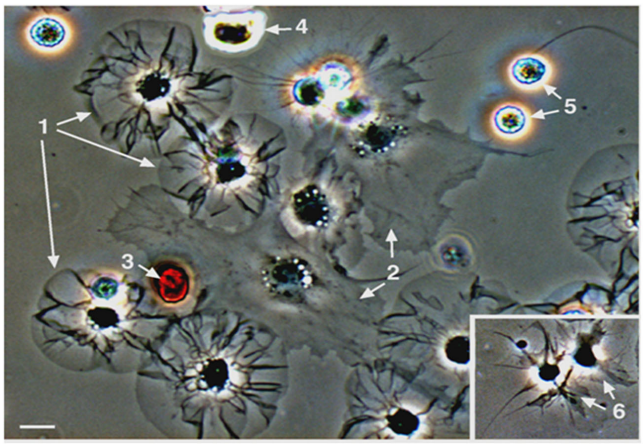


FIGURE 2 | Sea urchin coelomocytes. Coelomocytes from the green sea urchin, *Strongylocentrotus droebachiensis*, were settled onto a glass coverslip and imaged live. Cell types that are typical of echinoids include discoidal phagocytes (1), polygonal phagocytes (2), red spherule cells (3), colorless spherule cells (4), vibratile cells (5; the lower cell has lost the prominent flagellum that is present in the upper cell), and small phagocytes (6, inset). Bar = 10 μm. This is Figure 2 from Smith et al. (2010), reproduced with permission from Landes Bioscience and Springer Science + Business Media.

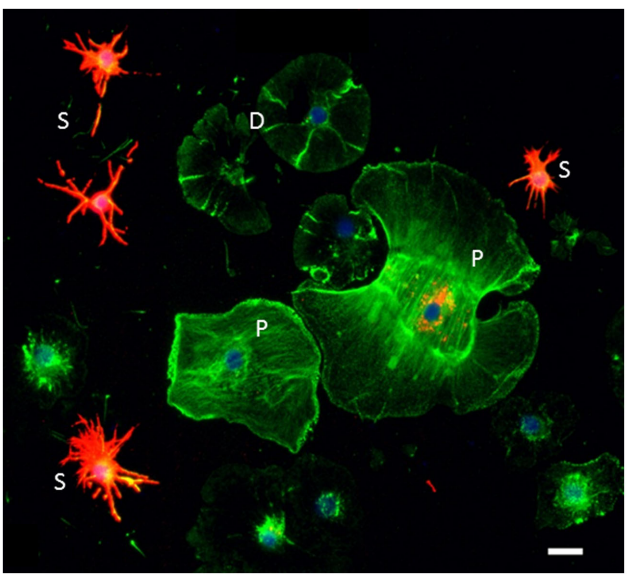


FIGURE 3 | Sp185/333 proteins are expressed by subsets of small phagocytes and polygonal phagocytes. Phagocytes are labeled for actin (green), Sp185/333 proteins (red), and DNA (blue). Small phagocytes (S) have different filopodial morphology and actin organization than the discoidal phagocytes (D) or polygonal phagocytes (P). A subset of small phagocytes are strongly labeled for Sp185/333 proteins. The larger polygonal cell has perinuclear vesicles that are Sp185/333^{+}. Bar = 10 μm. This is Figure 1D reproduced from Brockton et al. (2008).

sea urchin, there are three classes of coelomocytes: phagocytes, spherule cells, and vibratile cells (Smith et al., 2010; **Figure 2**). The phagocytes are composed of three morphotypes that vary in size and shape. The large phagocytes include the polygonal and discoidal cells that have different morphologies when spread on glass (Henson et al., 1992, 1999; Edds, 1993). The cytoskeleton of the polygonal phagocytes is shaped by parallel bundles of actin that give the cell an angular, polygonal shape. Discoidal phagocytes, which are slightly smaller than the polygonal cells, have bundles of actin that are aligned radially from the nucleus to the periphery resulting in a disk-like or "fried egg" morphology (**Figures 2** and **3**). Small phagocytes are significantly smaller than the large phagocytes and have perpetual filopodial morphology (Gross et al., 2000; Brockton et al., 2008; **Figures 2–4**). Large phagocytes, on the other hand, readily modify their morphology and appear as lamellipodial or filopodial depending on the microenvironmental conditions of the coelomic cavity (Edds, 1977, 1979, 1993; Henson et al., 1999). The spherule cells include cells with colorless inclusions and cells with red spherules that contain echinochrome A, which has antibacterial and antifungal activity (Johnson and Chapman, 1970; Messer and Wardlaw, 1980; Service and Wardlaw, 1984; Calestani et al., 2003). The vibratile cells have a single flagellum, are highly motile, and may be involved with clotting reactions (Johnson, 1969b; Bertheussen and Seijelid, 1978; Sacchi and Smith, unpublished). Immune functions mediated by coelomocytes include phagocytic activities of the large phagocytes, encapsulation by many types of cells, degranulation of echinochrome A from red spherule cells in response to bacterial contact (Johnson, 1969b; Johnson and Chapman, 1970) and injury (Coffaro and Hinegardner, 1977), and secretion of the complement C3 homolog by a subset of the discoidal cells (Gross et al., 2000; reviewed in Smith et al., 2010). Unfractionated coelomocytes from the sea urchin *Paracentrotus lividus*, show cytotoxic activity against red blood cells and K562 tumor cells, a function that may be mediated by the colorless spherule cells that is augmented by the presence of phagocytes (Arizza et al., 2007).

Early work to analyze gene expression in coelomocytes under conditions of immune challenge identified homologs of complement components including putative complement regulatory proteins, in addition to genes encoding transcription factors, lectins, ions channels, a Tie receptor homolog, lysosomal enzymes, cytoskeletal proteins, folding chaperones, mitochondrial enzymes, proteins that function in RNA splicing, signaling pathways, and secretion, plus a large number of unknowns (Smith et al., 1996, 1998; Al-Sharif et al., 1998; Rast et al., 2000; Multerer and Smith, 2004; Nair et al., 2005; Stevens et al., 2010). Many of the expressed sequence tags (ESTs) correlate with gene models annotated in the genome of the purple sea urchin, although the major finding from genome annotation was the striking level of complexity and sophistication of this invertebrate innate immune system (Hibino et al., 2006). For example, large gene families that are homologous to those in vertebrates that function in innate immune detection, such as the Toll-like receptors, NOD-like receptors and scavenger receptors are greatly expanded in the purple sea urchin (Rast et al., 2006; Rast and Messier-Solek, 2008). One gene family that is highly up-regulated in response to immune challenge is the *Sp185/333* family (Rast et al., 2000; Nair et al., 2005). This family has been estimated to have about 50 ± 10 gene loci using a variety of methods including quantitative PCR of genomic DNA, a statistical estimate based on the number of unique genes cloned and sequenced from three sea urchins, and an estimate from screens of two arrayed bacterial artificial chromosome (BAC) libraries (Terwilliger et al., 2006; Buckley et al., 2008a; Ghosh et al., 2010). BLAST analysis of 1025 randomly chosen cDNA clones that were up-regulated in response to lipopolysaccharide (LPS) showed that 73% matched to two sequences on GenBank; DD185 (Rast et al., 2000) and EST333 (Smith et al., 1996). Hence the name for the set of cDNAs, *Sp185/333*, which was chosen because none of the sequences matched to proteins with known function and homologs have not been identified in organisms outside of echinoids.

THE *Sp185/333* GENES

GENE STRUCTURE

Initial speculation on the *Sp185/333* gene structure was based on the extraordinary structure of the cDNAs (Nair et al., 2005). Optimal amino acid alignments of the translated cDNAs required artificial insertions of large gaps (see below), which seemed consistent with extensive alternative splicing of transcripts from one or a few large genes with many exons, similar to that for *Dscam* in arthropods (Schmucker et al., 2000; Watson et al., 2005; Dong et al., 2006; Brites et al., 2008; Schmucker and Chen, 2009). This prediction of large genes with many exons might have been consistent with Southern blots if all of the positive bands had been large (**Figure 5**). However, genomic DNA digested with *Pst*I released small fragments that were hybridized by probes from both the 5′ and 3′ ends of the cDNAs (Terwilliger et al., 2006). Furthermore, small amplicons were produced by PCR of genomic DNA using primers that hybridized to the ends of the cDNAs (Buckley and Smith, 2007). These results suggested that the cDNAs were expressed from many small genes rather than a few large genes with many exons. Accordingly, preliminary genome sequences from the purple sea urchin confirmed that the *Sp185/333* genes were small (1.2–2 kb) and had only two exons (Terwilliger et al., 2006; **Figure 6**). The first exon is very short (51 or 54 nucleotides),

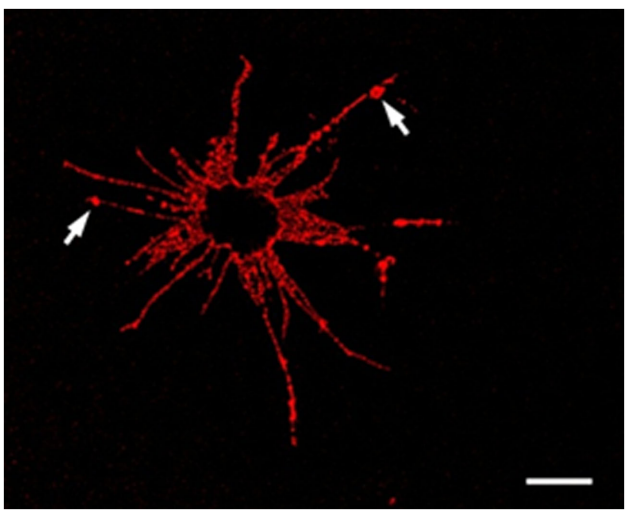


FIGURE 4 | An Sp185/333$^+$ Small Phagocyte. Confocal image of a Sp185/333$^+$ small phagocyte labeled with anti-Sp185/333 sera prior to fixation shows the filopodial morphology of the cell and indicates the presence of Sp185/333 proteins on the cell surface. This includes knobs on the filopodia that are strongly positive for Sp185/333 (arrows). The dark area in the center of the cell is the location of the nucleus. Bar = 10 μm. This is Figure 5B reproduced from Brockton et al. (2008).

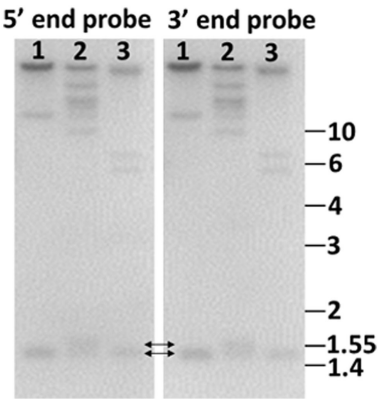


FIGURE 5 | Genome blot of *Sp185/333* genes indicates that the genes are small. Probes for the 5′ end and the 3′ end of the second exon hybridize to the same bands. This includes bands that are 1.4–1.55 kb (double headed arrows), demonstrating that the genes are small. Genomic DNA from three sea urchins (1, 2, 3) was digested to completion with *Pst*I and separated by electrophoresis. Duplicate gels were blotted onto nylon membranes and analyzed with ^{32}P-labeled riboprobes according to Terwilliger et al. (2006). Cloned templates used to generate the riboprobes were amplified by PCR from genomic DNA using primers that hybridized to elements 1 and 7 (5′ probe), and elements 7 and 25 (3′ probe). See **Figure 6A** for element positions. Size standards in kilobase are shown to the right. (Terwilliger and Smith, unpublished).

encodes the hydrophobic leader, and is followed by an intron of 380–413 nucleotides. The second exon encodes the mature protein and has the same unusual structure as the cDNAs, which is evident from alignments of the genes (**Figure 6**). Our initial speculation that alternative splicing generated the structures of the cDNAs is clearly improbable given that the genes only have two exons. Furthermore, splicing signals that might function within the intron or second exon to generate alternatively spliced transcripts are not present in the *Sp185/333* sequences (Buckley et al., 2009).

TWO ALIGNMENTS

The gaps that are required to align optimally the *Sp185/333* genes and cDNAs define recognizable blocks of sequence called *elements* (**Figure 6**). There are 25–27 different elements depending on the alignment that are shared among genes and cDNAs. They are variably present or absent within different genes and cDNAs resulting in recognizable mosaics of elements, called *element patterns*. Note that elements are not the result of alternative splicing during processing of the mRNAs, but are present within the second exon of the genes. Based on alignments of 121 unique gene sequences (of 171 sequenced genes), 31 different element patterns have been identified from three sea urchins (Buckley and Smith, 2007). When both genes and cDNA sequences from 16 individuals are evaluated together, 51 different element patterns can be identified. Although each type of element has a recognizable sequence, there are variants of each type that show sequence diversity. There can be between 1 and 28 (average = 11/element) different sequence versions for a given element based on genes cloned from individual animals (Buckley et al., 2008a). In addition

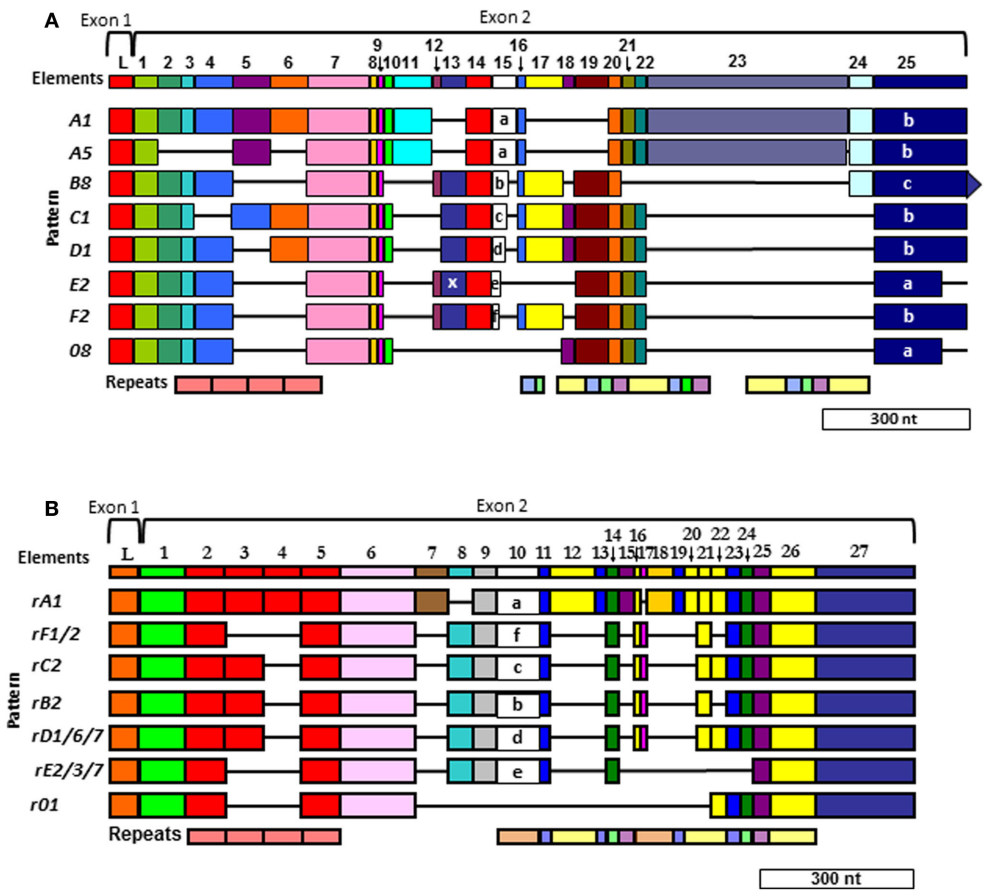


FIGURE 6 | Two different alignments for *Sp185/333* sequences are equally optimal. (A) The cDNA alignment was initially done with ESTs and full-length cDNA sequences (Terwilliger et al., 2006, 2007). **(B)** The Repeat-based alignment optimizes correspondence between repeats and elements whenever possible (Buckley and Smith, 2007). Optimal alignments require the insertion of artificial gaps (horizontal black lines) that delineate individual elements shown as colored blocks (the consensus of all elements are numbered across the top of each alignment; L, leader). Different element patterns are based on the variable presence or absence of elements. Designations of element patterns are listed to the left of each alignment. There are three types of element 25 in **(A)** (a, b, and c), which are defined by the location of three possible stop codons. A common, single nucleotide RNA edit in element pattern *E2* alters a glycine codon to a stop (X) in element 13 and is denoted as an *E2.1* pattern. Repeats, shown at the bottom of **(A,B)**, are shows as different colors and occur as tandem repeats and interspersed tandem repeats. The genes have two exons (brackets at the top of each alignment), of which the first encodes the leader and the second encodes the mature protein including all of the elements. A single intron of ~400 nt is positioned between the leader and the first element and is not shown. Modified from Ghosh et al. (2010) with permission from Elsevier.

to the complexity of the elements and element patterns, there are six types of repeats in the second exon of the genes (Nair et al., 2005; Terwilliger et al., 2006; Buckley et al., 2008a). Depending on the gene, there are two to four imperfect type 1 repeats near the 5′ end of the second exon that are direct, tandem repeats (**Figure 6**). Toward the 3′ end of the exon, there are five additional types of repeats that occur in an interspersed and partially duplicated pattern. Because of the repeats, the *Sp185/333* sequences can be aligned in two different ways that are equally optimal. Initially, the alignment was done with cDNA sequences and no attempt was made to correlate the elements with the repeats (Terwilliger et al., 2006, 2007; **Figure 6A**). The second alignment was done with both cDNA and gene sequences and efforts were made to make the repeats and elements correspond as much as possible (Buckley and Smith, 2007; **Figure 6B**). The two alignments are somewhat similar for the type 1 tandem repeats, but are quite different in the region of the interspersed repeats near the 3′ end of the exon with different numbers of interspersed repeats in the two alignments. Comparisons between the two alignments shows that some of the sequences could be interdigitated, essentially collapsing some of the repeats together, which reduced their numbers in the repeat-based alignment. (Compare the interspersed repeats shown in **Figures 6A,B**. For a detailed illustration, see Buckley and Smith, 2007 and Figure 7 within that publication.) Although the number of elements in the two alignments differ because of the collapse of some of the repeats (25 vs. 27 elements), the overall length of the two alignments is similar and both are about equally optimal (Buckley and Smith, 2007).

ORIGINS OF GENE STRUCTURE

The extraordinary sequence diversity of the *Sp185/333* gene family is due primarily to the mosaic element patterns in the second exon. Sequence variations within individual elements adds a second level of sequence diversity to the genes. The diverse structure of the genes is extraordinary and comparisons of full-length sequences among three sea urchins has not identified an identical shared gene (Buckley et al., 2008a). Yet surprisingly, 28% of individual element sequences are shared among genes from two or more animals (**Figure 7**). This paradox of element sequences shared among genes, but genes that are not shared among individuals, is quite remarkable and hints at an intriguing mechanism to generate the diversity within this gene family. Initial efforts to determine whether the type 1 repeats might be involved in generating diversity predicted three theoretical ancestral type 1 repeats (5′ end of the second exon; Buckley et al., 2008a). A computational estimate of the evolutionary history suggests that the extant type 1 repeats may have been derived from these ancestral sequences through modification by recombination, duplication, deletion, and point mutations. In the repeat-based alignment, the five types of interspersed repeats in the 3′ end of the exon are present in a block that is repeated twice with an extra type 4 repeat suggesting origins from duplication events (**Figure 6B**). Because there is no correlation between the patterns of tandem type 1 repeats and the patterns of the interspersed repeats in the 3′ end of the exon, this suggests swift recombination among genes plus recombination among the repeats within the genes. Several examples of

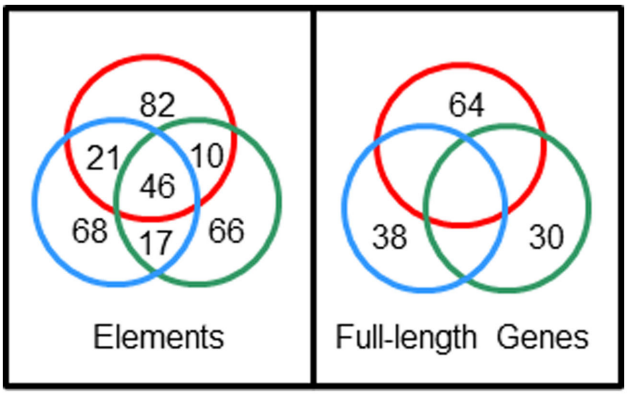


FIGURE 7 | *Sp185/333* gene sequences are not shared among sea urchins but element sequences are shared among genes. Full-length genes were cloned and sequenced from three sea urchins; 38 unique genes from animal 10 (blue circle), 64 unique genes from animal 2 (red circle), 30 unique genes from animal 4 (green circle). Nucleotide sequences were compared among full-length genes and among individual elements, and the numbers of shared unique sequences are shown in the intersections of the circles. Unpublished figure provided by Katherine Buckley.

putative gene recombination could be identified because different regions of a few genes matched optimally to corresponding regions of two other genes. Because the recombination rate among *Sp185/333* genes may be quite high, identifying the ancestral genes that contributed sequence to a given recombinant gene is quite difficult. Consequently, sequences evaluation to identify recombination sites or hotspots using incongruence length difference (ILD; Michevich and Ferris, 1981) and incongruence permutation (IP; Huelsenbeck and Bull, 1996) both show that recombination is likely to occur at any position within and between elements and repeats and that there are no recombination hotspots (Buckley et al., 2008a). The level of putative gene recombination among members of the *Sp185/333* gene family is swift, appearing surprisingly similar to somatic recombination of gene segments that is known to occur for T cell receptor genes (Siu et al., 1984). Given that these genes appear to encode proteins with immune effector function, ongoing diversification, based in part by gene recombination, will be advantageous for the host in the arms race with pathogens.

AGE ESTIMATE FOR THE GENE FAMILY

Given the apparently swift recombination of the *Sp185/333* genes, molecular clock analysis of the evolutionary history of the *Sp185/333* family suggests that the last common ancestral *Sp185/333* sequence of the extant genes may have been present in the population of purple sea urchins 2.7–10 million years ago (Ghosh et al., 2010). This time frame overlaps with the estimated time period when *Strongylocentrotus purpuratus* diverged from its sister species (Grula et al., 1982; Biermann, 1998; Lee, 2003; Buckley et al., 2008a). When trace sequences of *185/333* genes from the genomes of two other sea urchins, *S. franciscanus* and *Allocentrotus fragilis* are compared to the *Sp185/333* sequences from *S. purpuratus*, molecular clock analysis shows that the divergence times for these three gene families appear older than the

three species of sea urchins in which they function. This apparent paradox is another illustration of a rapid rate of evolution of the *185/333* genes in all three species, a characteristic that is consistent with immune function, participation in an arms race with pathogens, and mechanisms to promote sequence diversification. All three of these echinoid species live in the marine environment at the intersection of water and sediment. This habitat is estimated to have 10^5–10^6 microbes per milliliter of sea water with two or three orders of magnitude more in the sediment (Massana et al., 1997; Llobet-Brossa et al., 1998; Whitman et al., 1998). It is likely that some of the microbes that co-habit with sea urchins are pathogens or opportunists and function as selection pressure to drive the diversification of the *Sp185/333* gene family and the complexity of the innate immune system in these invertebrates.

SIZE AND ORGANIZATION OF THE *Sp185/333* GENE FAMILY

The three different approaches used to estimate the size of the *Sp185/333* family included qPCR of genomic DNA, a computational likelihood method to evaluate sequenced genes, and screens of two BAC libraries. Results from each evaluation was in general agreement and provided a family size of $\sim 50 \pm 10$ genes (Buckley et al., 2008a; Ghosh et al., 2010). Quantitative PCR of genomic DNA samples from three sea urchins indicated 40–60 genes (Terwilliger et al., 2006). A computational likelihood method evaluates the probability of sequencing the same gene twice given the total number of genes cloned and sequenced and resulted in an estimate of 45–71 genes (Buckley et al., 2008a). Screens of two BAC libraries for clones with *Sp185/333* sequences estimated a haploid genomic region of 200–250 kb in which the genes might be located, given the fold coverage of the genome for each library, the average insert size for the BACs, and the number of positive clones obtained (Buckley et al., 2008a). For a region of 200–250 kb, if all the genes (∼2 kb) in the haploid genome are closely linked and the intergenic region is ∼3 kb, this estimates ∼40–50 genes. However, evaluation of the purple sea urchin genome assembly (including build versions 0.5, 2.1, and 3.1) shows only five or six *Sp185/333* genes, depending on the version, that are positioned on a few separate scaffolds. Based on the level of sequence diversity that has been identified from 171 genes cloned and sequenced from three individuals (Buckley and Smith, 2007), six *Sp185/333* genes assembled in the genome is insufficient to explain this diversity, and does not agree with the our estimated number of ∼50 genes in the family. The underrepresentation of *Sp185/333* genes assembled in the genome is likely the result of a computational collapse of similar sequence reads into consensus genes based on the shortcomings of shotgun assembly methods that incorrectly assemble regions that appear as repeats (Sodergren et al., 2006; Ghosh et al., 2010; Miller et al., 2010).

To begin to understand the *Sp185/333* gene family structure, one BAC clone (GenBank accession number AC178508.1) was re-sequenced and independently assembled (GenBank accession number 7096) with careful attention given to avoiding artifacts introduced by the assembly methods (Miller et al., 2010). The 7096 version of the BAC assembly that was experimentally validated, shows six *Sp185/333* genes closely linked within 34 kb near one end of the BAC insert (**Figure 8**). Five of the genes are tightly clustered within 20 kb and the sixth is located 14 kb away. The outer genes are oriented in the same direction, while the four internal genes are oriented in the opposite direction with intergenic regions as short as 3.2 kb. The element patterns of the genes in the cluster are, in order, *A2*γ, *B8*β, three *D1*α and *E2*δ (Greek letters indicate intron class; see Buckley and Smith, 2007). These genes do not match the genes assembled in the sea urchin genome. If gene duplication and/or gene conversion functions in generating the diversity observed within the *Sp185/333* gene family, closely linked genes would be expected to show more sequence similarity than unlinked genes. However, the diversity of the linked genes on the 7096 BAC is not different from the diversity of 121 unique genes that have been randomly isolated and sequenced from three sea urchins (Miller et al., 2010). Proximity is not associated with sequence similarity, which also holds true for the three linked *D1* genes in the BAC compared to randomly cloned and sequenced *D1* genes from three different animals. This suggests that if gene conversion is involved with sequence diversification, genes may be sharing sequences from the entire family rather than just those that are most tightly linked and/or that conversion tracts may be shorter than the length of a gene.

With the availability of the sequence for the intergenic regions between the *Sp185/333* genes on the BAC assembly, additional oddities, and similarities were discovered. All six genes are flanked by microsatellites on both the 5′ side (30–60 GA repeats) and the 3′ side (140–165 GA repeats; **Figure 8**; Miller et al., 2010). In addition, there are GAT microsatellites that surround the *D1* genes and are positioned at the edges of three ∼13.5 kb tandem segmental duplications. These duplications have very low sequence diversity including the three *D1* genes that are 99.7% identical. This not only suggests a recent duplication event of the regions bounded by GAT microsatellites, but it also infers the potential of an unknown

FIGURE 8 | Six *Sp185/333* genes are linked on the 7096 BAC insert. The finished-level assembly of the region containing the *Sp185/333* genes was experimentally confirmed by PCR, pulsed field gel electrophoresis, *Ase*I digests, and subclone sequences. The six *Sp185/333* genes include one gene with the *A2* element pattern (red), one *B8* gene (orange), three *D1* genes (yellow, green, blue), and one *E2* gene (purple; see Figure for element patterns). All are located near the 3′ end of the BAC insert. Gene orientations are indicated and spacing is to scale unless otherwise noted. GA microsatellites flank each gene and GAT microsatellites flank segmental duplications that are positioned on 5′ side of *B8* and include the three *D1* genes. This is Figure 4 from Miller et al. (2010) reproduced with permission from BioMed Central.

number of identical *Sp185/333* genes in any given genome. Because previous estimates of family size (see above) have assumed that all genes have different sequences, the demonstration of genes on the 7096 BAC that are almost identical indicates that our estimates of family size are likely are likely to be too low.

Speculation on diversification mechanisms that act on the *Sp185/333* gene family is based on a number of unusual aspects of the clustered genes including sequence similarity, repeats, tight linkage, and microsatellites. Instability has been documented for regions of genomes that harbor tandem gene arrays, microsatellites, minisatellites, and other types of repeats (Gendrel et al., 2000; Wilson et al., 2000; Trifonov, 2003; Bagshaw et al., 2008; Smith, 2010; Despons et al., 2011). It is noteworthy that *D1* is the most common element pattern identified from randomly cloned and sequenced genes (Buckley and Smith, 2007), which corresponds with multiple *D1* genes on the BAC insert that are positioned within segmental duplications. This suggests that the *D1* genes in other regions of the genome may also be surrounded by GAT microsatellites that drive segmental duplications. The sequence between the GA microsatellites, which includes the *Sp185/333* genes and their flanking regions, has much lower sequence diversity than the intergenic regions positioned outside of the microsatellites (Miller et al., 2010). The sequence conservation may be driven by the presence of the microsatellites on either side of the genes plus element sequences that are shared within the second exons. The shared elements may promote gene recombination, conversion, and crossover, whereas the microsatellites may block the progression of DNA strand exchange during gene conversion, thereby protecting the region of the genome harboring clusters of *Sp185/333* genes from sequence homogenization (Gendrel et al., 2000; Miller et al., 2010). This model predicts that the exchanged sequences would be limited to the regions located between microsatellites, which would show higher sequence conservation. In general, the genomic structure of the family predicts mechanisms for sequence homogenization through gene duplication and/or conversion, yet maintains sequence diversity among the members of the family. Other families of clustered genes tend to generate pseudogenes, such as primate odorant receptors, sea urchin TLRs, and fish aquaporins (Gilad et al., 2003; Rast and Messier-Solek, 2008; Dong et al., 2009; Tingaud-Sequeira et al., 2010) however, only a single pseudogene has been identified from the 171 sequenced *Sp185/333* genes (Buckley and Smith, 2007). The *Sp185/333* gene family appears to employ intriguing mechanisms for sharing sequences, maintaining diversification, and blocking or correcting disruption of the open reading frame.

THE DIVERSITY OF THE *Sp185/333* mRNAs

The striking up-regulation of *Sp185/333* gene expression in response to a variety of immune challenges was the initial basis for our interest in investigating this immune response system. The genes are expressed in coelomocytes within 3–6 h of challenge from LPS, β-1,3-glucan, double stranded RNA (dsRNA), peptidoglycan (PDG), and heat-killed marine bacteria (Rast et al., 2000; Nair et al., 2005; Terwilliger et al., 2007). Blastocoelar cells, larval phagocytes in *S. purpuratus*, express the *Sp185/333* genes when cultured with marine microbes (Ghosh et al., 2010). The coding regions of the *Sp185/333* mRNAs range in size from 0.16 to 1.5 kb, depending on the element pattern. In addition to the wide variety of element patterns encoded in the genes, half of the mRNAs show changes to the reading frame, compared to the genes, and appear as single nucleotide polymorphisms (SNPs) that encode non-synonymous amino acids or introduce early stop codons. Furthermore, small indels shift the reading frame to missense sequence, typically ending with an early stop (Terwilliger et al., 2007; Ghosh et al., 2010). Unexpectedly, rather than being degraded as non-functional mRNAs, at least some are translated because both truncated and missense Sp185/333 proteins are produced (Dheilly et al., 2009). The most common element pattern observed in the cDNAs is *E2*, which is also present in a truncated version, *E2.1*, that has a SNP in element 13 that changes a glycine to a stop and deletes about half of the protein (Terwilliger et al., 2006, 2007; see **Figure 6A**). Prior to immune challenge, 81% of the sequenced cDNAs encode a truncated protein with the E2.1 pattern, whereas after challenge this is decreased to 58% (Ghosh et al., 2011). This variation can be interpreted as directed or regulated RNA editing rather than a change in gene expression because no gene has been identified with a stop codon in element 13.

Although many mRNAs encode truncated and missense proteins (Terwilliger et al., 2007), this is strikingly different from the sequenced genes of which all but one (of 171) have perfect open reading frames (Buckley and Smith, 2007). More detailed comparisons of genes and mRNAs from three individual sea urchins show a number of additional unexpected findings (Buckley et al., 2008b). The sequences of the genes and mRNAs both show high sequence diversity, which is present throughout and is not concentrated in diversity hotspots. However, for individual animals, the level of sequence diversity for genes is different from that for mRNAs. Similarly, the level of sequence diversity for elements that are shared between genes and mRNAs is also different. This means that the most variable nucleotide positions in the genes do not correspond with the most variable positions in the mRNAs. Very few of the mRNAs isolated from an individual animal match identically to any of the genes cloned from the same animal. Furthermore, estimates to identify the most likely gene as the source for sets of messages that share the same element pattern (based on the fewest number of substitutions and indels) show that 67–98% of the messages from individual sea urchins match best to a single gene with an *E2* element pattern. Comparisons between the variable positions in the gene and its most likely set of transcribed messages indicate that 73% of the changes are transitions, significantly higher than the expected 33% that would occur randomly. Furthermore, 30% of these transitions change a cytidine in the gene to uridine the mRNA. This is consistent with RNA editing by a cytidine deaminase, a family of enzymes for which several gene models have been identified in the genome (Hibino et al., 2006). In addition to possible RNA editing, transcription of the *Sp185/333* genes may be performed by a low fidelity polymerase, such as polμ, which has also been annotated in the genome (Hibino et al., 2006).

THE Sp185/333 PROTEINS

STRUCTURE

The structure of the Sp185/333 proteins, as deduced from the genes and mRNAs, have an N-terminal hydrophobic leader that is

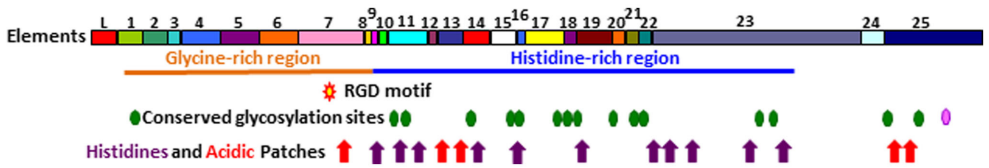


FIGURE 9 | Representation of Sp185/333 protein structure showing all possible elements. The deduced protein contains a leader (L), a glycine-rich region (orange line), and a histidine-rich region (blue line). Symbols indicate the presence of an RGD motif within element 7 (red star with a yellow center), N-linked glycosylation sites (green heptagons), an O-linked glycosylation site (pink heptagon), patches of acidic amino acids (red arrows), and histidines (purple arrows). Modified from Terwilliger et al. (2006).

encoded by the first exon and is assumed to be cleaved during processing. The mature proteins are encoded by the second exon and have a generic structure even though there is significant sequence diversity among them (Terwilliger et al., 2006; **Figure 9**). The N-terminal region of the mature, full-length proteins is glycine-rich and includes two to four imperfect tandem repeats, and the central region has an integrin binding motif; arginine–glycine–aspartic acid (RGD). In addition to the interspersed, imperfect, tandem repeats, the C-terminal region is histidine-rich with patches of poly-histidines that are interspersed with proline, glycine, arginine, and glutamine, and patches of acidic amino acids composed of short stretches of glutamic acid often combined with aspartic acid (Terwilliger et al., 2006). Up to three stop codons are present in conserved positions at the 3′ end of the genes, and alter the length of the C-terminus for different proteins. Conserved, predicted N-linked glycosylation sites are present throughout the length of the proteins, and O-linked sites are predicted in those sequences that employ the third stop codon and have the longest C-terminal region (Terwilliger et al., 2006, 2007). The deduced proteins have no recognizable transmembrane region, no conserved motif for a glycosylphosphatidylinositol (GPI) linkage, and no apparent means to be associated with membranes. The native proteins have no predictable secondary structure or folding characteristics, and consequently may be present in an extended, unfolded, or intrinsically disordered conformation. The amino acid sequence predictions from the genes indicates that the proteins have no cysteines, although frame shifts introduced by RNA editing predicts missense cysteines in some truncated proteins (Terwilliger et al., 2007), which have been confirmed for native Sp185/333 proteins by proteomic analysis (Dheilly et al., 2009; **Table 1**).

Table 1 | Mass spectrometric identification of native Sp185/333 proteins from coelomic fluid*.

Sp185/333 sequence	Element[1]
DFNERREKENDTERGQGGFGGRPGGMQMGGP	1–2
RGRGQGRFGGRPGGMQMGGPRQDGGPMG	1–2–3
GQGGFGGRPGGMQ[M/T]GG[P/L]R	2
RFDGPESGAPQM[E/G]GRRQNGVPMGGR	4
MGGRnstnPeFGGSRPDGAG[2]	6–7
(Overlap with above)FGGSRPDGAGGRPFFGQGGR[2]	6–7
RGDGEEETDAAQQIGDGLGGPGQFDG[P/H]GR	7–8–9–10
PQTDQRNNRLVSATKAAMRM[3]	13–14
MAVLTLATMAATTSIIIATTQKVTK[4]	14–15
KPFGDHPFGR	23
adVVEIAVNEEDVN[2]	25b

**Amino acid sequences of peptides identified by LC–MS/MS from gel slices after one dimensional separation of coelomic fluid proteins from three different sea urchins as reported in Dheilly et al. (2009). Modified from Table 1 in Dheilly et al. (2009).*

*[1] Element in which the sequence is present. For description of elements, see Buckley and Smith (2007), Terwilliger et al. (2007), and **Figure 6A**.*

[2] Exact matches were not found from the translated cDNAs reported by Terwilliger et al., 2007, see Supplemental Table 2). Lower case letters indicate mismatched amino acids between proteomic result and deduced sequence.

[3] Missense sequence matches to the amino acid sequence predicted from one cDNA (GenBank accession number; EF065834) reported by Terwilliger et al. (2007).

[4] Missense sequence matches to amino acid sequences predicted from five cDNAs (GenBank accession numbers; EF065834, EF065907, EF066020, EF066222, EF066214) reported by Terwilliger et al. (2007).

DIVERSITY

The combination of the mosaic structure of the elements plus RNA editing and the introduction of early stop codons results in a deduced size range of the Sp185/333 proteins of 4–55 kDa (Terwilliger et al., 2006, 2007). This includes varying numbers of histidine and acidic patches, the presence or absence of the RGD motif, and the deletion of the entire C-terminal histidine-rich region. However, the observed size range of the native proteins is unexpectedly large relative to deduced predictions, and likely results in part, from multimerization (Brockton et al., 2008). The presence of truncated and missense proteins in coelomic fluid is demonstrated on two dimensional (2D) Western blots by fewer spots for antisera that recognize the central and C-terminal regions of the proteins (Dheilly et al., 2009; **Figure 10**). Multimerization and putative post-translational modifications are likely underlying mechanisms for generating up to 264 different Sp185/333-positive ($Sp185/333^{+}$) spots for individual sea urchins (**Figure 11**). Variations in the arrays of Sp185/333 proteins evaluated from different animals are striking (**Figure 12**) and suggestive of variable gene expression, perhaps differences in the numbers and/or versions of the genes in different individuals, in addition to mRNA editing and post-translational modifications to the proteins. It is apparent that the array of proteins from individual sea urchins shows more

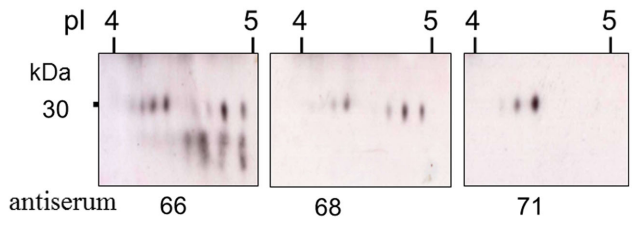


FIGURE 10 | Truncated Sp185/333 proteins are present in the coelomic fluid. Enlarged regions of three different two dimensional Western blots of coelomic fluid proteins from the same sea urchin were analyzed with different anti-Sp185/333 sera; anti-66, anti-68, or anti-71. Anti-66 recognizes AHAQRDFNERRGKENDTER from element 1; anti-68 recognizes GGRRGDGEEETDAAQQIGDGLC from element 7; anti-71 recognizes TEEGSPRRDGQRRPYGNR from element 25 (see **Figure 6A** for element positions). Decreasing numbers of spots in blots analyzed with antisera that recognize peptides in more C-terminal regions of the proteins suggests that many are either truncated or have missense sequence toward the C-terminus. Reprinted from Dheilly et al. (2009) with permission from the American Association of Immunologists, Inc., copyright 2009.

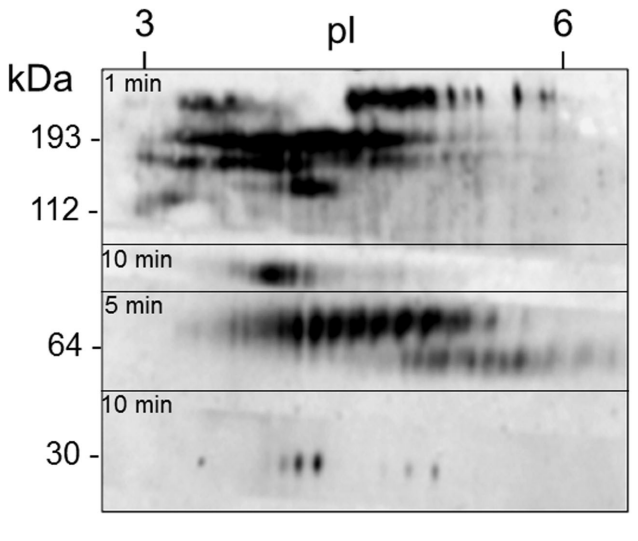


FIGURE 11 | A single sea urchin can have as many as 264 spots that are Sp185/333$^+$. Coelomic fluid proteins (200 μg) from sea urchin 12 (see **Figure 12**) were separated by 2D electrophoresis and transferred to a filter. The filter was immunostained with an equal mixture of the three anti-Sp185/333 sera (see legend to **Figure 10**) and exposed to autoradiographic film for 1, 5, or 10 min. The different exposures were merged to give a final composite image. Isoelectric points (pI) are shown at the top and the molecular weight standards (kDa) are shown to the left. Reprinted from Dheilly et al. (2009) with permission from the American Association of Immunologists, Inc., copyright 2009.

variation than can likely be encoded by a gene family of ∼50 members.

EXPRESSION

Subsets of the polygonal and small phagocytes express Sp185/333 proteins, which are present in the trans cisternae of the Golgi Apparatus, in perinuclear vesicles, and are associated with the

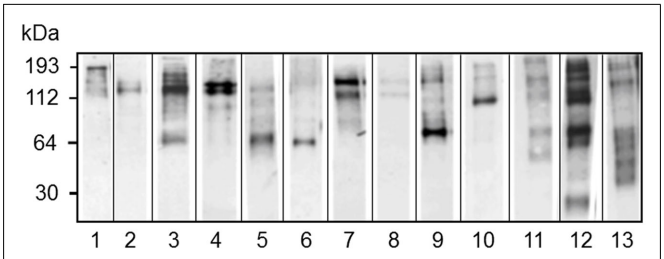


FIGURE 12 | Different sea urchins express different arrays of Sp185/333 proteins. Western blots of coelomic fluid from 13 different sea urchins sampled 96 h after challenge with LPS show different arrays of Sp185/333 proteins. The blots were immunostained with an equal mixture of the three different anti-Sp185/333 sera (see legend for **Figure 10**). Reprinted from Dheilly et al. (2009) with permission from the American Association of Immunologists, Inc., copyright 2009.

membrane of transport vesicles that appear to fuse with the plasma membrane and are present on the extracellular surface of the small phagocytes (Brockton et al., 2008; Dheilly et al., 2011; **Figures 3** and **4**). It is noteworthy and although the Sp185/333 proteins do not have a predicted transmembrane region, they are also associated with the membranes of internal organelles and vesicles, and with the plasma membrane. This localization may be mediated by interactions between the RGD motif and one or more of the integrin family members in the purple sea urchin (Whittaker et al., 2006) that are expressed in coelomocytes. A second sea urchin species, *Heliocidaris erythrogramma*, expresses He185/333 proteins, which are very similar to those in *S. purpuratus*, and are present in phagocytes and in colorless spherule cells (Dheilly et al., 2011).

FUNCTION

The putative immune function of the Sp185/333 proteins has been predicted based on the diversity of the genes, mRNAs, and proteins, and the striking up-regulation of gene expression in response to immunological challenge (Rast et al., 2000; Nair et al., 2005; Brockton et al., 2008; Dheilly et al., 2009). The amount of Sp185/333 protein in coelomic fluid increases in response to injection of LPS, which is reflected by an increase in the number of Sp185/333$^+$ cells (Brockton et al., 2008; Dheilly et al., 2009). These changes appear to be due, in part, to an increase in the number of small phagocytes that includes an increase in cells that are Sp185/333$^+$. On the other hand, while there is no increase in the number of polygonal cells in response to immune challenge, there is an increase in the percentage of Sp185/333$^+$ polygonal cells. Organs also have Sp185/333$^+$ cells including esophagus, gut, testis, ovary, pharynx, and axial organ (Majeske and Smith, unpublished). It is noteworthy that the number of Sp185/333$^+$ cells in the axial organ increases significantly in response to injections of LPS, whereas the other tissues do not show a similar response. The axial organ is located at the axis of the somewhat spherical echinoid (Hyman, 1955) and has no clearly defined function. However, cells in the axial organ of the sea star, *Asterias rubens*, proliferate in response to LPS and concanavalin A (Holm et al., 2008). Some gut associated cells in the sea urchin *H. erythrogramma*, are He185/333$^+$ and may be phagocytes (Dheilly et al.,

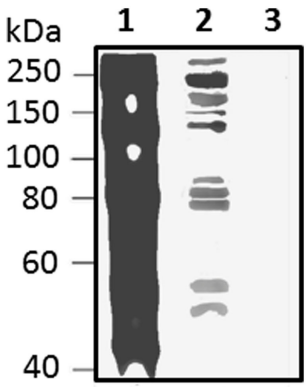


FIGURE 13 | Native Sp185/333 proteins bind *Vibrio diazotrophicus* (*Vd*), a gram negative marine bacterial species. Whole coelomic fluid (wCF) lysate from a sea urchin was incubated with *Vd*. Bacteria were pelleted, washed, and analyzed by Western blot using equal amounts of all three anti-Sp185/333 sera (see legend for **Figure 10**). wCF and *Vd* alone are shown for comparison. Lane 1, wCF; lane 2, wCF proteins bound to *Vd*; lane 3, *Vd*. Protein standard is to the left. Unpublished figure provided by Catherine Schrankel.

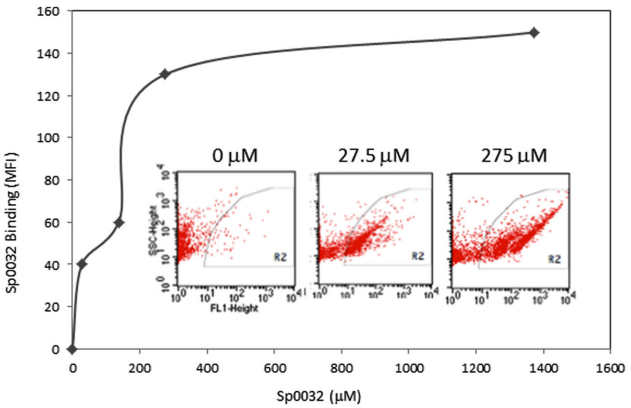


FIGURE 14 | A recombinant Sp185/333 protein (rSp0032) binds *Vibrio diazotrophicus* (*Vd*). Biotinylated rSp0032 incubated with 10^9 *Vd* cells and post-labeled with Neutravidin–FITC shows increased binding with increased protein concentration within the gate area (R2). Binding plateaus at about 400 mM of rSp0032 indicating the saturation point. MFI, mean fluorescence intensity. Results from flow cytometry (inserted images) show fluorescence (*X*-axis; fluorescent events measuring FITC) associated with bacterial cells (*Y*-axis; side scatter or cell counts) for increasing concentrations of rSp0032. Unpublished figure provided by Catherine Schrankel.

2011). To date, it appears Sp185/333$^+$ cells are present throughout echinoid tissues, but it is not known whether they are wandering coelomocytes or stromal cells of the organs. The increase in Sp185/333$^+$ cells and the increase in protein content post-immune challenge is consistent with an immune function of this protein family.

Speculation that the Sp185/333 proteins have antibacterial activity has been based on the diversity of the proteins and their expression kinetics in response to challenge from bacterial molecular patterns or from heat-killed marine gram negative bacterial species, *Vibrio diazotrophicus*. Native Sp185/333 proteins bind tightly to *V. diazotrophicus* (**Figure 13**), are not eluted by high or low pH, and in some cases, show an increase in size upon binding the target cells (Schrankel and Smith, unpublished). Perhaps the oligomerization that is commonly observed on Western blots is an aspect of binding to targets. Because predictions of Sp185/333 folding suggests that they may be unordered and unfolded, this has led to questions of whether the proteins conform to the target for binding and whether binding is specific. Testing these functions with a recombinant Sp185/333 protein shows that it binds to *V. diazotrophicus* and Baker's yeast in the absence of the coelomic fluid, but does not bind to the gram positive species, *Bacillus cereus* or *B. subtilis*. Binding to *V. diazotrophicus* appears to be saturable (**Figure 14**) and preliminary competition assays suggest that the binding site may be specific (Lun and Smith, unpublished). Future evaluation of different versions of recombinant Sp185/333 proteins with different element patterns, different numbers of repeats and histidine patches (see **Figure 9**), or that are truncated may have a variety of anti-pathogen functions besides binding such as opsonization and/or agglutination of bacteria to augment phagocytosis, plus bacteriostatic or bactericidal activity.

CONCLUSION

The rapid evolution of pathogens imposes significant burdens on metazoan immune systems to prevent or combat effectively all possible infections. Vertebrate adaptive immune systems employ somatic recombination of the immunoglobulin gene family (Hozumi and Tonegawa, 1976; Brack et al., 1978; Weigert et al., 1978) or assembly of the variable lymphocyte receptor genes (Nagawa et al., 2007; Rogozin et al., 2007; Kishishita et al., 2010) as mechanisms that enable specific recognition and neutralization of pathogens. Until recently, immune diversification mechanisms were thought to be restricted to the vertebrates and that invertebrates relied on innate immunity based entirely on receptors and antimicrobial peptides with broad specificities for pathogen detection and antibacterial capabilities. However, this may not be adequate for host protection, and several highly variable immune response gene families have been identified in a few invertebrate groups that may significantly expand capabilities for non-self detection and protection against pathogen attack (reviewed in Ghosh et al., 2011). The central question that emerges from evidence of invertebrate immune gene diversity is – what are the mechanisms that function to enhance diversification in these groups? DSCAM variability in arthropods is based on extensive alternative splicing (Watson et al., 2005; Brites et al., 2008; Schmucker and Chen, 2009; Lee et al., 2010), which may be function in hemocytes to generate proteins that bind specifically to the invading pathogen (Graveley, 2005; Dong and Dimopoulos, 2009). Some immune genes in invertebrates are members of multi-gene families and include those encoding the variable chitin binding proteins in protochordates (Cannon et al., 2004; Dishaw et al., 2008, 2010) and fibrinogen related proteins

(FRePs) in mollusks (Stout et al., 2009; Hanington and Zhang, 2010; Loker, 2010). Other invertebrate immune genes are single copy, such as *fuhc* in compound tunicates, in which the diversity is encoded in the many hundreds of alleles in the population and is enhanced by alternative splicing (De Tomaso et al., 2005; McKitrick and De Tomaso, 2010). Except for alternative splicing of *Dscam* mRNAs, diversification mechanisms beyond gene duplication are not understood for invertebrate immunity, and mechanisms to generate and/or maintain diversity may be different for each group of organisms or even for individual species. The beginnings for understanding invertebrate immune diversification has come from investigations of the *FReP3* gene subfamily (Loker et al., 2004; Zhang et al., 2004). The number of *FReP3* loci in the genome of an individual snail is insufficient to account for the level of sequence diversity observed in the *FReP3* mRNAs. Consequently, *FReP3* genes may undergo point mutations and/or somatic recombinatorial diversity, which is predicted to result from the concatenation of pairs of gene segments in the hemocytes.

Whether sea urchins have specific mechanisms to manipulate and diversify the members of the *Sp185/333* gene family, or whether diversification is the result of genomic instability and DNA repair, or both, is not known. However, multiple levels sequence diversification in the *Sp185/333* system appears to act on the genes, the mRNAs, and the proteins (**Figure 15**). Gene diversification may occur because region(s) of the genome that harbor clusters of *Sp185/333* genes (assuming that they have structure similar to that observed for the 7096 BAC sequence; **Figure 8**) are highly unusual due of the preponderance of various sized repeats. These include (i) large repeats in the form of segmental duplications and the genes themselves based on their sequence similarity and tight linkage, (ii) medium repeats that appear as shared elements and both tandem and interspersed repeats within the genes, and (iii) small repeats or microsatellites that tightly flank the genes in addition to simple repeats that also appear within the genes (Buckley and Smith, 2007; Miller et al., 2010). This range of clustered repeats may act together, perhaps synergistically, to drive genomic instability that appears as gene duplication, deletion, conversion, and recombination. This may change the sequences of the genes including changes in the mosaic patterns of the elements plus SNPs within the elements. In addition, the *Sp185/333* family may show size variation among individuals because the genomic structure that harbors the family may promote both meiotic mispairing of genes on sister chromatids and drive unequal crossovers. On the other hand, instability and diversification may be controlled so that gene or regional sequence homogenization from gene conversion is blocked from progressing through the microsatellites. Furthermore, only a single pseudogenes and no gene fragments have been detected from the 171 *Sp185/333* genes that have been cloned and sequenced, which suggests mechanisms to either block pseudogene formation or to correct or delete them once formed.

The diversity that is encoded within the members of the gene family is expanded by apparent RNA editing that changes the nucleotide sequence at single positions to alter codons and insert stops, in addition to introducing small indels that change the reading frame to encode missense sequence (**Figure 15**). Although RNA editing is typically observed for a single or a few nucleotide positions in other systems (Chan et al., 1997; Maas and Rich, 2000), editing of the *Sp185/333* mRNAs is a combination of random edits throughout the entire length plus site specific edits to certain nucleotides. The overall result broadens the array of expressed full-length proteins to include some that are altered at single amino acids, some with missense sequence, and many that are truncated. This is a noteworthy deviation from the central dogma in biology in which the proteins are only partially encoded by the genes.

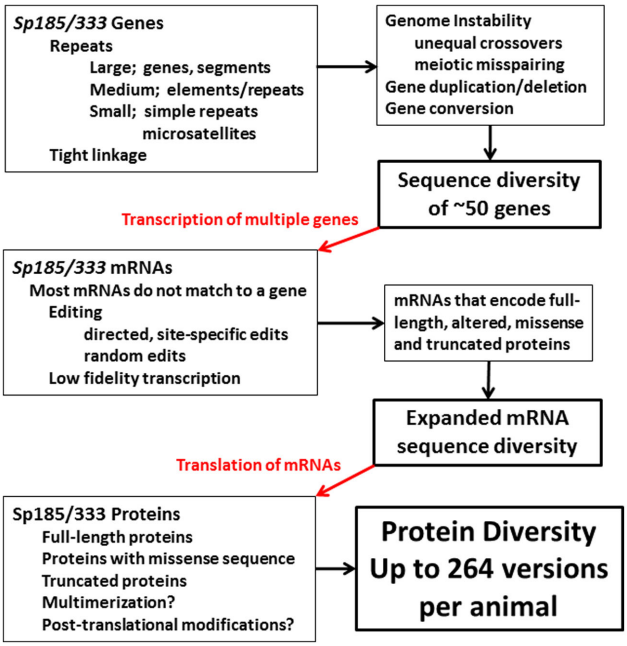


FIGURE 15 | Diversification in the *Sp185/333* system in the purple sea urchin. Putative diversification mechanisms based on genomic instability may function to increase gene sequence diversity and may also vary the size of the gene family among individuals. mRNA sequence diversity imparted by the genes is expanded by editing some of the mRNAs in addition to possible low fidelity transcription. The mRNAs are translated to generate a broad array of proteins that may multimerize and be post-translationally modified. The end result is protein diversity that is much broader than what is encoded by the genes.

The Sp185/333 proteins are larger than predicted from the cDNAs and genes suggesting post-translational diversification (**Figure 15**). This is evident for both native proteins and for individual recombinant proteins expressed in bacteria or insect cells (Brockton et al., 2008). The proteins may multimerize with themselves and/or with other Sp185/333 variants, both full-length and truncated, and possibly with other proteins in the coelomic fluid. Post-translational modifications may also act to diversify the proteins, perhaps as an underlying mechanism to produce a preponderance of acidic proteins (Dheilly et al., 2009) rather than the predicted full spectrum of isoelectric points. Given this dizzying array of putative diversification mechanisms, one must keep in mind that the proteins functioning in the immune

response of the sea urchin must withstand selection pressures from pathogens and not be altered to autoreactive forms. The arms race between pathogen virulence and host self protection takes different forms in this coevolutionary process and its direction depends on the capabilities of the pathogens to evade or counteract the immune system and the diversification mechanisms employed by the host to identify, deter, or destroy the invaders.

ACKNOWLEDGMENTS

The author is appreciative of the preliminary and unpublished results provided by Katherine Buckley, Catherine Schrankel, and Cheng Man Lun. Julie Ghosh, Cheng Man Lun, and three reviewers kindly provided improvements to the manuscript. Lauren Sherman, Preethi Golconda, and Hung-Yen Chou helped with the early organization of the information for this review. Work on sea urchin immunology is supported by the National Science Foundation.

REFERENCES

Al-Sharif, W. Z., Sunyer, J. O., Lambris, J. D., and Smith, L. C. (1998). Sea urchin coelomocytes specifically express a homologue of the complement component C3. *J. Immunol.* 160, 2983–2997.

Arizza, V., Giaramita, F. T., Parrinello, D., Cammarata, M., and Parrinello, N. (2007). Cell cooperation in coelomocyte cytotoxic activity of *Paracentrotus lividus* coelomocytes. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 147, 389–394.

Bagshaw, A. T., Pitt, J. P., and Gemmell, N. J. (2008). High frequency of microsatellites in *S. cerevisiae* meiotic recombination hotspots. *BMC Genomics* 9, 49. doi:10.1186/1471-2164-9-49

Bertheussen, K., and Seijelid, R. (1978). Echinoid phagocytes in vitro. *Exp. Cell Res.* 111, 401–412.

Biermann, C. H. (1998). The molecular evolution of sperm binding in six species of sea urchins (Echinoidea: Strongylocentrotidae). *Mol. Biol. Evol.* 15, 1761–1771.

Booloottian, R. A., and Geise, C. A. (1958). Coelomic corpuscles of echinoderms. *Biol. Bull.* 15, 53–56.

Brack, C., Hirama, M., Lenhard-Schuuler, R., and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. *Cell* 15, 1–14.

Brites, D., Mctaggart, S., Morris, K., Anderson, J., Thomas, K., Colson, I., Fabbro, T., Little, T. J., Ebert, D., and Du Pasquier, L. (2008). The Dscam homologue of the crustacean Daphnia is diversified by alternative splicing like in insects. *Mol. Biol. Evol.* 25, 1429–1439.

Brockton, V., Henson, J. H., Raftos, D. A., Majeske, A. J., Kim, Y. O., and Smith, L. C. (2008). 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.* 121, 339–348.

Buckley, K. M., Florea, L. D., and Smith, L. C. (2009). 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* 10, 318. doi:10.1186/1471-2164-10-318

Buckley, K. M., Munshaw, S., Kepler, T. B., and Smith, L. C. (2008a). The 185/333 gene family is a rapidly diversifying host-defense gene cluster in the purple sea urchin, *Strongylocentrotus purpuratus*. *J. Mol. Biol.* 379, 912–928.

Buckley, K. M., Terwilliger, D., and Smith, L. C. (2008b). Sequence variations in 185/333 messages from the purple sea urchin suggest posttranscriptional modifications to increase immune diversity. *J. Immunol.* 181, 8585–8594.

Buckley, K. M., and Smith, L. C. (2007). Extraordinary diversity among members of the large gene family, 185/333, from the purple sea urchin, *Strongylocentrotus purpuratus*. *BMC Mol. Biol.* 8, 68. doi:10.1186/1471-2199-8-68

Calestani, C., Rast, J. P., and Davidson, E. H. (2003). Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* 130, 4587–4596.

Cannon, J. P., Haire, R. N., Schnitker, N., Mueller, M. G., and Litman, G. W. (2004). Individual protochordates have unique immune-type receptor repertoires. *Curr. Biol.* 14, R465–R466.

Chan, L., Chang, B. H.-J., Makamuta, M., Li, W.-H., and Smith, L. C. (1997). Apobec-1 and apolipoprotein B mRNA editing. *Biochim. Biophys. Acta* 1345, 11–26.

Coffaro, K. A. (1979). *Memory and Specificity in the Sea Urchin Lytechinus pictus*. Doctoral Dissertation, University of California, Santa Cruz.

Coffaro, K. A. (1980). *Memory and Specificity in the Sea Urchin Lytechinus pictus*. New York, NY: Elsevier/North Holland Biomedical Press.

Coffaro, K. A., and Hinegardner, R. T. (1977). Immune response in the sea urchin *Lytechinus pictus*. *Science* 197, 1389–1390.

De Tomaso, A. W., Nyholm, S. V., Palmeri, K. J., Ishizuka, K. J., Ludington, W. B., Mitchel, K., and Weissman, I. L. (2005). Isolation and characterization of a protochordate histocompatibility locus. *Nature* 438, 454–945.

Despons, L., Uzunov, Z., and Louis, V. L. (2011). Tandem gene arrays, plastic chromosomal organizations. *C. R. Biol.* 334, 639–646.

Dheilly, N., Nair, S. V., Birch, D., and Raftos, D. A. (2011). 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.* 89, 861–869.

Dheilly, N. M., Nair, S. V., Smith, L. C., and Raftos, D. A. (2009). Highly variable immune-response proteins (185/333) from the sea urchin *Strongylocentrotus purpuratus*: proteomic analysis identifies diversity within and between individuals. *J. Immunol.* 182, 2203–2212.

Dishaw, L., Ota, T., Mueller, M., Cannon, J., Haire, R., Gwatney, N., Litman, R., and Litman, G. (2010). The basis for haplotype complexity in VCBPs, an immune-type receptor in amphioxus. *Immunogenetics* 62, 623–631.

Dishaw, L. J., Mueller, M. G., Gwatney, N., Cannon, J. P., Haire, R. N., Litman, R. T., Amemiya, C. T., Ota, T., Rowen, L., Glusman, G., and Litman, G. W. (2008). Genomic complexity of the variable region-containing chitin-binding proteins in amphioxus. *BMC Genet.* 9, 78. doi:10.1186/1471-2156-9-78

Dong, D., He, G., Zhang, S., and Zhang, Z. (2009). Evolution of olfactory receptor genes in primates dominated by birth-and-death process. *Genome Biol. Evol.* 1, 258–264.

Dong, Y., and Dimopoulos, G. (2009). Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *J. Biol. Chem.* 284, 9835–9844.

Dong, Y., Taylor, H. E., and Dimopoulos, G. (2006). AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol.* 4, e229. doi:10.1371/journal.pbio.0040229

Edds, K. T. (1977). Dynamic aspects of filopodial formation by reorganization of microfilaments. *J. Cell Biol.* 73, 479–491.

Edds, K. T. (1979). Isolation and characterization of two forms of a cytoskeleton. *J. Cell Biol.* 83, 109–115.

Edds, K. T. (1993). Cell biology of echinoid coelomocytes. *J. Invertebr. Biol.* 61, 173–178.

Furukawa, R., Takahashi, Y., Kanajima, Y., Dan-Sohkawa, M., and Kaneko, H. (2009). Defense system by mesenchyme cells in bipinnaria larvae of the starfish, *Asterina pectinifera*. *Dev. Comp. Immunol.* 33, 205–215.

Gendrel, C.-G., Boulet, A., and Dutreix, M. (2000). (CA/GT)n microsatellites affect homologous recombination during yeast meiosis. *Genes Dev.* 14, 1261–1268.

Ghosh, J., Lun, C. M., Majeske, A. J., Sacchi, S., Schrankel, C. S., and Smith, L. C. (2011). Invertebrate immune diversity. *Dev. Comp. Immunol.* 35, 959–974.

Ghosh, J. G., Buckley, K. M., Nair, S. V., Raftos, D. A., Miller, C. A., Majeske, A. J., Hibino, T., Rast, J. P., Roth, M., and Smith, L. C. (2010). Sp185/333: a novel family of genes and proteins involved in the purple sea urchin immune response. *Dev. Comp. Immunol.* 34, 235–245.

Gilad, Y., Man, O., Paabo, S., and Lancet, D. (2003). Human specific loss of olfactory receptor genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3324–3327.

Graveley, B. R. (2005). Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* 123, 65–73.

Gross, P. S., Clow, L. A., and Smith, L. C. (2000). SpC3, the complement homologue from the purple sea urchin, *Strongylocentrotus purpuratus*, is expressed in two subpopulations of the phagocytic coelomocytes. *Immunogenetics* 51, 1034–1044.

Grula, J. W., Hall, T. J., Hunt, J. A., Guiugni, D. F., Graham, G. J., Davidson, E. H., and Britten, R. J. (1982).

Sea urchin DNA sequence variation and reduced interspecies differences of the less variable DNA sequences. *Evolution* 36, 655–676.

Hanington, P. C., and Zhang, S.-M. (2010). The primary role of fibrinogen-related proteins in invertebrates is defense, not coagulation. *J. Innate Immun.* 3, 17–27.

Henson, J. H., Nesbitt, D., Wright, B. D., and Scholey, J. M. (1992). Immunolocalization of kinesin in sea urchin coelomocytes. Association of kinesin with intracellular organelles. *J. Cell. Sci.* 103(Pt 2), 309–320.

Henson, J. H., Svitkina, T. M., Burns, A. R., Hughes, H. E., Macpartland, K. J., Nazarian, R., and Borisy, G. G. (1999). Two components of actin-based retrograde flow in sea urchin coelomocytes. *Mol. Biol. Cell* 10, 4075–4090.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A., Terwilliger, D. P., Buckley, K. M., Brockton, V., Nair, S., Berney, K., Fugmann, S. D., Anderson, M. K., Pancer, Z., Cameron, R. A., Smith, L. C., and Rast, J. P. (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* 300, 349–365.

Hildemann, W. H., and Dix, T. G. (1972). Transplantation reactions of tropical Australian echinoderms. *Transplantation* 14, 624–633.

Holm, K., Dupont, S., Skold, H., Stenius, A., Thorndyke, M., and Hernroth, B. (2008). Induced cell proliferation in putative haematopoietic tissues of the sea star, *Asterias rubens* (L.). *J. Exp. Biol.* 211, 2551–2558.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U.S.A.* 73, 3628–3632.

Huelsenbeck, J. P., and Bull, J. J. (1996). A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* 45, 92–98.

Hyman, L. H. (1955). *The Invertebrates: Echinodermata, the Coelomate Bilateria*, Vol. 4. New York, NY: McGraw-Hill Book Co., Inc.

Johnson, P. T. (1969a). The coelomic elements of sea urchins (*Strongylocentrotus*). I. The normal coelomocytes; their morphology and dynamics in hanging drops. *J. Invertebr. Pathol.* 13, 25–41.

Johnson, P. T. (1969b). The coelomic elements of the sea urchins (*Strongylocentrotus*) III. In vitro reaction to bacteria. *J. Invertebr. Pathol.* 13, 42–62.

Johnson, P. T., and Chapman, F. A. (1970). Abnormal epithelial growth in sea urchin spines (*Strongylocentrotus franciscanus*). *J. Invertebr. Pathol.* 16, 116–122.

Karp, R. D., and Hildemann, W. H. (1976). Specific allograft reactivity in the sea star *Dermasterias imbricata*. *Transplantation* 22, 434–439.

Kishishita, N., Matsuno, T., Takahashi, Y., Takaba, H., Nishizumi, H., and Nagawa, F. (2010). Regulation of antigen-receptor gene assembly in hagfish. *EMBO Rep.* 11, 126–132.

Lee, C., Kim, N., Roy, M., and Graveley, B. R. (2010). Massive expansions of Dscam splicing diversity via staggered homologous recombination during arthropod evolution. *RNA* 16, 91–105.

Lee, Y. H. (2003). Molecular phylogenies and divergence times of sea urchin species of Strongylocentrotidae, Echinoidea. *Mol. Biol. Evol.* 20, 1211–1221.

Llobet-Brossa, E., Rossello-Mora, R., and Amann, R. (1998). Microbial community composition of Wadden sea sediments as revealed by fluorescence in situ hybridization. *Appl. Environ. Microbiol.* 64, 2691–2696.

Loker, E. S. (2010). "Gastropod immunobiology," in *Invertebrate Immunity*, ed. K. Soderhall (Austin: Landes Bioscience and Springer Science+Business Media), 17–43.

Loker, E. S., Adema, C. M., Zhang, S.-M., and Kepler, T. B. (2004). Invertebrate immune systems – not homogeneous, not simple, not well understood. *Immunol. Rev.* 198, 10–24.

Maas, S., and Rich, A. (2000). Changing genetic information through RNA editing. *Bioessays* 22, 790–802.

Massana, R., Murray, A. E., Preston, C. M., and Delong, E. F. (1997). Vertical distribution and phylogenetic characterization of marine palnktonic Archaea in the Santa Barbara chanel. *Appl. Environ. Microbiol.* 63, 50–56.

McKitrick, T. R., and De Tomaso, A. W. (2010). Molecular mechanisms of allorecognition in a basal chordate. *Semin. Immunol.* 22, 34–38.

Messer, L. I., and Wardlaw, A. C. (1980). "Separation of coelomocytes of *Echinus esculentus* by density gradient centrifugation," in *Proceedings of the European Colloquium on Echinoderms*, ed. M. Jangoux (Rotterdam: A. A. Balkema), 319–323.

Metchnikoff, E. (1893). *Lectures on the Comparative Pathology of Inflammation: Delivered at the Pasteur Institute in 1891*. Kegan Paul, Trench, Trubner & Co. Ltd, London.

Michevich, M. F., and Ferris, J. S. (1981). The implications of congruence in Mendida. *Syst. Zool.* 30, 351–370.

Miller, C. A., Buckley, K. M., Easley, R. L., and Smith, L. C. (2010). An Sp185/333 gene cluster from the purple sea urchin and putative microsatellite-mediated gene diversification. *BMC Genomics* 11, 575. doi:10.1186/1471-2164-11-575

Multerer, K. A., and Smith, L. C. (2004). 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* 56, 89–106.

Nagawa, F., Kishishita, N., Shimizu, K., Hirose, S., Miyoshi, M., Nezu, J., Nishimura, T., Nishizumi, H., Takahashi, Y., Hashimoto, S., Takeuchi, M., Miyajima, A., Takemori, T., Otsuka, A. J., and Sakano, H. (2007). Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. *Nat. Immunol.* 8, 206–213.

Nair, S. V., Del Valle, H., Gross, P. S., Terwilliger, D. P., and Smith, L. C. (2005). Macroarray analysis of coelomocyte gene expression in response to LPS in the sea urchin. Identification of unexpected immune diversity in an invertebrate. *Physiol. Genomics* 22, 33–47.

Plytycz, B., and Seljelid, R. (1993). Bacterial clearance by the sea urchin, *Strongylocentrotus droebachiensis*. *Dev. Comp. Immunol.* 17, 283–289.

Rast, J. P., and Messier-Solek, C. (2008). Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol. Bull.* 214, 274–283.

Rast, J. P., Pancer, Z., and Davidson, E. H. (2000). New approaches towards an understanding of deuterostome immunity. *Curr. Top. Microbiol. Immunol.* 248, 3–16.

Rast, J. P., Smith, L. C., Loza-Coll, M., Hibino, T., and Litman, G. W. (2006). Genomic insights into the immune system of the sea urchin. *Science* 314, 952–956.

Reinisch, C. L., and Bank, F. B. (1971). Cell recognition: reactions of the sea star (*Asterias vulgaris*) to the injection of amebocytes of sea urchin (*Arbacia punctulata*). *Cell. Immunol.* 2, 496–503.

Rogozin, I. B., Iyer, L. M., Liang, L., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L., and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat. Immunol.* 8, 647–656.

Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals yet simple genes. *Genes Dev.* 23, 147–156.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, J. E., and Zipursky, S. L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101, 671–684.

Service, M., and Wardlaw, A. C. (1984). Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L). *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 79, 161–165.

Silva, J. R. (2000). The onset of phagocytosis and identity in the embryo of *Lytechinus variegatus*. *Dev. Comp. Immunol.* 24, 733–739.

Siu, G., Clark, S. P., Yoshikai, Y., Malissen, M., Yanagi, Y., Strauss, E., Mak, T. W., and Hood, L. (1984). The human T cell antigen receptor is encoded by variable, diversity, and joining gene segments that rearrange to generate a complete V gene. *Cell* 37, 393–401.

Smith, L. C. (2010). Diversification of innate immune genes: lessons from the purple sea urchin. *Dis. Model Mech.* 3, 274–279.

Smith, L. C., Chang, L., Britten, R. J., and Davidson, E. H. (1996). 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.* 156, 593–602.

Smith, L. C., and Davidson, E. H. (1992). The echinoid immune system and the phylogenetic occurrence of immune mechanisms in deuterostomes. *Immunol. Today* 13, 356–362.

Smith, L. C., Ghosh, J., Buckley, K. M., Clow, L. A., Dheilly, N. M., Haug, T., Henson, J. H., Li, C., Lun, C. M., Majeske, A. J., Matranga, V., Nair, S. V., Rast, J. P., Raftos, D. A., Roth, M., Sacchi, S., Schrankel, C. S., and Stensvag, K. (2010). "Echinoderm immunity," in *Invertebrate Immunity*, ed. K. Soderhall (Austin: Landes Bioscience and Spring Science+Business Media), 260–301.

Smith, L. C., Shih, C. S., and Dachenhausen, S. G. (1998). Coelomocytes express SpBf, a homologue of factor B, the second component in the

sea urchin complement system. *J. Immunol.* 161, 6784–6793.

Smith, V. J. (1981). "The echinoderms," in *Invertebrate Blood Cells*, eds N. A. Ratcliffe and A. F. Rowley (New York, NY: Academic Press), 513–562.

Sodergren, E., Weinstock, G. M., Davidson, E. H., Cameron, R. A., Gibbs, R. A., Angerer, R. C., Angerer, L. M., Arnone, M. I., Burgess, D. R., Burke, R. D., Coffman, J. A., Dean, M., Elphick, M. R., Ettensohn, C. A., Foltz, K. R., Hamdoun, A., Hynes, R. O., Klein, W. H., Marzluff, W., Mcclay, D. R., Morris, R. L., Mushegian, A., Rast, J. P., Smith, L. C., Thorndyke, M. C., Vacquier, V. D., Wessel, G. M., Wray, G., Zhang, L., Elsik, C. G., Ermolaeva, O., Hlavina, W., Hofmann, G., Kitts, P., Landrum, M. J., Mackey, A. J., Maglott, D., Panopoulou, G., Poustka, A. J., Pruitt, K., Sapojnikov, V., Song, X., Souvorov, A., Solovyev, V., Wei, Z., Whittaker, C. A., Worley, K., Durbin, K. J., Shen, Y., Fedrigo, O., Garfield, D., Haygood, R., Primus, A., Satija, R., Severson, T., Gonzalez-Garay, M. L., Jackson, A. R., Milosavljevic, A., Tong, M., Killian, C. E., Livingston, B. T., Wilt, F. H., Adams, N., Belle, R., Carbonneau, S., Cheung, R., Cormier, P., Cosson, B., Croce, J., Fernandez-Guerra, A., Geneviere, A. M., Goel, M., Kelkar, H., Morales, J., Mulner-Lorillon, O., Robertson, A. J., Goldstone, J. V., Cole, B., Epel, D., Gold, B., Hahn, M. E., Howard-Ashby, M., Scally, M., Stegeman, J. J., Allgood, E. L., Cool, J., Judkins, K. M., Mccafferty, S. S., Musante, A. M., Obar, R. A., Rawson, A. P., Rossetti, B. J., Gibbons, I. R., Hoffman, M. P., Leone, A., Istrail, S., Materna, S. C., Samanta, M. P., Stolc, V., Tongprasit, W., Tu, Q., Bergeron, K. F., Brandhorst, B. P., Whittle, J., Berney, K., Bottjer, D. J., Calestani, C., Peterson, K., Chow, E., Yuan, Q. A., Elhaik, E., Graur, D., Reese, J. T., Bosdet, I., Heesun, S., Marra, M. A., Schein, J., Anderson, M. K., Brockton, V., Buckley, K. M., Cohen, A. H., Fugmann, S. D., Hibino, T., Loza-Coll, M., Majeske, A. J., Messier, C., Nair, S. V., Pancer, Z., Terwilliger, D. P., Agca, C., Arboleda, E., Chen, N., Churcher, A. M., Hallböök, F., Humphrey, G. W., Idris, M. M., Kiyama, T., Liang, S., Mellott, D., Mu, X., Murray, G., Olinski, R. P., Raible, F., Rowe, M., Taylor, J. S., Tessmar-Raible, K., Wang, D., Wilson, K. H., Yaguchi, S., Gaasterland, T., Galindo, B. E., Gunaratne, H. J., Juliano, C., Kinukawa, M., Moy, G. W., Neill, A. T., Nomura, M., Raisch, M., Reade, A., Roux, M. M., Song, J. L., Su, Y. H., Townley, I. K., Voronina, E., Wong, J. L., Amore, G., Branno, M., Brown, E. R., Cavalieri, V., Duboc, V., Duloquin, L., Flytzanis, C., Gache, C., Lapraz, F., Lepage, T., Locascio, A., Martinez, P., Matassi, G., Matranga, V., Range, R., Rizzo, F., Röttinger, E., Beane, W., Bradham, C., Byrum, C., Glenn, T., Hussain, S., Manning, G., Miranda, E., Thomason, R., Walton, K., Wikramanayke, A., Wu, S. Y., Xu, R., Brown, C. T., Chen, L., Gray, R. F., Lee, P. Y., Nam, J., Oliveri, P., Smith, J., Muzny, D., Bell, S., Chacko, J., Cree, A., Curry, S., Davis, C., Dinh, H., Dugan-Rocha, S., Fowler, J., Gill, R., Hamilton, C., Hernandez, J., Hines, S., Hume, J., Jackson, L., Jolivet, A., Kovar, C., Lee, S., Lewis, L., Miner, G., Morgan, M., Nazareth, L. V., Okwuonu, G., Parker, D., Pu, L. L., Thorn, R., and Wright, R. (2006). The genome of the sea urchin, *Strongylocentrotus purpuratus. Science* 314, 941–952.

Stevens, M. E., Dhillon, J., Miller, C. A., Messier-Solek, C., Rast, J. P., Majeske, A. J., Zuelke, D., and Smith, L. C. (2010). SpTie1/2 from the sea urchin *Strongylocentrotus purpuratus*, is an orthologue of vertebrate Tie1 and Tie2 and is expressed in coelomocytes, axial organ and embryos. *Dev. Comp. Immunol.* 34, 884–895.

Stout, B. A., Adema, C. M., Zhang, S. M., and Loker, E. S. (2009). "The biology of FREPs: diversified lectins with fibrinogen-related domains from the freshwater snail *Biomphalaria glabrata*," in *Animal Lectins: A Functional View*, ed. V. A. Ahmen (New York: CRC Press, Taylor & Francis), 475–491.

Terwilliger, D. P., Buckley, K. M., Brockton, V., Ritter, N. J., and Smith, L. C. (2007). 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.* 8, 16. doi:10.1186/1471-2199-8-16

Terwilliger, D. P., Buckley, K. M., Mehta, D., Moorjani, P. G., and Smith, L. C. (2006). Unexpected diversity displayed in cDNAs expressed by the immune cells of the purple sea urchin, *Strongylocentrotus purpuratus. Physiol. Genomics* 26, 134–144.

Tingaud-Sequeira, A., Calusinska, M., Finn, R. M., Chauvigne, F., Lozano, J., and Cerda, J. (2010). The zebrafish genome encodes the largest vertebrate repertoire of functional aquaporins with dual paralogy and substrate specificities similar to mammals. *BMC Evol. Biol.* 10, 38. doi:10.1186/1471-2148-10-38

Trifonov, E. N. (2003). "Tuning function of tandemly repeating sequences: a molecular device for fast adaptation," in *Evolutionary Theory and Processes: Modern Horizons, Papers in Honor of Eviatar Nevo*, ed. S. P. Wassser (Amsterdam: Kluwer Academic Publishers), 1–24.

Varadarajan, J., and Karp, R. D. (1983). Histological versus morphological assessment of graft rejection in invertebrates. *Transplantation* 36, 629–631.

Watson, F. L., Puttmann-Holgado, R., Thomas, F., Lamar, D. L., Hughes, M., Kondo, M., Rebel, V. I., and Schmucker, D. (2005). Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 309, 1874–1878.

Weigert, M., Gatmaitan, L., Loh, E., Schilling, J., and Hood, L. (1978). Rearrangement of genetic information may produce immunoglobulin diversity. *Nature* 276, 785–790.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583.

Whittaker, C. A., Bergeron, K. F., Whittle, J., Brandhorst, B. P., Burke, R. D., and Hynes, R. O. (2006). The echinoderm adhesome. *Dev. Biol.* 300, 252–266.

Wilson, M. J., Torkar, M., Haude, A., Milne, S., Jones, T., Sheer, D., Beck, S., and Trowsdale, J. (2000). Plasticity in the organization and sequences of human KIR/ILT gene families. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4778–4783.

Yui, M., and Bayne, C. (1983). Echinoderm immunity:bacterial clearance by the sea urchin *Strongylocentrotus purpuratus. Biol. Bull.* 165, 473–485.

Zhang, S. M., Adema, C. M., Kepler, T. B., and Loker, E. S. (2004). Diversification of Ig superfamily genes in an invertebrate. *Science* 305, 251–254.

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.

Received: 01 December 2011; accepted: 21 March 2012; published online: 12 April 2012.

Citation: Smith LC (2012) Innate immune complexity in the purple sea urchin: diversity of the Sp185/333 system. Front. Immun. ***3****:70. doi: 10.3389/fimmu.2012.00070*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

REVIEW ARTICLE
published: 13 July 2012
doi: 10.3389/fimmu.2012.00199

Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox

Gerardo R. Vasta[1]*, Hafiz Ahmed[2], Mihai Nita-Lazar[1], Aditi Banerjee[1], Marta Pasek[1], Surekha Shridhar[1,3], Prasun Guha[2] and José A. Fernández-Robledo[1]

[1] *Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, Baltimore, MD, USA*

[2] *Department of Biochemistry and Molecular Biology, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, Baltimore, MD, USA*

[3] *The Community College of Baltimore County, Catonsville, MD, USA*

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Carlo Pucillo, University of Udine, Italy
George Hajishengallis, University of Pennsylvania, USA

****Correspondence:***
Gerardo R. Vasta, Department of Microbiology and Immunology, Institute of Marine and Environmental Technology, University of Maryland School of Medicine, Columbus Center, 701 East Pratt Street, Baltimore, MD 21202-3101, USA.
e-mail: gvasta@som.umaryland.edu

Galectins are characterized by their binding affinity for β-galactosides, a unique binding site sequence motif, and wide taxonomic distribution and structural conservation in vertebrates, invertebrates, protista, and fungi. Since their initial description, galectins were considered to bind endogenous ("self") glycans and mediate developmental processes and cancer. In the past few years, however, numerous studies have described the diverse effects of galectins on cells involved in both innate and adaptive immune responses, and the mechanistic aspects of their regulatory roles in immune homeostasis. More recently, however, evidence has accumulated to suggest that galectins also bind exogenous ("non-self") glycans on the surface of potentially pathogenic microbes, parasites, and fungi, suggesting that galectins can function as pattern recognition receptors (PRRs) in innate immunity. Thus, a perplexing paradox arises by the fact that galectins also recognize lactosamine-containing glycans on the host cell surface during developmental processes and regulation of immune responses. According to the currently accepted model for non-self recognition, PRRs recognize pathogens via highly conserved microbial surface molecules of wide distribution such as LPS or peptidoglycan (pathogen-associated molecular patterns; PAMPs), which are absent in the host. Hence, this would not apply to galectins, which apparently bind similar self/non-self molecular patterns on host and microbial cells. This paradox underscores first, an oversimplification in the use of the PRR/PAMP terminology. Second, and most importantly, it reveals significant gaps in our knowledge about the diversity of the host galectin repertoire, and the subcellular targeting, localization, and secretion. Furthermore, our knowledge about the structural and biophysical aspects of their interactions with the host and microbial carbohydrate moieties is fragmentary, and warrants further investigation.

Keywords: galectin, C-type lectin, microbial recognition, glycan ligands

INTRODUCTION

The functional interplay between lectins and their "self" or "non-self" carbohydrate receptors implicated in various aspects of immune responses of both vertebrates and invertebrates have been characterized in considerable detail in recent years (Akira et al., 2001; Liu and Rabinovich, 2005; Ludwig et al., 2006). It is now firmly established C-type lectins, ficolins, siglecs, and galectins are not only key players in innate immune processes that lead to pathogen recognition, endocytosis, complement activation, and antigen processing, but are also involved in adaptive immune functions, including B and T cell clonal selection, maturation, activation, and apoptosis. In addition, lectins also participate in other intracellular and extracellular biological processes such as glycoprotein trafficking, protein folding, cell–cell or cell–ECM interactions, signal transduction, fertilization, and development (reviewed in Vasta and Ahmed, 2009). The identification of strikingly diverse lectin repertoires in virtually every animal species, including the presence of multiple lectin families with numerous members and lectin isoforms, has enabled rationalization of their roles in immunity. Unlike immunoglobulins, however, lectins do not generate diversity in recognition by genetic recombination and therefore, considerable interest has arisen on the germline-encoded diversity of the lectin repertoires, the somatic mechanisms leading to expansion of their ligand recognition spectrum, and the structural/functional aspects their carbohydrate recognition domains (CRDs; Vilches and Parham, 2002; Garred et al., 2006).

It is now firmly established that in vertebrates innate immunity carries a substantial burden of the defense functions against infections disease, and in the past few years the instructive roles of innate immunity on adaptive immunity have been widely recognized. Furthermore, invertebrates and protochordates rely solely on innate immunity for defense against microbial infection. Thus, great interest has been generated in the structural-functional

aspects of its various components, particularly on complement, lectins, and Toll-like receptors (Fujita et al., 2004; Khalturin et al., 2004; Iliev et al., 2005). Both the *Drosophila* Toll and the mammalian Toll-like receptors recognize pathogens via highly conserved and widely distributed microbial surface molecules such as lipopolysaccharide, flagellin, lipoteichoic acid, or peptidoglycan ("pathogen-associated molecular patterns"; PAMPs), which are essential for the microbe but absent in the host. By recognizing such non-self molecular patterns, these receptors were designated as "pattern recognition receptors" (PRRs; Medzhitov and Janeway, 2002). Given that non-pathogenic microbes also share these surface molecules it has been suggested that these may be more accurately described as "microbe-associated molecular patterns" (MAMPs; Bittel and Robatzek, 2007). More recently, the term "virulence-associated molecular pattern" (VAMP) has been introduced to describe those factors (e.g., microbial toxins, flagellin) that enable the host to discriminate pathogenic microbes from the non-pathogenic ones (Miao and Warren, 2010). Finally, endogenous factors such as nuclear or cytosolic components that are released during tissue stress or necrosis can trigger inflammatory responses have been designated as "danger-associated molecular patterns" (DAMPs; Seong and Matzinger, 2004).

THE MANNOSE-BINDING LECTIN AS A PROTOTYPICAL PATTERN RECOGNITION RECEPTOR

Since the PRR/PAMP paradigm was initially established for Toll and TLRs, it has been progressively extended to other innate immune recognition proteins. Among the best-characterized animal lectins, the mannose-binding lectin (MBL) a member of the C-type lectin family has been described as a prototypical PRR (Garred et al., 2006). C-type lectins are characterized by their Ca^{2+} requirement for ligand binding and their structural fold (C-type lectin domain fold, CTLD), and in most family members, the presence of multiple, unrelated structural domains in the polypeptide (Zelensky and Gready, 2005). They comprise the collectins (MBLs, ficolin, conglutinin, pulmonary surfactant), proteoglycan core proteins, selectins, endocytic receptors, the mannose-macrophage receptor, and DC-SIGN (Zelensky and Gready, 2005; Ip et al., 2009). Although some C-type lectins such as selectins and DC-SIGN bind "self" glycans, others such as collectins recognize exposed sugar ligands on the microbial surface. Collectins are lectins with a collagenous region linked to the CRD that recognizes sugars on microbial surfaces, and upon binding to a serine protease (MBL-associated serine proteases; MASPs) may activate the complement cascade (Weis et al., 1998; Wallis, 2002; Nonaka, 2011; Kingeter and Lin, 2012; **Figure 1**). Several lectins homologous of MBLs and ficolins, MASPs, and complement components have been identified in invertebrates and ectothermic vertebrates, suggesting that C-type lectins and the complement system played a pivotal role in innate immunity long before the emergence of adaptive immunity in vertebrates (Weis et al., 1998; Wallis, 2002; Nonaka, 2011). The CTLD fold has a double-loop structure with its N- and C-terminal β strands (β1, β5) coming close together to form an antiparallel β-sheet (**Figure 1A**). The second loop that lies within the domain is long and it enters and exits the core domain at the same location. Four cysteine residues (C1–C4), the most conserved residues in the CTLD, form

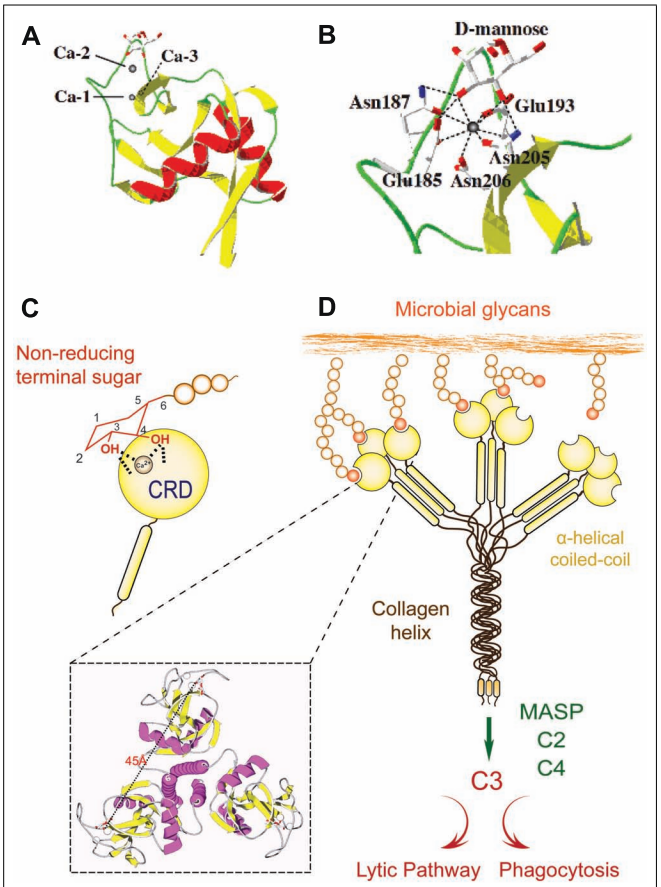


FIGURE 1 | Recognition and effector activities of the mannose-binding lectin (MBL). (A) A schematic representation of CTLD organization; **(B)** Crystallographic model of CTLD; **(C)** Ca^{2+}-dependent carbohydrate binding of CTLD. The CRD recognizes equatorial hydroxyls on C3 and C4 of non-reducing terminal mannose with participation of the Ca^{2+} atom. **(D)** The MBL trimer binds to ligands that are displayed about 45 Å apart on the microbial surface, and via association with the MASP) may activate the complement cascade, leading to opsonization or lysis of the microbe.

disulfide bridges at the bases of the loops. The residues C1 and C4 link β5 and α1 (the whole domain loop), while C2 and C3 residues link β3 and β5 (the long loop region). The rest of the chain contains two flanking α helices (α1 and α2) and the second β-sheet, formed by strands β2, β3, and β4 (Weis et al., 1998; Feinberg et al., 2000; Liu and Eisenberg, 2002). The long loop region is involved in Ca^{2+}-dependent carbohydrate binding, and in domain-swapping dimerization of some CTLDs. Four Ca^{2+}-binding sites are present in the CTLD structures, of which only one (site 2) is known to participate in binding to the carbohydrate ligand (Loeb and Drickamer, 1988; Weis et al., 1991; Feinberg et al., 2000). Resolution of the structure of the rat MBP-A/Man_6-$GalNAc_2$-Asn complex revealed that a ternary complex is formed between the protein, the Ca^{2+} ion bound in site 2, and the terminal mannose moiety of the oligosaccharide (Weis et al., 1992). The complex is stabilized by a network of coordination and hydrogen bonds: oxygen atoms from 4- and 3-hydroxyls of the mannose form two coordination bonds with the Ca^{2+} ion and four hydrogen bonds with the carbonyl side chains that form

the Ca^{2+}-binding site 2 (**Figure 1B**). The amino acid residues flanking the conserved *cis*-proline in the long loop region, which are involved in Ca^{2+}-binding site 2 formation, determine the specificity for either galactose or mannose (**Figure 1C**). In most mannose-binding CTLDs, the sequence of the motif is EPN (E^{185} and N^{185} in MBP-A), while in the galactose-specific CTLDs it is QPD. The oligomerization of the MBL subunits results in binding multivalency that enables the protein to recognize ligands that are displayed 45 Å apart on the microbial surface, thereby increasing the MBL's avidity (**Figure 1D**). The density of the surface ligands and their scaffolding (as glycoproteins or glycolipids) modulates affinity of the interaction via negative cooperativity (Dam and Brewer, 2008). Thus, the binding of MBL to multiple non-reducing terminal carbohydrate ligands on the microbial surface, which are not readily exposed in the mammalian host, leads to agglutination and immobilization of the potential pathogen. Further, the interactions of other MBL domains with additional factors such as the MASP trigger downstream effector functions including complement activation and opsonization or lysis of the agglutinated microbes (Weis et al., 1998; Nonaka, 2011). In addition to C-type lectins, other lectins families have been identified as PRRs, including ficolins, F-lectins, pentraxins, and more recently, the galectins (reviewed in Vasta and Ahmed, 2009).

GALECTINS: A STRUCTURALLY CONSERVED LECTIN FAMILY

Galectins constitute a family of animal lectins defined by their affinity for β-galactosides, and a characteristic CRD sequence motif (Cooper, 2002). The galectin family members are widely distributed in eukaryotic taxa from fungi and sponges to both protostome and deuterostome lineages of metazoans and from the structural standpoint are strikingly conserved (Vasta et al., 1999; Cooper, 2002). Based on their domain organization, galectins have been classified in three types: "proto," "chimera," and "tandem-repeat" (Hirabayashi and Kasai, 1993; **Figure 2A**). Proto-type galectins contain one CRD per subunit and are non-covalently linked homodimers. The chimera galectins have a C-terminal CRD and an N-terminal domain rich in proline and glycine. In tandem-repeat (TR) galectins, two CRDs are joined by a functional linker peptide. Recently, a novel TR-type galectin with four CRDs has been described (Tasumi and Vasta, 2007). Proto and TR types comprise several distinct galectin subtypes, which have been numbered following the order of their discovery. At present time, 15 galectin subtypes have been identified in mammals. Galectin-1, -2, -5, -7, -10, -11, -13, -14, and -15 are proto type. Galectin-3 is the only chimera type. Galectin-4, -6, -8, -9, and -12 are TR type. In solution, galectins can form multivalent species in a concentration-dependent equilibrium (Morris et al., 2004). The association of proto-type galectin monomers as non-covalently bound dimers via a hydrophobic interphase is critical for their function in mediating cell–cell or cell–ECM interactions, lattice formation at the cell surface, and downstream effector functions (Rabinovich et al., 2002b). For the chimera-type galectins, oligomerization takes place via the N-terminus domain to form trimers or pentamers that in the presence of multivalent oligosaccharides in solution or at the cell surface display binding cooperativity (Brewer et al., 2002; Rabinovich et al., 2002b; Dam and Brewer, 2008). Proto-type galectins associate as non-covalently bound dimers via a hydrophobic interphase, while the bivalent TR-type galectins can recognize different saccharide ligands with a single polypeptide, although they can also form higher order aggregates that enhances their avidity (Liu et al., 2012; Troncoso et al., 2012). Most galectins are non-glycosylated soluble proteins, although a few exceptions have transmembrane domains (Gorski et al., 2002; Lipkowitz et al., 2004). Although galectins lack a typical secretion signal peptide, they are present not only in the cytosol and the nucleus, but also in the extracellular space (Cho and Cummings, 1995). From the cytosol, galectins may be targeted for secretion by non-classical mechanisms, possibly by direct translocation across the plasma membrane (Sato and Hughes, 1994; Cleves et al., 1996; Vyakarnam et al., 1997). In the extracellular space, galectins can bind to glycans at the cell surface and/or the

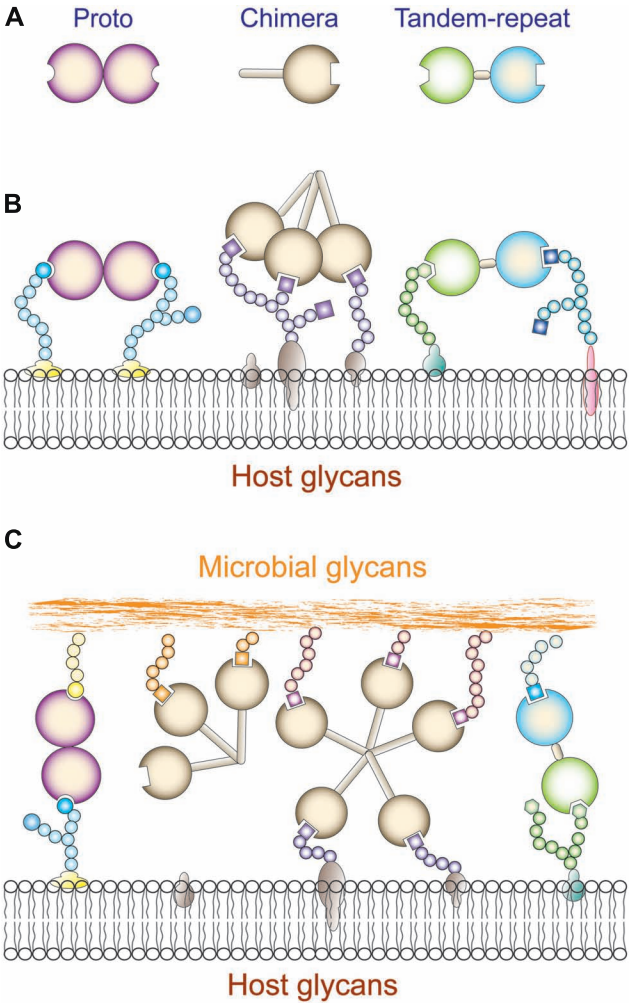


FIGURE 2 | Galectin recognition. (A) Schematic representation of galectin domain organization; **(B)** Schematic illustration of *cis*-interactions of proto, chimera, and tandem repeat galectins with host cell surface glycans. **(C)** Schematic illustration of *trans*-interactions of proto, chimera, and tandem repeat galectins with host cell surface and microbial glycans. Proto- and trandem repeat-type galectins can cross-link host and microbial glycans. Chimera galectins (galectin-3) can recognize microbial glycans but it is not clear that they can cross-link them to the host cell surface.

extracellular matrix (Elola et al., 2007; Rabinovich and Toscano, 2009) and to potential pathogens (Mercier et al., 2008; Vasta, 2009). Galectins preferentially bind to *N*-acetyllactosamine (LacNAc; Galβ1,4GlcNAc) and related disaccharides, including lactose (Lac), T-disaccharide (Galβ1,3GalNAc), and human ABH blood group oligosaccharides. Thus, glycans that contain *N*-acetyllactosamine and polylactosamine chains [(Galβ1,4GlcNAc)n], such as laminin, fibronectin, lysosome-associated membrane proteins, and mucins, are the preferred endogenous glycans recognized by galectins (Fang et al., 1993; Seetharaman et al., 1998). The biological function of a particular galectin, however, may vary from site to site, depending on the availability of suitable ligands. The binding properties and biological functions of galectins in the oxidative extracellular environment, however, may depend on their immediate binding to ligand, which prevents the oxidation of free cysteine residues, as well as galectin susceptibility to proteolysis (Lobsanov et al., 1993; Liao et al., 1994). The binding of galectins to cell surface β-galactoside-containing glycolipids and glycoproteins can lead to the formation of lattices that cluster these ligands into lipid raft microdomains required for optimal transmission of signals relevant to cell function (Brewer et al., 2002; Partridge et al., 2004; Rabinovich et al., 2007b; **Figures 2B and C**). Galectin-mediated lipid raft assembly may modulate turnover of endocytic receptors, signal transduction pathways leading to T cell activation and cytokine secretion, or apoptosis, B cell maturation, activation and tolerance, and neutrophil activation leading to phagocytosis, oxidative burst, and protease and cytokine release. Thus, galectin-glycoprotein lattices at the cell surface have been proposed to function as an "on-an-off switch" that regulates cell proliferation, differentiation and survival, including immune cell responsiveness and tolerance (Brewer et al., 2002; Dam and Brewer, 2008).

Among mammalian galectins, gene organization and primary structures of the encoded proteins are substantially conserved (**Figure 3A**). Prior to or during early in chordate evolution, duplication of a mono-CRD galectin gene would have led to a bi-CRD galectin gene, in which the N- and C-terminal CRDs subsequently diverged into two different subtypes, defined by exon–intron structure (F4-CRD and F3-CRD). All vertebrate single-CRD galectins belong to either the F3 (e.g., gal-1, -2, -3, -5) or F4 (e.g., gal-7, -10, -13, -14) subtype, whereas TR galectins such as gal-4, -6, -8, -9, and -12 contain both F4 and F3 subtypes (Houzelstein et al., 2004). However, phylogenetic analysis of the galectin (CvGal) from the eastern oyster *Crassostrea virginica*, that displays four tandemly arrayed CRDs, revealed that these are closely related to the single CRD galectins, suggesting that the CvGal gene is the product of two consecutive gene duplications of a single-CRD galectin gene (Tasumi and Vasta, 2007; **Figure 3B**). Like for C- and F-type lectins, most mammals are endowed with a complex galectin repertoire, including members that exhibit multiple isoforms and more or less subtle variations in carbohydrate specificity, which together with a certain degree of plasticity in sugar binding of each CRD, suggests a substantial diversity in recognition properties (Sparrow et al., 1987; Zhou and Cummings, 1990; Sato and Hughes, 1992; Fang et al., 1993; Ahmed et al., 2002; Shoji et al., 2003). In general, ectothermic vertebrates and invertebrates, and earlier taxa such as parazoa

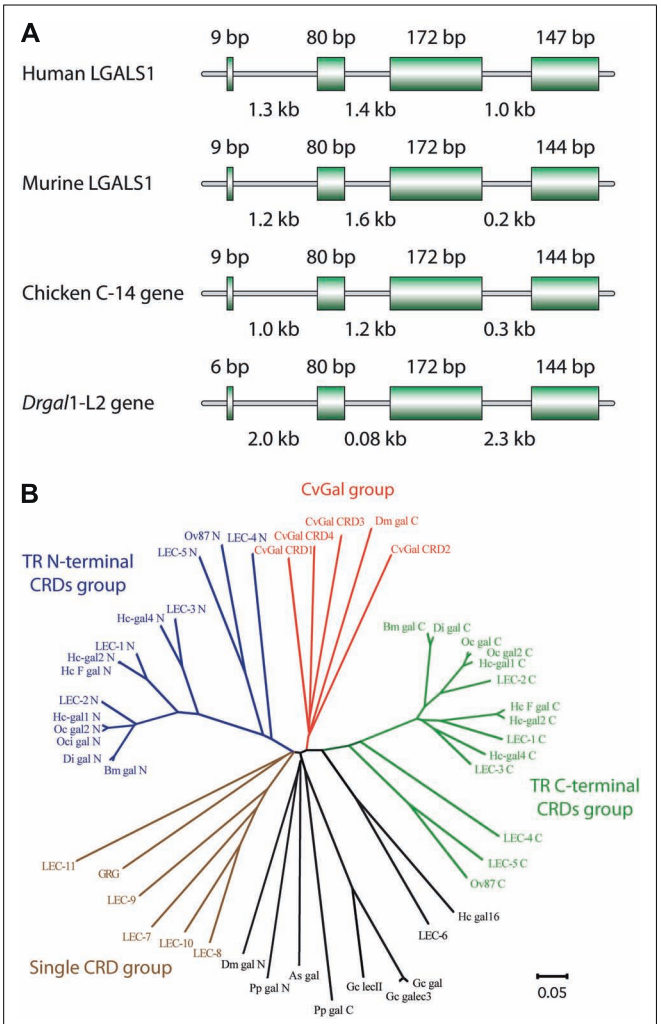


FIGURE 3 | Gene organization and phylogeny of galectins. (A) Genomic structures of galectin-1 or galectin-1 like protein from human, murine, chicken, and zebrafish (*Danio rerio*). Coding sequences (exons) are represented by boxes, with their sizes noted above each box. Intron sizes are shown below. **(B)** Phylogenetic analysis of invertebrate galectins modified based on Tasumi and Vasta (2007). The unrooted tree constructed by the N-J distance method is shown.

(sponges), fungi, and protista appear to have a smaller galectin repertoire. The presence of a galectin fold in the protistan parasite *Toxoplasma gondii*, and galectin-like proteins in the fungus *Coprinopsis cinerea* and in the sponge *Geodia cydonium* reveals the early emergence and structural conservation of galectins in eukaryotic evolution (Saouros et al., 2005; Walser et al., 2005; Stalz et al., 2006). Among the ectothermic vertebrates, proto-type galectins have been characterized in toads (*Bufo arenarum*), frogs (*Rana catesbeiana*, *Xenopus laevis*), salamanders (*Ambystoma mexicanum*), and zebrafish, among numerous other species (Ahmed et al., 1996, 2004; Vasta et al., 1997; Cooper, 2002; Shoji et al., 2003). The proto-type galectin from *B. arenarum* resembles the mammalian galectin-1 with respect to its carbohydrate-binding profiles and its carbohydrate-binding sequence motif (all nine residues responsible for carbohydrate binding are present: His45,

Asn47, Arg49, His53, Asp55, Asn62, Trp69, Glu72, and Arg74; Ahmed et al., 1996). The structures of the *B. arenarum* galectin complexed with LacNAc or TDG further support this close similarity (Bianchet et al., 2000). Multiple proto-type galectins from *X. laevis* and zebrafish contain all the above nine amino acid residues for carbohydrate binding, and are likely to exhibit a binding profile similar to the mammalian galectin-1. Some divergent family members exhibit unique features, ranging from sequence replacements in the CRD to major structural differences. In a proto galectin from *X. laevis*, His52 and Arg73 (numbered as the bovine galectin-1) are replaced by Ser and Lys, respectively (Marschal et al., 1992; Shoji et al., 2003), leading to a distinct sugar-binding profile (Marschal et al., 1992; Ahmed and Vasta, 1994). In congerin I, a galectin from the conger eel *Conger myriaster*, one of the β strands is exchanged between the two subunits (Shirai et al., 1999, 2002). This "strand-swap" contributes to stabilize the dimer by increasing inter-subunit interactions, and perhaps explains the high thermostability of the protein. CGL2, a galectin from the fungus *Coprinus cinereus* forms a tetramer characterized by two perpendicular twofold axes of rotation, with the C-terminal amino acids of the four monomers meeting at the center of the tetramer interface (Walser et al., 2004). Furthermore, some galectin-like proteins such as the mammalian lens crystalline protein GRIFIN (galectin-related inter-fiber protein) and the galectin-related protein GRP (previously HSPC159; hematopoietic stem cell precursor) lack carbohydrate-binding activity, and are considered products of evolutionary co-option (Ogden et al., 1998; Ahmed and Vasta, 2008).

Resolution of the structure of galectin-1–LacNAc complex revealed a jellyroll topology typical of legume lectins, and enabled the identification of the amino acid residues and the hydroxyl groups of the ligands that participate in protein–carbohydrate interactions (Lobsanov et al., 1993; Liao et al., 1994; Bianchet et al., 2000). The subunit of galectin-1 is composed of an 11-strand antiparallel β-sandwich and contains a single CRD (**Figure 4A**). The carbohydrate binding site is formed by three continuous concave strands (β4–β6) containing all important residues such as histidine 44, asparagine 46, arginine 48, histidine 52, asparagine 61, tryptophan 68, glutamic acid 71, and arginine 730 that are involved in direct interactions with LacNAc (Liao et al., 1994; **Figure 4B**). Additional interactions of water molecules are involved in bridging the nitrogen of the NAc group with His52, Asp54, and Arg73 resulting in the increased affinity of LacNAc over Lac. Unlike galectin-1, galectin-3 has an extended carbohydrate-binding site formed by a cleft open at both ends, in which the LacNAc is positioned in such a way that the reducing end of the LacNAc (GlcNAc) is open to solvent, but the non-reducing Gal moiety is in close proximity to residues in the β3 strand (Seetharaman et al., 1998). The extended binding site leads to increased affinity for glycans with multiple lactosamine units, and with their substitution of the non-reducing terminal galactose moiety such as ABH blood group oligosaccharides [Fucα1, 2; GalNAcα1,3(Fucα1,2); and Galα1,3(Fucα1,2)]. Thermodynamic approaches have been used not only to assess the galectins' carbohydrate-binding properties, but also the oligomeric organization of the protein. On microcalorimetric studies, the dissociation constants for the interactions of bovine galectin-1 with the preferred ligands (lactose,

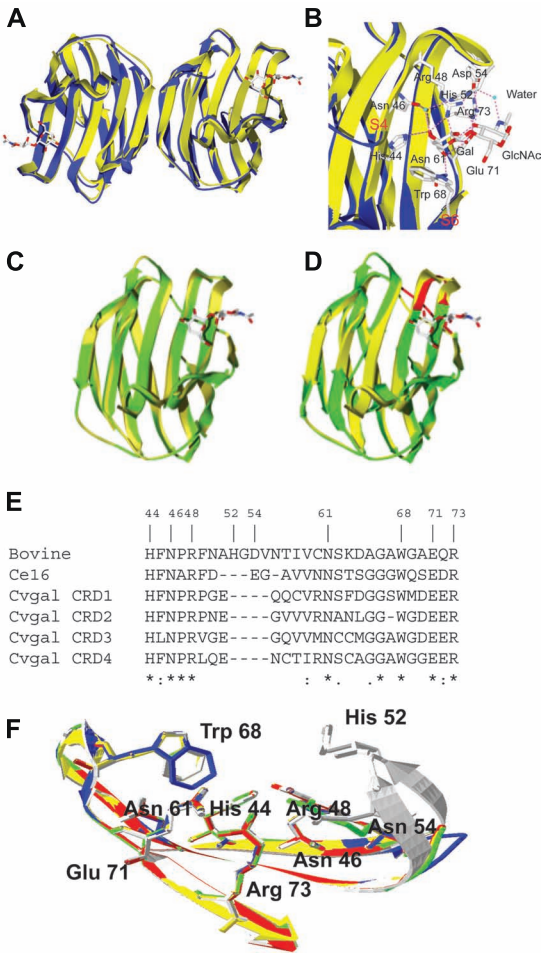


FIGURE 4 | Structures of bovine galectin-1 and *Bufo arenarum* galectin-1 like protein. (A) The ribbon diagram shows the overlap of the toad (*B. arenarum*) galectin-1 like protein (blue, PDB 1GAN) and bovine (*Bos taurus*) galectin-1 (yellow, PDB 1slt) in complex with LacNAc (stick representation). **(B)** Carbohydrate-binding sites of *B. arenarum* galectin-1 like protein (blue) and bovine galectin-1 (yellow). The interactions of amino acid residues with LacNAc are shown for the bovine galectin-1. The OH at C4′ of Gal (in LacNAc) makes hydrogen bond interactions with the highly conserved residues His44, Asn46, and Arg48. The OH at C6′ makes similar interactions with the Asn61 and Glu71. Trp68 participates in a stacking interaction with the Gal ring carbons and restricts orientation of the OH at C4′ to the axial form. In GlcNAc moiety of the LacNAc, the hydrogen bond interactions are involved with the protein through the C3-OH with Agr48, Glu71, and Arg73. Additional interactions are involved via a water molecule that bridges the nitrogen of the NAc group with His52, Asp54, and Arg73. **(C)** *Drgal*1-L2 (green**)** was modeled at the SWISS-MODEL Protein Modeling Server (http://swissmodel.expasy.org) based on the bovine galectin-1 structure (yellow, PDB 1slt). All nine residues that form the carbohydrate-binding cassette in mammalian galectin-1 are present in the putative binding site of *Drgal*1-L2. All side chains of these residues were within 0.5 Å of the equivalent side chains of the bovine galectin-1. **(D)** *C. elegans* 16-kDa galectin (Lec-6; shown in green) was modeled at the SWISS-MODEL Protein Modeling Server (http://swissmodel.expasy.org) based on the bovine galectin-1 structure (shown in yellow, PDB 1slt). The model reveals that a shorter loop (indicated by arrow) between strands 4 and 5 is responsible for its unique binding profile. **(E)** Alignment of bovine galectin-1, Ce16 (*C. elegans* 16 kDa galectin), and CRD1 to -4 of CvGal. **(F)** Homology modeling of CvGal CRDs. Bovine galectin-1, CRD-1, -2, -3, and -4 are shown in white, blue, yellow, red, and green, respectively. Numbering of amino acid residues is based on bovine galectin-1.

N-acetyllactosamine, thiodigalactoside) were in the range of 10^{-5} M, with two binding sites per molecule (Schwarz et al., 1998).

The galectin-1 like protein (Drgal-L2) from zebrafish (*Danio rerio*) shows extensive sequence homology and structural similarity to vertebrate galectin-1 (**Figure 4C**). All nine residues that form the carbohydrate-binding cassette in the mammalian galectin-1 are present in the putative binding site of Drgal1-L2 (Ahmed et al., 2002). All side chains of these residues were within 0.5 Å of the equivalent side chains of the bovine galectin-1. Although, the model of the *Caenorhabditis elegans* 16-kDa galectin (Ce16) is similar to the bovine galectin-1 structure (**Figure 4D**), the carbohydrate specificity of the Ce16 is interesting compared to the bovine galectin-1 (Ahmed et al., 2002). In Ce16, amino acid substitutions at positions 46 and 48 (29 and 31 in bovine galectin) suggest that unlike in bovine galectin-1, these positions (strand 3) are involved in sugar binding. R^{46} (S^{29} in the bovine galectin sequence) interacts with 3′- and 4′-OH of the Gal residue. Interestingly, a shorter loop 2 (containing residues 66–69 between strands 4 and 5; **Figure 4E**) allows E^{67} to interact with equatorial −OH at C-3 of GlcNAc (in Galβ1,4GlcNAc) as well as axial −OH at C-4 of GalNAc (in Galβ1,3GalNAc). The model of the binding site of the *C. elegans* 16-kDa galectin also provides a rationale for the binding to α and β derivatives of Galβ1,3GalNAc (Ahmed et al., 2002). The sequence alignment of the oyster *Crassostrea virginica* galectin CvGal with the bovine galectin-1 revealed that among nine aa residues responsible for ligand binding, seven are conserved in all four CvGal CRDs (His^{44}, Asn^{46}, Arg^{48}, Asn^{61}, Trp^{68}, Glu^{71}, and Arg^{73}; numbers corresponding to the bovine galectin-1; Tasumi and Vasta, 2007; **Figure 4E**). Homology modeling of all four CvGal CRDs revealed that the seven conserved residues from all four CvGal CRDs maintain their positions and orientations in the binding cleft, relative to the bovine galectin-1 (**Figure 4F**).

Although relatively conserved from a structural standpoint, galectins display a surprising functional diversification (**Figure 5**). Proto-type galectins such as galectin-1 have been reported as mediators of cell adhesion (Weis et al., 1998; Wallis, 2002), B cell differentiation (He and Baum, 2004; Martinez et al., 2004), development (Poirier, 2002), inflammation (Rubinstein et al., 2004), mRNA splicing (Patterson et al., 2004), leukocyte apoptosis (Perillo et al., 1998; Rabinovich et al., 2002a), neutrophil turnover (Dias-Baruffi et al., 2003), and cancer metastasis (Takenaka et al., 2004; Liu and Rabinovich, 2005). Furthermore, given the complexity of the mammalian galectin repertoire and the initial difficulties in identifying clear phenotypes in null mice for selected galectin family members, it was believed that a certain degree of functional redundancy exists among the members of the galectin repertoire. More recently, however, as the subtle aspects of their binding properties and natural ligands are identified and characterized, and their biological roles are elucidated in increasing detail, it has become clear that this is not the case (Rabinovich and Toscano, 2009). Soon after their discovery, galectins were characterized as developmentally regulated proteins, and proposed to participate in embryogenesis and early development. This was based on their binding to "self" carbohydrate moieties, such as polylactosamine-containing glycans, abundant at the cell surface and the ECM (**Figure 5**). Chicken galectins have been proposed to participate in myoblast fusion, whereas murine galectin-1 and galectin-3 have roles in notochord development, somitogenesis, and development of muscle tissue and central nervous system (Poirier et al., 1992; Colnot et al., 1996; Puche et al., 1996; Shoji et al., 2009).

Recent studies show that the gene encoding human galectin-2, a close relative of galectin-1, is mutated in patients with increased susceptibility to myocardial infarction, implicating galectin functions in vascular inflammation and atherosclerosis (Ozaki et al., 2004). Further, expression of galectin-3 at the onset of chondrification suggests its role in bone development, which was supported by finding of abnormalities of the hypertrophic zones of cells and a reduction of the total hypertrophic chondrocytes in galectin-3 null mice (Colnot et al., 2001).

GALECTINS AS REGULATORS OF IMMUNE HOMEOSTASIS

It is now firmly established that galectins participate in both innate and adaptive immune responses (Lund and Olafsen, 1999; Sato et al., 2003; Cartwright et al., 2004; Ohtsubo and Marth, 2006; Marth and Grewal, 2008; Mascanfroni et al., 2011). Galectins are ubiquitously expressed and distributed in mammalian tissues, including most cells of the innate (dendritic cells (DCs), macrophages, mast cells, natural killer cells, gamma/delta T cells, and B-1 cells) and adaptive (activated B and T cells) immune system, and as in other cell types (Li et al., 2011; Cedeno-Laurent and Dimitroff, 2012).

Since the early 1990s a growing body of experimental (*in vivo* and *in vitro*) evidence has accumulated to support the roles of galectins expressed by these cells and neighboring stromal cells in the development and regulation of innate and adaptive immune homeostasis as well as responses to infectious and allergic challenge, and cancer. Galectins released by stromal cells in central compartments contribute to the differentiation of immune cell precursors. Immune challenge and several pathological conditions may lead to further activation and differentiation of immune cells, and modulate the expression and release of galectins to the extracellular space where they may have autocrine or paracrine effects on immune regulation. Galectins released by immune cells can oligomerize and form lattices at the cell surface leading to activation of transmembrane signaling pathways that modulate immune cell functions, including for example, cell adhesion and migration, T cell apoptosis, and the Th1/Th2 cytokine balance (Rabinovich and Toscano, 2009). Further, galectins released into the extracellular environment under abnormal situations may themselves constitute "danger signals," or by exerting their activities on other cells, such as mast cells, induce degranulation and release of factors (e.g., histamine) that represent the "danger signals" leading to activation of immune mechanisms in the absence of antigenic challenge (Sato and Nieminen, 2004).

Galectins have diverse effects on cells involved in innate immune responses (Rabinovich et al., 2007a; Di Lella et al., 2011) including macrophages and DCs, neutrophils, eosinophils, and mast cells. Galectin-1 participates in acute and allergic inflammation and displays anti-inflammatory activities by blocking or attenuating signaling events that lead to leukocyte infiltration, migration, and recruitment (Guevremont et al., 2004). It also displays various other effects on innate immunity, including

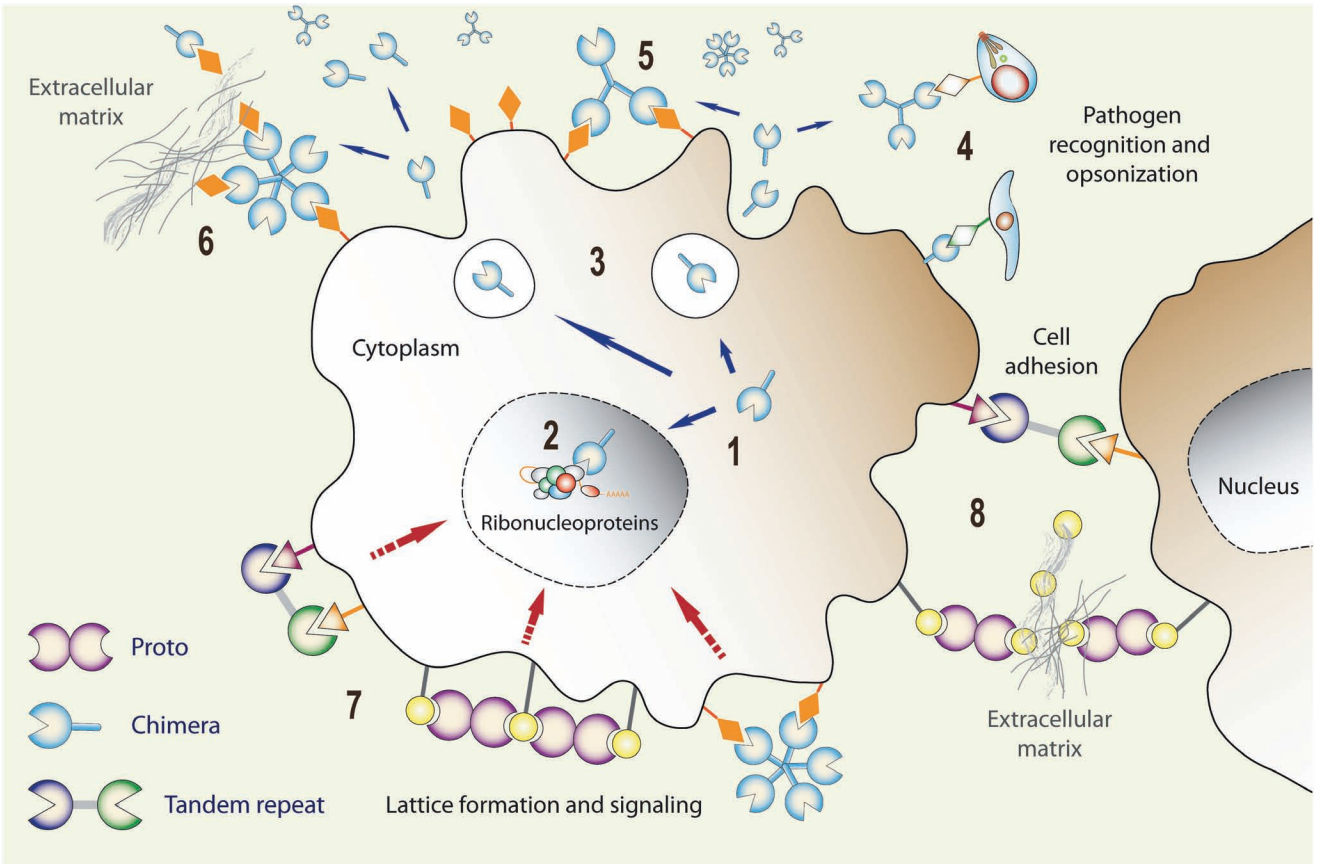


FIGURE 5 | Expression, secretion, and functional diversification of galectins: **(1)** Galectin transcripts are translated in the cytoplasm, and the proteins can be translocated into the nucleus **(2)** where they can associate with ribonucleoproteins. *Via* unconventional mechanism(s), galectins can be secreted to the extracellular space **(3)** where they can function as pattern recognition receptors for microbial glycans **(4)**, bind to the host cell surface glycans **(5)**, and cross-link them with ECM glycans **(6)** thereby, for example, promoting cell migration. Galectins can also cross-link cell surface glycans and induce clustering of microdomains and lattice formation at the cell surface **(7)** that can trigger signaling cascades, or cross-link neighboring cells **(8)** and promote cell–cell interactions/adhesion.

cell surface exposure of phosphatidylserine in activated neutrophils, a process that leads to neutrophil removal by phagocytic cells without causing apoptosis, and activation/deactivation of macrophages on a concentration-dependent manner. In contrast to the anti-inflammatory effects of galectin-1, galectin-3 shows pro-inflammatory activity. Galectin-3 is normally expressed in various epithelia and inflammatory cells, such as activated macrophages, DCs, and Kupffer cells, and is upregulated during inflammation, cell proliferation, and cell differentiation. Galectin-3 also exhibits anti-apoptotic activity for macrophages and enhances their interactions with basal lamina glycans, such as laminin and fibronectin. Taken together, these observations strongly suggest that galectin-3 enhances macrophage survival, and positively modulates their recruitment and anti-microbial activity. Galectin-9 is a selective chemoattractant for eosinophils, highly expressed in various tissues of the immune system, such as bone marrow, spleen, thymus, and lymph nodes. Gal-9 released from activated T cells induces chemotaxis, activation, oxidative activity, and degranulation of eosinophils, and monocyte-derived DC maturation (Partridge et al., 2004; Ohtsubo et al., 2005; Rabinovich et al., 2007b; Stowell et al., 2008). Galectin-9 also promotes tissue inflammation through interaction with TIM-3 on macrophages (Anderson et al., 2007; Di Lella et al., 2011).

Concerning adaptive immune responses galectins function as regulators of immune cell homeostasis (Cartwright et al., 2004; Rabinovich et al., 2007a; Ilarregui et al., 2009; Di Lella et al., 2011). Interactions between stromal cells from the bone marrow and thymic compartments and lymphocyte precursors are critical to their development, selection, and further progression to the periphery. In this regard, interactions mediated by galectins can modulate B cell maturation and differentiation both at the central and peripheral immune compartments (Rabinovich et al., 2007a). Similarly, from their early developmental stages in the thymic compartment to the removal of the mature activated T cells in the periphery, the regulation of T cell survival is critical to a controlled immune response. Galectin-1 can regulate T cell proliferation and apoptosis through binding and clustering of lactosamine-rich cell surface glycoconjugates into segregated membrane microdomains (Rabinovich and Toscano, 2009). Galectin-1 may have pro- or anti-apoptotic effects on T cells depending on the developmental

stage and activation status of the cell, and the microenvironment in which the exposure takes place. An immunoregulatory circuit involving galectin-1, DCs, and T cells has recently been described by Rabinovich and his associates (Ilarregui et al., 2009). Upon exposure to galectin-1, DCs acquired an IL-27-dependent regulatory function and promoted IL-10-mediated T cell tolerance, Th1, and Th17 responses, and suppressed autoimmune neuroinflammation (Ilarregui et al., 2009). Effects of galectin-3 in T cell survival are dependent on whether protein is produced endogenously (anti-apoptotic) or by exogenous exposure (pro-apoptotic; Shoji et al., 2003). Several studies suggest that intracellular galectin-3 confers resistance to apoptosis in response to chemotherapeutic drugs (Fukumori et al., 2007). In contrast, extracellular galectin-3 has been shown to induce T cell apoptosis through caspase-3 activation (Fukumori et al., 2003). Galectin-3 can also modulate T cell activation. Recent studies demonstrated that endogenous galectin-3 can directly control T cell activation at sites of immunological synapse (Chen et al., 2009).

Galectins also exert regulatory functions in T cell homeostasis, and signaling cascades triggered by their binding and lattice formation at the T cell surface has implications in a variety of downstream events that modulate their differentiation, functional activation, and production of pro- and anti-inflammatory cytokines. The effects of galectins on T cell cytokine synthesis and secretion ultimately determine the Th1/Th2 polarization of the immune response. By reducing IFN-γ and IL-2 and enhancing IL-5, IL-10, and TGF-β production, galectin-1 skews the balance from a Th1- toward a Th2-polarized response, whereas by reducing IL-5 levels, galectin-3 drives have the opposite effect (Liu and Hsu, 2007). Finally, given the regulatory roles of galectins on cells that mediate both innate and adaptive immune responses, their effects can be beneficial or detrimental to pathological conditions that have a basis on exacerbated or depressed immune function, such as inflammatory, allergic, and autoimmune disorders, and cancer (Liu and Hsu, 2007).

GALECTINS AS PATTERN RECOGNITION RECEPTORS

Insight into the multiple roles of galectins in both innate and adaptive immune functions has further expanded recently by discovery of their ability to directly recognize microbial pathogens (Lund and Olafsen, 1999), a property shared with other lectin types, such as C- and F-lectins, ficolins, and pentraxins. The roles of lectins in recognition of microbial glycans are particularly critical in invertebrates, since these organisms lack immunoglobulins and rely solely in innate immune mechanisms for recognition of potential microbial pathogens (Gerwick et al., 2007). This seems to be the case even in vertebrate taxa, since it has been reported that susceptibility/resistance to several infectious diseases in humans are determined by the presence of certain lectin alleles (Fuller et al., 2004). As discussed above, recognition of "non-self" carbohydrate moieties on the surface of microbial pathogens and parasites by lectins such as the MBL can be rationalized under the PRR/PAMP concept. In contrast, for those lectins that recognize endogenous glycans such as galectins, understanding recognition of microbial glycans requires additional considerations. In some cases, the ligands on foreign cells can be similar to those displayed on host cells such as ABH or Le blood group oligosaccharides (Cartwright et al., 2004; Stowell et al., 2010) and LacNAc present in viral and bacterial glycans, or structurally different and absent from the host glycome, such as α1-2-mannans in *Candida* (Cook et al., 2005) and LacdiNAc in *Schistosoma* (Cummings and Nyame, 1999). While the first scenario can be conceptualized as molecular mimicry by the microbial pathogens (Vasta, 2009), understanding the molecular basis of galectin binding to distinct self and non-self glycans via the same CRD requires further discussion. In this regard, galectins with tandemly arrayed CRDs such as the TR galectins from vertebrates, and the 4-CRD galectins from invertebrates are intriguing both in their binding properties and functional aspects. Vertebrate TR galectins such as galectin-4, -8, and -9 differ from the proto and chimera types in that they display two tandemly arrayed CRDs, and like the binary CRD F-type lectins, the N- and C-CRDs of TR galectins are similar but not identical, suggesting that they have distinct recognition properties (Carlsson et al., 2007). The structures of the TR galectin-4, -8, and -9 have been partially resolved either by crystallization of NMR analysis of their isolated N- or C-CRDs and revealed differences in their binding specificity and/or affinity for oligosaccharides or their scaffolding as glycolipids or glycoproteins (Tomizawa et al., 2005; Nagae et al., 2009; Krejcirikova et al., 2011). The structure of the N-CRD of the mouse galectin-4 revealed binding sites for lactose with different affinities, while the galectin-8 binds preferentially to larger glycans such as glycosphingolipids (Tomizawa et al., 2005; Nagae et al., 2009; Krejcirikova et al., 2011). The capacity of TR galectins to cross-link cells with different synthetic glycoconjugates (Tomizawa et al., 2005; Ideo et al., 2011) strongly suggests significant differences in the binding properties of their N- and C-CRDs. From the functional standpoint, the most striking example is the recognition and killing of *Escherichia coli* O86 that display B-blood group oligosaccharides (BGB+ *E. coli*) by the TR galectin-4 and -8. Mutation of key residues in either CRD revealed that the C-CRD mediates recognition the BGB+ *E. coli* but does not affect its viability, while the N-CRD was not affected, suggesting that N-CRD might be endowed with killing activity. This was confirmed by examining the binding and killing activity of the separate CRDs. This observation suggests that in galectin-4 and -8 the N- and C-CRDs not only have different recognition properties, but also they are functionally different in their interactions with the BGB+ *E. coli* (Stowell et al., 2009).

A different scenario is presented by the recognition of glycans on the phagocyte (self) and parasite (non-self) surface by the oyster galectin CvGal (Tasumi and Vasta, 2007). Although the four CRDs of CvGal are not identical, homology modeling suggests that the minor differences in sequence among the CRDs do not result in significant differences in binding properties. Therefore, the cross-linking of similar glycans on the host phagocytic cells and on the surface of the parasite trophozoite by CRDs of similar binding properties can only be rationalized by the cross-linking of similar self and non-self glycans by the folding of the 4-CRD polypeptide in a particular geometry, such as a tetrahedral architecture, where the CRDs are oriented in opposite directions. A similar model could be proposed by single-CRD proto type and chimera galectins for which the requisite multivalency and the spatial arrangement and orientation of the CRDs is achieved by oligomerization of the single-CRD subunits. However, questions

about the molecular and structural basis for recognition and binding to either similar or distinct self or non-self glycans remain open, and several aspects concerning the protein, the carbohydrate ligands, the thermodynamics of lectin–ligand interaction, and the particular microenvironment in which this interaction takes place merit further discussion. With regards to the protein, recognition specificity and affinity are key aspects of the lectin–ligand interaction(s) that modulate downstream effector functions.

In addition to the structural aspects of the primary and extended carbohydrate binding sites by any CRD in monovalent recognition, multiple factors contribute to the energetics of the interaction, which are further complicated when addressing multivalent recognition (Ideo et al., 2011). Although as discussed above for C-, and F-lectins and TR galectins multivalency is a property of tandemly arrayed CRDs in the lectin polypeptide subunit, oligomerization of the monomers results in displays of clustered binding sites that either in their geometry or dimensions can be unique to a particular lectin. For example, the "bouquet"-like display of binding sites in the MBL trimer results in multivalency for sugars spaced about 45 Å apart, and increases avidity of the binding (Rapoport et al., 2008). Although a similar CRD display is presented in F-type lectins, the geometry of the ligand display requires sugars spaced 26 Å apart (Bianchet et al., 2002, 2010). The tissue localization and subcellular compartmentalization of lectins can influence their oligomerization and in turn, influence the selective binding of lectins to particular ligands. For galectins, oligomerization and clustering of CRDs enables binding to and cross-linking of multivalent glycoproteins and glycolipids on the cell surface leading to formation of microdomains and lattices that signal vial specific pathways depending on the glycans and galectins involved (Greenspan, 2001; Rabinovich et al., 2007b; Vokhmyanina et al., 2011). In addition, the redox properties of the intracellular and extracellular environments can modulate the activity of lectin–ligand interactions, and the biological outcome. Most galectins are active in the reducing environment, but are susceptible to inactivation by oxidation of their free Cys residues present in the CRD (Stowell et al., 2009). In the reducing intracellular environment, galectins remain stable, but upon secretion to the mostly oxidative extracellular space, the oxidation of the free cysteines compromises not only the binding activity but also the oligomerization of the galectin subunits (Stowell et al., 2009). It is noteworthy that the binding of galectin-9 to the protein disulfide isomerase at the cell surface, increases retention of the enzyme which in turn modulates the redox status at the plasma membrane (Rabinovich et al., 2007b). In addition to the lectin quaternary structure as established via structural or hydrodynamic approaches, higher orders of aggregation may occur in environments where the lectin concentration reaches above a certain threshold. The galactosyl-binding lectins DCL-I and DCL-II from the protochordate *Didemnum candidum* behave as homotetramers of 14.5 and 15.5 kDa subunits, respectively, as revealed by sedimentation velocity and sedimentation equilibrium studies. However, even at concentrations below 1 mg/ml, a positive dependence of the sedimentation coefficient with protein concentration was observed, suggesting that at increasing concentrations the protein associates at orders above the tetramer (Vasta et al., 1986). For the dimeric galectin-1, further association to form tetramers is temperature driven (Dam et al., 2009). Furthermore, the ligand density at the cell surface can drive oligomerization of the lectins subunits. In solution, galectin-3 monomers are in equilibrium with higher order oligomers, and when binding to multivalent glycoproteins or cell surface glycans they may recruit additional monomers to form a complex of multivalent interactions via galectin-3 trimers and pentamers (Ahmad et al., 2004). The density of the carbohydrate ligands and their protein or lipid scaffolding context on the cell surface are key factors that determine lectin recognition and affinity. As proposed by Dam et al. (2009), in contrast to the well-defined dissociation constants that describe the binding of lectin monomers to monovalent glycans, a wide range of relative dissociation constants is needed to describe lectin binding to multivalent surface glycans, that depends on the density and number of glycans on a surface and increased negative cooperativity. The authors propose that this relative affinity should replace the term avidity (Brewer et al., 2002). Galectin binding to ligand is enthalpically driven, exhibits enthalpy-entropy compensation, and follows a van't Hoff dependence of the binding constant on temperature, properties that are shared by other lectins. Further, the solvation properties of the CRD, the reorganization or displacement of the water shell, and the establishment of defined water-mediated interactions between the protein and the carbohydrate are factors that affect the thermodynamic aspects of the recognition process and stability of the lectin–ligand interaction (Di Lella et al., 2007; Dam and Brewer, 2010; Echeverria and Amzel, 2011). As discussed above, many functional properties of lectins such as the triggering of signaling cascades result from clustered CRDs recognizing multivalent carbohydrate moieties in different scaffolds exposed at the cell surface, followed by the formation of microdomains and lattices (Rabinovich et al., 2007b). Therefore, a detailed knowledge of the thermodynamic aspects of lectin-mediated cross-linking of complex glycans is critical for the understanding of lectin-mediated selective recognition of self and non-self ligands. In this regard, recent studies have shown that high-affinity lectin–mucin interactions are driven by favorable binding entropy of binding associated with a "bind and jump" mechanism. This consists of a dynamic binding process in which the bound lectins jump from carbohydrate to carbohydrate moiety in the multivalent glycans, which by enhancing entropic effects facilitates binding and subsequent complex formation (Dam et al., 2009). The factors discussed above suggest that in the natural context, multiple factors that concern the protein binding site and lectin oligomerization, the display geometry and density of the recognized carbohydrate moieties in a particular protein or lipid scaffold, and the particular properties of the environment where the lectin–sugar interaction takes place, may determine not only the binding dynamics and affinity, but also the lectin's preference for any one of the (self or non-self) ligands available.

GALECTINS AS RECEPTORS FOR MICROBIAL ADHESION AND INFECTION

Whether galectin-mediated recognition is an effective defence mechanism with a clear benefit for the host is not entirely clear,

except for a few examples discussed above (Cartwright et al., 2004; Cook et al., 2005; Kohatsu et al., 2006; Stowell et al., 2010). In some cases, however, the microbe's recognition by the vector or host galectins promotes its adhesion, host cell entry, or infection persistence, in addition to modulating the host's immune responses. Thus, these pathogens and parasites would "subvert" the roles of host or vector galectins as PRRs, to attach to or gain entry into their cells. This is clearly illustrated by the participation of galectin interactions in the infection mechanisms of HIV. In contrast to the inhibitory role of galectin-1 in paramixovirus-mediated cell fusion, galectin-1, which is abundant in organs that represent major reservoirs for HIV-1, such as the thymus and lymph nodes, promotes infection by HIV-1 by facilitating viral attachment to CD4 receptor, and increasing infection efficiency (Ouellet et al., 2005; Mercier et al., 2008). Recent studies showed that galectin-1 enhances HIV adsorption kinetics on monocyte-derived host macrophages, which facilitates HIV-1 infectivity by shortening the time required to establish an infection. Further, galectin-1 would also function as a soluble scavenger receptor and enhance the uptake of the virus by macrophages, which together with evidence that galectin-1 is present in the ejaculate and the heads and tails of late spermatids, led to extend the proposal that galectin-1 may also facilitate sexual transmission of HIV-1 (Mercier et al., 2008). This would take place through enhancement of viral adsorption kinetics on the target cells' surface by the galectin-1 released by sheared fibroblasts and epithelial cells following sex-related microabrasions. Gal-3 has no effect on HIV-1 adsorption, entry, or infection, although its expression is upregulated by the HIV Tat protein in several human cell lines, and in cells infected with other retroviruses, suggesting that it may participate in regulation of antiviral immunity (Schroder et al., 1995; Hsu et al., 1996; Fogel et al., 1999). This underscores the relevance of the subtle differences in galectin specificity and affinity that may determine very different recognition and effector outcomes. It is noteworthy that HIV also uses recognition by DC-SIGN, a C-type lectin, to enter DCs, thereby underscoring the multiple adaptations of the viral glycome for host infection (Hijazi et al., 2011; van der Vlist et al., 2011).

Several *Leishmania* species, which spend part of their life cycle in phlebotomine sandflies that constitute vectors for transmission to the vertebrate hosts, are also illustrative examples. Upon the sandfly feeding on blood from an infected host, the ingested amastigotes mature into promastigotes, which attach to the insect midgut epithelium to prevent their excretion along with the digested bloodmeal, and undergo numerous divisions before differentiating into free-swimming infective metacyclics (Volf et al., 2007). Although the involvement of the parasite LPG in this interaction had been suspected from prior studies, the specific *Phlebotomus papatasi* sandfly midgut receptor for the procyclic *L. major* LPG was identified as a 35.4-kDa TR galectin (PpGalec) only expressed by epithelial midgut cells, and upregulated in the blood-feeding females (Myskova et al., 2007). Because the binding specificity of PpGalec is restricted to *Leishmania* promastigotes bearing poly-Gal (β1–3) side chains on their LPG, it was proposed that it is the carbohydrate moiety responsible for specific binding of *L. major* to *P. papatasi* midgut linings. The assembly of polygalactose epitopes is downregulated during *L. major* metacyclogenesis, and thus, unable to bind to rPpGalec the free-swimming infective metacyclic promastigotes are released from the midgut for transmission from the sandfly to the mammalian host (Myskova et al., 2007).

The protozoan parasite *Perkinsus marinus* is a facultative intracellular parasite that causes "Dermo" disease in the eastern oyster *Crassostrea virginica*, and is responsible for catastrophic damage to shellfisheries and the estuarine environment in North America (Perkins, 1996). The infection mechanism remains unclear, but it is likely that while filter feeding, the healthy oysters ingest *P. marinus* trophozoites released to the water column by the infected neighboring individuals. Inside oyster phagocytic cells (hemocytes), trophozoites resist oxidative killing, proliferate, and spread throughout the host. It was recently discovered that oyster hemocytes recognize *P. marinus* via a novel galectin (CvGal) that displays four canonical galectin CRDs, a domain organization unlike any of the known galectin types (Tasumi and Vasta, 2007). Two amino acid residues (His^{53} and Asp^{55}) that interact with the NAc group via a water molecule are missing in all four CvGal CRDs resulting in broader carbohydrate specificity. CvGal is present in the cytoplasm of circulating granulocytes, and upon their attachment and spreading it is translocated to the periphery, secreted, and binds to the cell surface. The remaining galectin is released to the extracellular environment, where it may bind to all other circulating (non-activated) granulocytes and hyalinocytes. The most surprising observation, however, was that the soluble CvGal also binds in a carbohydrate-specific manner to a wide variety of microorganisms, phytoplankton components, and preferentially, to *Perkinsus* spp. trophozoites, suggesting a direct role in recognition and opsonization of potential microbial pathogens, as well as algal food. The partial inhibition of phagocytosis of *P. marinus* trophozoites by pre-treatment of hemocytes with anti-CvGal revealed that the hemocyte surface-associated CvGal is a phagocytosis receptor for *P. marinus*. Thus, *P. marinus* may have evolved to adapt the trophozoite's glycocalyx to be selectively recognized by the oyster hemocyte CvGal, thereby subverting the oyster's innate immune/feeding recognition mechanism to gain entry into the host cells (Tasumi and Vasta, 2007). A galectin of similar domain organization was recently identified in the snail *Biomphalaria glabrata*, the intermediate host of the human blood fluke *Schistosoma mansoni*, and was proposed to mediate interactions with the parasite (Yoshino et al., 2008).

A recent study identified galectin-1 as the receptor for the protozoan parasite *Trichomonas vaginalis* (Okumura et al., 2008) the causative agent of the most prevalent non-viral sexually transmitted human infection in both women and men. As an obligate extracellular parasite, establishment and persistence of *T. vaginalis* infection requires adherence to the host epithelial cell surface. Like *Leishmania* spp., *T. vaginalis* displays a surface LPG rich in galactose and *N*-acetyl glucosamine, which is recognized in a carbohydrate-dependent manner by galectin-1 expressed by the epithelial cells in the cervical linings, as well as placenta, prostate, endometrial, and decidual tissue, also colonized by the parasite (Okumura et al., 2008).

CONCLUSION

By recognizing highly conserved and widely distributed microbial surface glycans (PAMPs or MAMPs) which are essential for the microbe but absent in the host, lectins like the MBL, a member of the C-type lectin family, fit accurately the definition of PRR. Recent studies clearly indicate that also galectins can function as PRRs that target oligosaccharides on the surface of virus, bacteria, protista and helminth pathogens, and parasites. A perplexing paradox arises, however, by the fact that galectins also recognize endogenous lactosamine-containing glycans on the host cell surface during development and regulation of immune homeostasis. For the TR-type galectins that display two CRDs of similar but distinct specificity in a single polypeptide monomer, the binding and cross-linking of endogenous and exogenous glycans can be rationalized by the distinct properties of their N- and C-terminus binding sites. For other lectins such as the proto- and chimera-type galectins that display a single CRD per monomer, their capacity to recognize both endogenous and exogenous glycans through the same binding site, can be explained by taking into consideration the multiple factors pertaining to the local lectin concentrations and oligomerization, the geometry of the presentation of the multivalent carbohydrate ligands on the host or microbial cell surface, and the properties of the microenvironment in which interactions take place. For example, in a local microenvironment where an infection is initiated, the particular concentration of galectins released to the extracellular space may lead to a distinctive oligomerization, CRD display geometry, and avidity that may shift galectin binding from the endogenous ligands to the glycans displayed on the microbial surface in a particular architectural presentation. It has been proposed that both the microbial and host glycomes and their receptors continuously evolve to escape mutual recognition, a process known as the "Red Queen effect" (reviewed in Vasta, 2009), by which the microbe avoids recognition by the host innate immune receptors (PRRs) and, the host by the microbial colonization factors (agglutinins, adhesins, and lectins). Given the key roles played by galectins in early development and regulation of immune homeostasis by the recognition of "self" glycans, strong functional constraints would prevent galectins from significant evolutionary changes in carbohydrate specificity. This is supported by the considerable structural conservation within this lectin family. Further, with the current evidence about how pathogens and parasites, which display a remarkable evolutionary plasticity, efficiently subvert the roles of galectins to attach or gain entrance into the host cells, it seems more plausible that instead of avoiding recognition by the host, they would have evolved their glycomes to mimic their hosts' in a "Trojan horse" model (reviewed in Vasta, 2009), and rely on the host's self-recognition molecules such as galectins for attachment to the vector or host invasion. It is noteworthy that most of (if not all) these pathogens and parasites are endowed with diverse and powerful mechanisms to evade intracellular killing by the host, and/or downregulate downstream immune responses. The complex strategies such as glycan mimicry developed by symbionts, commensals, and microbial pathogens to successfully colonize, enter, proliferate, and disseminate within and among their hosts or vectors, are the products of strong selective pressures that have led to adaptations that ensure their survival in the most hostile environment of all (Casadevall and Pirofski, 2000; Hughes and Sperandio, 2008). Thus, these fine tuned protein–carbohydrate interactions represent a significant challenge for the development of innovative strategies for intervention in infectious disease.

ACKNOWLEDGMENTS

The authors' research reviewed herein was supported by Grants IOS 1050518, IOB-0618409, and IOS0822257 from the National Science Foundation, and Grants 1R01GM070589-01 and 5R01GM070589-06 from the National Institutes of Health (NIH) to Gerardo R. Vasta, grant R41CA141970-01A2 from the NIH to Hafiz Ahmed, and grant 1R21AI076797-01A2 from the NIH to José A. Fernández-Robledo.

REFERENCES

Ahmad, N., Gabius, H.-J., Andre, S., Kaltner, H., Sabesan, S., Roy, R., Liu, B., Macaluso, F., and Brewer, C. F. (2004). Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. *J. Biol. Chem.* 279, 10841–10847.

Ahmed, H., Bianchet, M. A., Amzel, L. M., Hirabayashi, J., Kasai, K., Giga-Hama, Y., Tohda, H., and Vasta, G. R. (2002). Novel carbohydrate specificity of the 16-kDa galectin from *Caenorhabditis elegans*: binding to blood group precursor oligosaccharides (type 1, type 2, Talpha, and Tbeta) and gangliosides. *Glycobiology* 12, 451–461.

Ahmed, H., Du, S. J., O'Leary, N., and Vasta, G. R. (2004). Biochemical and molecular characterization of galectins from zebrafish (*Danio rerio*): notochord-specific expression of a prototype galectin during early embryogenesis. *Glycobiology* 14, 219–232.

Ahmed, H., Pohl, J., Fink, N. E., Strobel, F., and Vasta, G. R. (1996). The primary structure and carbohydrate specificity of a beta-galactosyl-binding lectin from toad (*Bufo arenarum* Hensel) ovary reveal closer similarities to the mammalian galectin-1 than to the galectin from the clawed frog *Xenopus laevis*. *J. Biol. Chem.* 271, 33083–33094.

Ahmed, H., and Vasta, G. R. (1994). Galectins: conservation of functionally and structurally relevant amino acid residues defines two types of carbohydrate recognition domains. *Glycobiology* 4, 545–548.

Ahmed, H., and Vasta, G. R. (2008). Unlike mammalian GRIFIN, the zebrafish homologue (DrGRIFIN) represents a functional carbohydrate-binding galectin. *Biochem. Biophys. Res. Commun.* 371, 350–355.

Akira, S., Takeda, K., and Kaisho, T. (2001). Toll-like receptors: critical proteins linking innate and acquired immunity. *Nat. Immunol.* 2, 675–680.

Anderson, A. C., Anderson, D. E., Bregoli, L., Hastings, W. D., Kassam, N., Lei, C., Chandwaskar, R., Karman, J., Su, E. W., Hirashima, M., Bruce, J. N., Kane, L. P., Kuchroo, V. K., and Hafler, D. A. (2007). Promotion of tissue inflammation by the immune receptor Tim-3 expressed on innate immune cells. *Science* 318, 1141–1143.

Bianchet, M. A., Ahmed, H., Vasta, G. R., and Amzel, L. M. (2000). Soluble beta-galactosyl-binding lectin (galectin) from toad ovary: crystallographic studies of two protein-sugar complexes. *Proteins* 40, 378–388.

Bianchet, M. A., Odom, E. W., Vasta, G. R., and Amzel, L. M. (2002). A novel fucose recognition fold involved in innate immunity. *Nat. Struct. Biol.* 9, 628–634.

Bianchet, M. A., Odom, E. W., Vasta, G. R., and Amzel, L. M. (2010). Structure and specificity of a binary tandem domain F-lectin from striped bass (*Morone saxatilis*). *J. Mol. Biol.* 401, 239–252.

Bittel, P., and Robatzek, S. (2007). Microbe-associated molecular patterns (MAMPs) probe plant immunity. *Curr. Opin. Plant Biol.* 10, 335–341.

Brewer, C. F., Miceli, M. C., and Baum, L. G. (2002). Clusters, bundles, arrays and lattices: novel mechanisms for lectin-saccharide-mediated cellular interactions. *Curr. Opin. Struct. Biol.* 12, 616–623.

Carlsson, S., Oberg, C. T., Carlsson, M. C., Sundin, A., Nilsson, U. J., Smith, D., Cummings, R. D., Almkvist, J., Karlsson, A., and Leffler, H. (2007). Affinity of galectin-8 and its carbohydrate recognition domains for ligands

in solution and at the cell surface. *Glycobiology* 17, 663–676.

Cartwright, J. R., Tharia, H. A., Burns, I., Shrive, A. K., Hoole, D., and Greenhough, T. J. (2004). Isolation and characterisation of pentraxin-like serum proteins from the common carp *Cyprinus carpio*. *Dev. Comp. Immunol.* 28, 113–125.

Casadevall, A., and Pirofski, L. A. (2000). Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infect. Immun.* 68, 6511–6518.

Cedeno-Laurent, F., and Dimitroff, C. J. (2012). Galectin-1 research in T cell immunity: past, present and future. *Clin. Immunol.* 142, 107–116.

Chen, H. Y., Fermin, A., Vardhana, S., Weng, I. C., Lo, K. F., Chang, E. Y., Maverakis, E., Yang, R. Y., Hsu, D. K., Dustin, M. L., and Liu, F. T. (2009). Galectin-3 negatively regulates TCR-mediated CD4+ T-cell activation at the immunological synapse. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14496–14501.

Cho, M., and Cummings, R. D. (1995). Galectin-1, a beta-galactoside-binding lectin in Chinese hamster ovary cells. II. Localization and biosynthesis. *J. Biol. Chem.* 270, 5207–5212.

Cleves, A. E., Cooper, D. N., Barondes, S. H., and Kelly, R. B. (1996). A new pathway for protein export in *Saccharomyces cerevisiae*. *J. Cell Biol.* 133, 1017–1026.

Colnot, C., Ripoche, M. A., Scaerou, F., Foulis, D., and Poirier, F. (1996). Galectins in mouse embryogenesis. *Biochem. Soc. Trans.* 24, 141–146.

Colnot, C., Sidhu, S. S., Balmain, N., and Poirier, F. (2001). Uncoupling of chondrocyte death and vascular invasion in mouse galectin 3 null mutant bones. *Dev. Biol.* 229, 203–214.

Cook, M. T., Hayball, P. J., Nowak, B. F., and Hayball, J. D. (2005). The opsonising activity of a pentraxin-like protein isolated from snapper (*Pagrus auratus*, Sparidae) serum. *Dev. Comp. Immunol.* 29, 703–712.

Cooper, D. N. (2002). Galectinomics: finding themes in complexity. *Biochim. Biophys. Acta* 1572, 209–231.

Cummings, R. D., and Nyame, A. K. (1999). *Schistosome* glysoconjugates. *Biochim. Biophys. Acta* 1455, 363–374.

Dam, T. K., and Brewer, C. F. (2008). Effects of clustered epitopes in multivalent ligand-receptor interactions. *Biochemistry* 47, 8470–8476.

Dam, T. K., and Brewer, C. F. (2010). Lectins as pattern recognition molecules: the effects of epitope density in innate immunity. *Glycobiology* 20, 270–279.

Dam, T. K., Gerken, T. A., and Brewer, C. F. (2009). Thermodynamics of multivalent carbohydrate-lectin cross-linking interactions: importance of entropy in the bind and jump mechanism. *Biochemistry* 48, 3822–3827.

Di Lella, S., Marti, M. A., Alvarez, R. M., Estrin, D. A., and Ricci, J. C. (2007). Characterization of the galectin-1 carbohydrate recognition domain in terms of solvent occupancy. *J. Phys. Chem. B* 111, 7360–7366.

Di Lella, S., Sundblad, V., Cerliani, J. P., Guardia, C. M., Estrin, D. A., Vasta, G. R., and Rabinovich, G. A. (2011). When galectins recognize glycans: from biochemistry to physiology and back again. *Biochemistry* 50, 7842–7857.

Dias-Baruffi, M., Zhu, H., Cho, M., Karmakar, S., Mcever, R. P., and Cummings, R. D. (2003). Dimeric galectin-1 induces surface exposure of phosphatidylserine and phagocytic recognition of leukocytes without inducing apoptosis. *J. Biol. Chem.* 278, 41282–41293.

Echeverria, I., and Amzel, L. M. (2011). Disaccharide binding to galectin-1: free energy calculations and molecular recognition mechanism. *Biophys. J.* 100, 2283–2292.

Elola, M. T., Wolfenstein-Todel, C., Troncoso, M. F., Vasta, G. R., and Rabinovich, G. A. (2007). Galectins: matricellular glycan-binding proteins linking cell adhesion, migration, and survival. *Cell. Mol. Life Sci.* 64, 1679–1700.

Fang, R., Mantle, M., and Ceri, H. (1993). Characterization of quail intestinal mucin as a ligand for endogenous quail lectin. *Biochem. J.* 293 (Pt 3), 867–872.

Feinberg, H., Park-Snyder, S., Kolatkar, A. R., Heise, C. T., Taylor, M. E., and Weis, W. I. (2000). Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor. *J. Biol. Chem.* 275, 21539–21548.

Fogel, S., Guittaut, M., Legrand, A., Monsigny, M., and Hebert, E. (1999). The tat protein of HIV-1 induces galectin-3 expression. *Glycobiology* 9, 383–387.

Fujita, T., Matsushita, M., and Endo, Y. (2004). The lectin-complement pathway – its role in innate immunity and evolution. *Immunol. Rev.* 198, 185–202.

Fukumori, T., Kanayama, H. O., and Raz, A. (2007). The role of galectin-3 in cancer drug resistance. *Drug Resist. Update* 10, 101–108.

Fukumori, T., Takenaka, Y., Yoshii, T., Kim, H. R., Hogan, V., Inohara, H., Kagawa, S., and Raz, A. (2003). CD29 and CD7 mediate galectin-3-induced type II T-cell apoptosis. *Cancer Res.* 63, 8302–8311.

Fuller, J. R., Pitzer, J. E., Godwin, U., Albertino, M., Machon, B. D., Kearse, K. P., and Mcconnell, T. J. (2004). Characterization of the molecular chaperone calnexin in the channel catfish, *Ictalurus punctatus*, and its association with MHC class II molecules. *Dev. Comp. Immunol.* 28, 603–617.

Garred, P., Larsen, F., Seyfarth, J., Fujita, R., and Madsen, H. O. (2006). Mannose-binding lectin and its genetic variants. *Genes Immun.* 7, 85–94.

Gerwick, L., Corley-Smith, G., and Bayne, C. J. (2007). Gene transcript changes in individual rainbow trout livers following an inflammatory stimulus. *Fish Shellfish Immunol.* 22, 157–171.

Gorski, J. P., Liu, F. T., Artigues, A., Castagna, L. F., and Osdoby, P. (2002). New alternatively spliced form of galectin-3, a member of the beta-galactoside-binding animal lectin family, contains a predicted transmembrane-spanning domain and a leucine zipper motif. *J. Biol. Chem.* 277, 18840–18848.

Greenspan, N. S. (2001). Dimensions of antigen recognition and levels of immunological specificity. *Adv. Cancer Res.* 80, 147–187.

Guevremont, M., Martel-Pelletier, J., Boileau, C., Liu, F. T., Richard, M., Fernandes, J. C., Pelletier, J. P., and Reboul, P. (2004). Galectin-3 surface expression on human adult chondrocytes: a potential substrate for collagenase-3. *Ann. Rheum. Dis.* 63, 636–643.

He, J., and Baum, L. G. (2004). Presentation of galectin-1 by extracellular matrix triggers T cell death. *J. Biol. Chem.* 279, 4705–4712.

Hijazi, K., Wang, Y., Scala, C., Jeffs, S., Longstaff, C., Stieh, D., Haggarty, B., Vanham, G., Schols, D., Balzarini, J., Jones, I. M., Hoxie, J., Shattock, R., and Kelly, C. G. (2011). DC-SIGN increases the affinity of HIV-1 envelope glycoprotein interaction with CD4. *PLoS ONE* 6, e28307. doi: 10.1371/journal.pone.0028307

Hirabayashi, J., and Kasai, K. (1993). The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. *Glycobiology* 3, 297–304.

Houzelstein, D., Goncalves, I. R., Fadden, A. J., Sidhu, S. S., Cooper, D. N., Drickamer, K., Leffler, H., and Poirier, F. (2004). Phylogenetic analysis of the vertebrate galectin family. *Mol. Biol. Evol.* 21, 1177–1187.

Hsu, D. K., Hammes, S. R., Kuwabara, I., Greene, W. C., and Liu, F. T. (1996). Human T lymphotropic virus-I infection of human T lymphocytes induces expression of the beta-galactoside-binding lectin, galectin-3. *Am. J. Pathol.* 148, 1661–1670.

Hughes, D. T., and Sperandio, V. (2008). Inter-kingdom signalling: communication between bacteria and their hosts. *Nat. Rev. Microbiol.* 6, 111–120.

Ideo, H., Matsuzaka, T., Nonaka, T., Seko, A., and Yamashita, K. (2011). Galectin-8-*N*-domain recognition mechanism for sialylated and sulfated glycans. *J. Biol. Chem.* 286, 11346–11355.

Ilarregui, J. M., Croci, D. O., Bianco, G. A., Toscano, M. A., Salatino, M., Vermeulen, M. E., Geffner, J. R., and Rabinovich, G. A. (2009). Tolerogenic signals delivered by dendritic cells to T cells through a galectin-1-driven immunoregulatory circuit involving interleukin 27 and interleukin 10. *Nat. Immunol.* 10, 981–991.

Iliev, D. B., Roach, J. C., Mackenzie, S., Planas, J. V., and Goetz, F. W. (2005). Endotoxin recognition: in fish or not in fish? *FEBS Lett.* 579, 6519–6528.

Ip, W. K., Takahashi, K., Ezekowitz, R. A., and Stuart, L. M. (2009). Mannose-binding lectin and innate immunity. *Immunol. Rev.* 230, 9–21.

Khalturin, K., Panzer, Z., Cooper, M. D., and Bosch, T. C. (2004). Recognition strategies in the innate immune system of ancestral chordates. *Mol. Immunol.* 41, 1077–1087.

Kingeter, L. M., and Lin, X. (2012). C-type lectin receptor-induced NF-kappaB activation in innate immune and inflammatory responses. *Cell. Mol. Immunol.* 9, 105–112.

Kohatsu, L., Hsu, D. K., Jegalian, A. G., Liu, F.-T., and Baum, L. G. (2006). Galectin-3 induces death of *Candida* species expressing specific beta-1,2-linked mannans. *J. Immunol.* 177, 4718–4726.

Krejcirikova, V., Pachl, P., Fabry, M., Maly, P., Rezacova, P., and Brynda, J. (2011). Structure of the mouse galectin-4 N-terminal carbohydrate-recognition domain reveals the mechanism of oligosaccharide recognition. *Acta Crystallogr. D Biol. Crystallogr.* 67, 204–211.

Li, Y., Feng, J., Geng, S., Wei, H., Chen, G., Li, X., Wang, L., Wang, R., Peng, H., Han, G., and Shen, B. (2011). The N- and C-terminal

carbohydrate recognition domains of galectin-9 contribute differently to its multiple functions in innate immunity and adaptive immunity. *Mol. Immunol.* 48, 670–677.

Liao, D. I., Kapadia, G., Ahmed, H., Vasta, G. R., and Herzberg, O. (1994). Structure of S-lectin, a developmentally regulated vertebrate beta-galactoside-binding protein. *Proc. Natl. Acad. Sci. U.S.A.* 91, 1428–1432.

Lipkowitz, M. S., Leal-Pinto, E., Cohen, B. E., and Abramson, R. G. (2004). Galectin 9 is the sugar-regulated urate transporter/channel UAT. *Glycoconj. J.* 19, 491–498.

Liu, F. T., and Hsu, D. K. (2007). The role of galectin-3 in promotion of the inflammatory response. *Drug News Perspect.* 20, 455–460.

Liu, F. T., and Rabinovich, G. A. (2005). Galectins as modulators of tumour progression. *Nat. Rev. Cancer* 5, 29–41.

Liu, F. T., Yang, R. Y., and Hsu, D. K. (2012). Galectins in acute and chronic inflammation. *Ann. N.Y. Acad. Sci.* 1253, 80–91.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. *Protein Sci.* 11, 1285–1299.

Lobsanov, Y. D., Gitt, M. A., Leffler, H., Barondes, S. H., and Rini, J. M. (1993). X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution. *J. Biol. Chem.* 268, 27034–27038.

Loeb, J. A., and Drickamer, K. (1988). Conformational changes in the chicken receptor for endocytosis of glycoproteins. Modulation of ligand-binding activity by Ca2+ and pH. *J. Biol. Chem.* 263, 9752–9760.

Ludwig, I. S., Geijtenbeek, T. B., and Van Kooyk, Y. (2006). Two way communication between neutrophils and dendritic cells. *Curr. Opin. Pharmacol.* 6, 408–413.

Lund, V., and Olafsen, J. A. (1999). Changes in serum concentration of a serum amyloid P-like pentraxin in Atlantic salmon, *Salmo salar* L., during infection and inflammation. *Dev. Comp. Immunol.* 23, 61–70.

Marschal, P., Herrmann, J., Leffler, H., Barondes, S. H., and Cooper, D. N. (1992). Sequence and specificity of a soluble lactose-binding lectin from *Xenopus laevis* skin. *J. Biol. Chem.* 267, 12942–12949.

Marth, J. D., and Grewal, P. K. (2008). Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* 8, 874–887.

Martinez, V. G., Pellizzari, E. H., Diaz, E. S., Cigorraga, S. B., Lustig, L., Denduchis, B., Wolfenstein-Todel, C., and Iglesias, M. M. (2004). Galectin-1, a cell adhesion modulator, induces apoptosis of rat Leydig cells in vitro. *Glycobiology* 14, 127–137.

Mascanfroni, I. D., Cerliani, J. P., Dergan-Dylon, S., Croci, D. O., Ilarregui, J. M., and Rabinovich, G. A. (2011). Endogenous lectins shape the function of dendritic cells and tailor adaptive immunity: mechanisms and biomedical applications. *Int. Immunopharmacol.* 11, 833–841.

Medzhitov, R., and Janeway, C. A. Jr. (2002). Decoding the patterns of self and nonself by the innate immune system. *Science* 296, 298–300.

Mercier, S., St-Pierre, C., Pelletier, I., Ouellet, M., Tremblay, M. J., and Sato, S. (2008). Galectin-1 promotes HIV-1 infectivity in macrophages through stabilization of viral adsorption. *Virology* 371, 121–129.

Miao, E. A., and Warren, S. E. (2010). Innate immune detection of bacterial virulence factors via the NLRC4 inflammasome. *J. Clin. Immunol.* 30, 502–506.

Morris, S., Ahmad, N., Andre, S., Kaltner, H., Gabius, H. J., Brenowitz, M., and Brewer, F. (2004). Quaternary solution structures of galectins-1, -3, and -7. *Glycobiology* 14, 293–300.

Myskova, J., Svobodova, M., Beverley, S. M., and Volf, P. (2007). A lipophosphoglycan-independent development of *Leishmania* in permissive sand flies. *Microbes Infect.* 9, 317–324.

Nagae, M., Nishi, N., Murata, T., Usui, T., Nakamura, T., Wakatsuki, S., and Kato, R. (2009). Structural analysis of the recognition mechanism of poly-*N*-acetyllactosamine by the human galectin-9 N-terminal carbohydrate recognition domain. *Glycobiology* 19, 112–117.

Nonaka, M. (2011). The complement C3 protein family in invertebrates. *ISJ* 8, 21–32.

Ogden, A. T., Nunes, I., Ko, K., Wu, S., Hines, C. S., Wang, A. F., Hegde, R. S., and Lang, R. A. (1998). GRIFIN, a novel lens-specific protein related to the galectin family. *J. Biol. Chem.* 273, 28889–28896.

Ohtsubo, K., and Marth, J. D. (2006). Glycosylation in cellular mechanisms of health and disease. *Cell* 126, 855–867.

Ohtsubo, K., Takamatsu, S., Minowa, M. T., Yoshida, A., Takeuchi, M., and Marth, J. D. (2005). Dietary and genetic control of glucose transporter 2 glycosylation promotes insulin secretion in suppressing diabetes. *Cell* 123, 1307–1321.

Okumura, C. Y., Baum, L. G., and Johnson, P. J. (2008). Galectin-1 on cervical epithelial cells is a receptor for the sexually transmitted human parasite *Trichomonas vaginalis*. *Cell. Microbiol.* 10, 2078–2090.

Ouellet, M., Mercier, S., Pelletier, I., Bounou, S., Roy, J., Hirabayashi, J., Sato, S., and Tremblay, M. J. (2005). Galectin-1 acts as a soluble host factor that promotes HIV-1 infectivity through stabilization of virus attachment to host cells. *J. Immunol.* 174, 4120–4126.

Ozaki, K., Inoue, K., Sato, H., Iida, A., Ohnishi, Y., Sekine, A., Odashiro, K., Nobuyoshi, M., Hori, M., Nakamura, Y., and Tanaka, T. (2004). Functional variation in LGALS2 confers risk of myocardial infarction and regulates lymphotoxin-alpha secretion in vitro. *Nature* 429, 72–75.

Partridge, E. A., Le Roy, C., Di Guglielmo, G. M., Pawling, J., Cheung, P., Granovsky, M., Nabi, I. R., Wrana, J. L., and Dennis, J. W. (2004). Regulation of cytokine receptors by Golgi *N*-glycan processing and endocytosis. *Science* 306, 120–124.

Patterson, R. J., Wang, W., and Wang, J. L. (2004). Understanding the biochemical activities of galectin-1 and galectin-3 in the nucleus. *Glycoconj. J.* 19, 499–506.

Perillo, N. L., Marcus, M. E., and Baum, L. G. (1998). Galectins: versatile modulators of cell adhesion, cell proliferation, and cell death. *J. Mol. Med. (Berl)* 76, 402–412.

Perkins, F. O. (1996). The structure of *Perkinsus marinus* (Mackin, Owen and Collier, 1950) Levine, 1978 with comments on taxonomy and phylogeny of *Perkinsus* spp. *J. Shellfish Res.* 15, 67–87.

Poirier, F. (2002). Roles of galectins in vivo. *Biochem. Soc. Symp.* 69, 95–103.

Poirier, F., Timmons, P. M., Chan, C. T., Guenet, J. L., and Rigby, P. W. (1992). Expression of the L14 lectin during mouse embryogenesis suggests multiple roles during pre- and post-implantation development. *Development* 115, 143–155.

Puche, A. C., Poirier, F., Hair, M., Bartlett, P. F., and Key, B. (1996). Role of galectin-1 in the developing mouse olfactory system. *Dev. Biol.* 179, 274–287.

Rabinovich, G. A., Baum, L. G., Tinari, N., Paganelli, R., Natoli, C., Liu, F. T., and Iacobelli, S. (2002a). Galectins and their ligands: amplifiers, silencers or tuners of the inflammatory response? *Trends Immunol.* 23, 313–320.

Rabinovich, G. A., Liu, F. T., Hirashima, M., and Anderson, A. (2007a). An emerging role for galectins in tuning the immune response: lessons from experimental models of inflammatory disease, autoimmunity and cancer. *Scand. J. Immunol.* 66, 143–158.

Rabinovich, G. A., Rubinstein, N., and Toscano, M. A. (2002b). Role of galectins in inflammatory and immunomodulatory processes. *Biochim. Biophys. Acta* 1572, 274–284.

Rabinovich, G. A., and Toscano, M. A. (2009). Turning "sweet" on immunity: galectin-glycan interactions in immune tolerance and inflammation. *Nat. Rev. Immunol.* 9, 338–352.

Rabinovich, G. A., Toscano, M. A., Jackson, S. S., and Vasta, G. R. (2007b). Functions of cell surface galectin-glycoprotein lattices. *Curr. Opin. Struct. Biol.* 17, 513–520.

Rapoport, E. M., Andre, S., Kurmyshkina, O. V., Pochechueva, T. V., Severov, V. V., Pazynina, G. V., Gabius, H.-J., and Bovin, N. V. (2008). Galectin-loaded cells as a platform for the profiling of lectin specificity by fluorescent neoglycoconjugates: a case study on galectins-1 and -3 and the impact of assay setting. *Glycobiology* 18, 315–324.

Rubinstein, N., Ilarregui, J. M., Toscano, M. A., and Rabinovich, G. A. (2004). The role of galectins in the initiation, amplification and resolution of the inflammatory response. *Tissue Antigens* 64, 1–12.

Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., Dowse, T. J., Jakle, U., Ramboarina, S., Shivarattan, T., Matthews, S., and Soldati-Favre, D. (2005). A novel galectin-like domain from *Toxoplasma gondii* micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex. *J. Biol. Chem.* 280, 38583–38591.

Sato, A., Mayer, W. E., Overath, P., and Klein, J. (2003). Genes encoding putative natural killer cell C-type lectin receptors in teleostean fishes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7779–7784.

Sato, S., and Hughes, R. C. (1992). Binding specificity of a baby hamster kidney lectin for H type I and II chains, polylactosamine glycans, and appropriately glycosylated forms of laminin and fibronectin. *J. Biol. Chem.* 267, 6983–6990.

Sato, S., and Hughes, R. C. (1994). Regulation of secretion and surface

expression of Mac-2, a galactoside-binding protein of macrophages. *J. Biol. Chem.* 269, 4424–4430.

Sato, S., and Nieminen, J. (2004). Seeing strangers or announcing "danger": galectin-3 in two models of innate immunity. *Glycoconj. J.* 19, 583–591.

Schroder, H. C., Ushijima, H., Theis, C., Seve, A. P., Hubert, J., and Muller, W. E. (1995). Expression of nuclear lectin carbohydrate-binding protein 35 in human immunodeficiency virus type 1-infected Molt-3 cells. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* 9, 340–348.

Schwarz, F. P., Ahmed, H., Bianchet, M. A., Amzel, L. M., and Vasta, G. R. (1998). Thermodynamics of bovine spleen galectin-1 binding to disaccharides: correlation with structure and its effect on oligomerization at the denaturation temperature. *Biochemistry* 37, 5867–5877.

Seetharaman, J., Kanigsberg, A., Slaaby, R., Leffler, H., Barondes, S. H., and Rini, J. M. (1998). X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-A resolution. *J. Biol. Chem.* 273, 13047–13052.

Seong, S.-Y., and Matzinger, P. (2004). Hydrophobicity: an ancient damage-associated molecular pattern that initiates innate immune responses. *Nat. Rev. Immunol.* 4, 469–478.

Shirai, T., Matsui, Y., Shionyu-Mitsuyama, C., Yamane, T., Kamiya, H., Ishii, C., Ogawa, T., and Muramoto, K. (2002). Crystal structure of a conger eel galectin (congerin II) at 1.45A resolution: implication for the accelerated evolution of a new ligand-binding site following gene duplication. *J. Mol. Biol.* 321, 879–889.

Shirai, T., Mitsuyama, C., Niwa, Y., Matsui, Y., Hotta, H., Yamane, T., Kamiya, H., Ishii, C., Ogawa, T., and Muramoto, K. (1999). High-resolution structure of the conger eel galectin, congerin I, in lactose-liganded and ligand-free forms: emergence of a new structure class by accelerated evolution. *Structure* 7, 1223–1233.

Shoji, H., Deltour, L., Nakamura, T., Tajbakhsh, S., and Poirier, F. (2009). Expression pattern and role of Galectin1 during early mouse myogenesis. *Dev. Growth Differ.* 51, 607–615.

Shoji, H., Nishi, N., Hirashima, M., and Nakamura, T. (2003). Characterization of the *Xenopus* galectin family. Three structurally different types as in mammals and regulated expression during embryogenesis. *J. Biol. Chem.* 278, 12285–12293.

Sparrow, C. P., Leffler, H., and Barondes, S. H. (1987). Multiple soluble beta-galactoside-binding lectins from human lung. *J. Biol. Chem.* 262, 7383–7390.

Stalz, H., Roth, U., Schleuder, D., Macht, M., Haebel, S., Strupat, K., Peter-Katalinic, J., and Hanisch, F. G. (2006). The *Geodia cydonium* galectin exhibits prototype and chimera-type characteristics and a unique sequence polymorphism within its carbohydrate recognition domain. *Glycobiology* 16, 402–414.

Stowell, S. R., Arthur, C. M., Dias-Baruffi, M., Rodrigues, L. C., Gourdine, J. P., Heimburg-Molinaro, J., Ju, T., Molinaro, R. J., Rivera-Marrero, C., Xia, B., Smith, D. F., and Cummings, R. D. (2010). Innate immune lectins kill bacteria expressing blood group antigen. *Nat. Med.* 16, 295–301.

Stowell, S. R., Cho, M., Feasley, C. L., Arthur, C. M., Song, X., Colucci, J. K., Karmakar, S., Mehta, P., Dias-Baruffi, M., Mcever, R. P., and Cummings, R. D. (2009). Ligand reduces galectin-1 sensitivity to oxidative inactivation by enhancing dimer formation. *J. Biol. Chem.* 284, 4989–4999.

Stowell, S. R., Qian, Y., Karmakar, S., Koyama, N. S., Dias-Baruffi, M., Leffler, H., Mcever, R. P., and Cummings, R. D. (2008). Differential roles of galectin-1 and galectin-3 in regulating leukocyte viability and cytokine secretion. *J. Immunol.* 180, 3091–3102.

Takenaka, Y., Fukumori, T., and Raz, A. (2004). Galectin-3 and metastasis. *Glycoconj. J.* 19, 543–549.

Tasumi, S., and Vasta, G. R. (2007). A galectin of unique domain organization from hemocytes of the Eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite *Perkinsus marinus*. *J. Immunol.* 179, 3086–3098.

Tomizawa, T., Kigawa, T., Saito, K., Koshiba, S., Inoue, M., and Yokoyama, S. (2005). Solution structure of the C-terminal gal-bind lectin domain from human galectin-4. Structural Genomics and Proteomics Initiative (RSGI), Riken.

Troncoso, M. F., Elola, M. T., Croci, D. O., and Rabinovich, G. A. (2012). Integrating structure and function of "tandem-repeat" galectins. *Front. Biosci. (Schol Ed)* 4, 864–887.

van der Vlist, M., van der Aar, A. M., Gringhuis, S. I., and Geijtenbeek, T. B. (2011). Innate signaling in HIV-1 infection of dendritic cells. *Curr. Opin. HIV AIDS* 6, 348–352.

Vasta, G. R. (2009). Roles of galectins in infection. *Nat. Rev. Microbiol.* 7, 424–438.

Vasta, G. R., and Ahmed, H. (2009). *Animal Lectins: A Functional View.* Boca Raton, FL: CRC Press (Taylor & Francis Group).

Vasta, G. R., Ahmed, H., Amzel, L. M., and Biahchet, M. A. (1997). Galectins from amphibian species: carbohydrate specificity, molecular structure and evolution. *Trends Blycosci. Glycotechnol.* 9, 131–144.

Vasta, G. R., Hunt, J. C., Marchalonis, J. J., and Fish, W. W. (1986). Galactosyl-binding lectins from the tunicate *Didemnum candidum*. Purification and physicochemical characterization. *J. Biol. Chem.* 261, 9174–9181.

Vasta, G. R., Quesenberry, M., Ahmed, H., and O'Leary, N. (1999). C-type lectins and galectins mediate innate and adaptive immune functions: their roles in the complement activation pathway. *Dev. Comp. Immunol.* 23, 401–420.

Vilches, C., and Parham, P. (2002). KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu. Rev. Immunol.* 20, 217–251.

Vokhmyanina, O. A., Rapoport, E. M., Ryzhov, I. M., Korchagina, E. Y., Pazynina, G. V., Severov, V. V., Kaltner, H., Andre, S., Gabius, H. J., and Bovin, N. V. (2011). Carbohydrate specificity of chicken and human tandem-repeat-type galectins-8 in composition of cells. *Biochemistry (Mosc)* 76, 1185–1192.

Volf, P., Benkova, I., Myskova, J., Sadlova, J., Campino, L., and Ravel, C. (2007). Increased transmission potential of *Leishmania major/Leishmania infantum* hybrids. *Int. J. Parasitol.* 37, 589–593.

Vyakarnam, A., Dagher, S. F., Wang, J. L., and Patterson, R. J. (1997). Evidence for a role for galectin-1 in pre-mRNA splicing. *Mol. Cell. Biol.* 17, 4730–4737.

Wallis, R. (2002). Structural and functional aspects of complement activation by mannose-binding protein. *Immunobiology* 205, 433–445.

Walser, P. J., Haebel, P. W., Kunzler, M., Sargent, D., Kues, U., Aebi, M., and Ban, N. (2004). Structure and functional analysis of the fungal galectin CGL2. *Structure* 12, 689–702.

Walser, P. J., Kues, U., Aebi, M., and Kunzler, M. (2005). Ligand interactions of the *Coprinopsis cinerea* galectins. *Fungal Genet. Biol.* 42, 293–305.

Weis, W. I., Drickamer, K., and Hendrickson, W. A. (1992). Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature* 360, 127–134.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., and Hendrickson, W. A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. *Science* 254, 1608–1615.

Weis, W. I., Taylor, M. E., and Drickamer, K. (1998). The C-type lectin superfamily in the immune system. *Immunol. Rev.* 163, 19–34.

Yoshino, T. P., Dinguirard, N., Kunert, J., and Hokke, C. H. (2008). Molecular and functional characterization of a tandem-repeat galectin from the freshwater snail *Biomphalaria glabrata*, intermediate host of the human blood fluke *Schistosoma mansoni*. *Gene* 411, 46–58.

Zelensky, A. N., and Gready, J. E. (2005). The C-type lectin-like domain superfamily. *FEBS J.* 272, 6179–6217.

Zhou, Q., and Cummings, R. D. (1990). The S-type lectin from calf heart tissue binds selectively to the carbohydrate chains of laminin. *Arch. Biochem. Biophys.* 281, 27–35.

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.

Received: 26 April 2012; accepted: 26 June 2012; published online: 13 July 2012.
*Citation: Vasta GR, Ahmed H, Nita-Lazar M, Banerjee A, Pasek M, Shridhar S, Guha P and Fernández-Robledo JA (2012) Galectins as self/non-self recognition receptors in innate and adaptive immunity: an unresolved paradox. Front. Immun. **3**:199. doi: 10.3389/fimmu.2012.00199*
This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

ORIGINAL RESEARCH ARTICLE
published: 10 May 2012
doi: 10.3389/fimmu.2012.00091

Understanding the role of host hemocytes in a squid/*Vibrio* symbiosis using transcriptomics and proteomics

Andrew J. Collins*†, *Tyler R. Schleicher*†, *Bethany A. Rader* and *Spencer V. Nyholm*

Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Ioannis Eleftherianos, The George Washington University, USA
Virginia Weis, Oregon State University, USA

****Correspondence:***
Spencer V. Nyholm, Department of Molecular and Cell Biology, University of Connecticut, 91 N. Eagleville Road, Unit 3125, Storrs, CT 06269, USA.
e-mail: spencer.nyholm@uconn.edu

†Andrew J. Collins and Tyler R. Schleicher have contributed equally to this work.

The symbiosis between the squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*, serves as a model for understanding interactions between beneficial bacteria and animal hosts. The establishment and maintenance of the association is highly specific and depends on the selection of *V. fischeri* and exclusion of non-symbiotic bacteria from the environment. Current evidence suggests that the host's cellular innate immune system, in the form of macrophage-like hemocytes, helps to mediate host tolerance of *V. fischeri*. To begin to understand the role of hemocytes in this association, we analyzed these cells by high-throughput 454 transcriptomic and liquid chromatography/tandem mass spectrometry (LC-MS/MS) proteomic analyses. 454 high-throughput sequencing produced 650, 686 reads totaling 279.9 Mb while LC-MS/MS analyses of circulating hemocytes putatively identified 702 unique proteins. Several receptors involved with the recognition of microbial-associated molecular patterns were identified. Among these was a complete open reading frame to a putative peptidoglycan recognition protein (EsPGRP5) with conserved residues for amidase activity. Assembly of the hemocyte transcriptome showed EsPGRP5 had high coverage, suggesting it is among the 5% most abundant transcripts in circulating hemocytes. Other transcripts and proteins identified included members of the conserved NF-κB signaling pathway, putative members of the complement pathway, the carbohydrate binding protein galectin, and cephalotoxin. Quantitative Real-Time PCR of complement-like genes, cephalotoxin, EsPGRP5, and a nitric oxide synthase showed differential expression in circulating hemocytes from adult squid with colonized light organs compared to those isolated from hosts where the symbionts were removed. These data suggest that the presence of the symbiont influences gene expression of the cellular innate immune system of *E. scolopes*.

Keywords: innate immunity, symbiosis, *Euprymna scolopes*, *Vibrio fischeri*, hemocytes, transcriptomics, proteomics

INTRODUCTION

All animals have co-evolved associations with microorganisms. These symbiotic associations have been a major driving force in eukaryotic evolution, including the vastly influential bacterial endosymbiotic origin of essential organelles as championed by the late Lynn Margulis (Sagan, 1967). Given that all metazoans evolved in close and intimate association with microorganisms, it is not surprising that co-evolved microbiota also influenced the evolution of other aspects of animal biology. There has been a long-standing view that the immune system evolved to counter infection by pathogens and to exclude microorganisms from animals. However, recent evidence suggests that the immune system also plays an important role as an interface between hosts and their natural co-evolved benign or beneficial microbiota (Gross et al., 2009; McFall-Ngai et al., 2010; Royet et al., 2011). This interface involves a "dialog" between host and symbiont(s) that ensures that these associations are established and maintained. Growing evidence suggests that these beneficial microorganisms have influenced the evolution of both the innate and adaptive immune systems (McFall-Ngai, 2007; Lee and Mazmanian, 2010; McFall-Ngai et al., 2010). Gnotobiotic studies in mammals have demonstrated that proper development of components of the adaptive immune system require the presence of the host's normal microbiota. For example the development of the gut-associated lymphoid tissue and differentiation of various T-cell classes are greatly influenced by the bacterial communities in the gut (reviewed in Bauer et al., 2006; Duan and Kasper, 2011; Olszak et al., 2012). It has also been proposed that perhaps the adaptive immune system itself evolved to deal with an increasingly complex gut microbiota (McFall-Ngai, 2007).

Invertebrates, which comprise the vast majority of animals, lack an adaptive immune system and must rely solely on innate immunity to interface with a microbial world. Despite any obvious mechanism to achieve the precision with which the adaptive immune system targets specific pathogens, there are many examples of highly specific associations formed between microorganisms and invertebrates (e.g., as found in insect bacteriocytes, termite hindguts, the trophosomes of hydrothermal vent tubeworms, and the light organs of squids; Nyholm and McFall-Ngai, 2004; Stewart and Cavanaugh, 2006; Moran et al., 2008; Feldhaar and Gross, 2009; Watanabe and Tokuda, 2010). What are the mechanisms for achieving such specificity? Many of these associations

exist by creating specialized microenvironments that favor colonization by microorganisms and sequester them to specific cells or tissues (e.g., bacteriocytes; anaerobic, pH, or nutrient-specific environments as in the gut; or specialized tissues as in light organs; Nyholm and McFall-Ngai, 2004; Bäckhed et al., 2005; Stoll et al., 2010). However, increasing evidence suggests that the innate immune system of animal hosts is also important in maintaining these associations (Gross et al., 2009; Nyholm et al., 2009; McFall-Ngai et al., 2010; Krasity et al., 2011; Weiss et al., 2011).

The light organ symbiosis between the Hawaiian bobtail squid, *Euprymna scolopes*, and the bioluminescent bacterium, *Vibrio fischeri*, is used as a model system to study the mechanisms by which beneficial bacteria colonize animal hosts (**Figure 1**; McFall-Ngai, 2002; Nyholm and McFall-Ngai, 2004; Visick and Ruby, 2006). The symbionts reside within the epithelia-lined crypt spaces of a specialized light organ (to densities of 10^9 *V. fischeri* cells/adult squid). This organ is directly connected to the environment via ciliated ducts that facilitate the environmental transmission of the symbionts. Juvenile squid must select *V. fischeri* from a background of 10^6 non-symbiotic bacteria/ml of seawater (Nyholm and McFall-Ngai, 2004). There are many mechanisms by which the partners ensure this specificity and the host's innate immune system is important in this process. For example, innate immunity effectors like reactive oxygen species (ROS), putative complement members, and phagocytic hemocytes, along with the recognition of microbial-associated molecular patterns (MAMPs) by host pattern recognition receptors (PRRs) have all been implicated in playing a role in this association (Weis et al., 1996; Nyholm and McFall-Ngai, 1998; Davidson et al., 2004; Koropatnick et al., 2004, 2007; Castillo et al., 2009; Nyholm et al., 2009; Troll et al., 2009, 2010; McFall-Ngai et al., 2010, 2012; Heath-Heckman and McFall-Ngai, 2011; Krasity et al., 2011; Schleicher and Nyholm, 2011). Because the symbiosis is binary with a single host and symbiont, it offers the opportunity to ask how such high specificity is established and maintained in the context of interactions with the innate immune system of the host.

Euprymna scolopes, like other cephalopods, has a single type of macrophage-like hemocyte (**Figure 1**). These hemocytes can traverse between tissues and enter into the crypt spaces where the symbionts reside and appear to "sample" these spaces (Nyholm and McFall-Ngai, 1998). Previous studies of hemocytes from *E. scolopes* have indicated that they behave like macrophages, binding, and phagocytosing bacteria (Nyholm and McFall-Ngai, 1998; Nyholm et al., 2009; Heath-Heckman and McFall-Ngai, 2011). Although the crypts of juvenile symbiotic *E. scolopes* may contain hemocytes with engulfed bacterial cells, hemocytes in adult crypts have never been observed with phagocytosed bacteria despite being found in a microenvironment with high densities of *V. fischeri*. These observations suggest that the squid's hemocyte response may change during development of the association. A study of adult hemocytes revealed that these cells could also differentiate between symbiotic and non-symbiotic bacteria (Nyholm et al., 2009). *In vitro* binding assays demonstrated that hemocytes differentially bound various species within the *Vibrio* group. *V. fischeri* bound poorly to hemocytes from hosts with fully colonized light organs and significantly more to blood cells from hosts from which the symbionts were removed by curing with antibiotics. No significant change was observed for non-symbiotic bacteria. Taken together, these data suggest that hemocytes can differentiate between *V. fischeri* and other closely related members of the Vibrionaceae and colonization alters the hemocytes' ability to recognize the symbiont. The phenomenon of hemocytes changing their binding affinity to *V. fischeri* after colonization may be analogous to vertebrate immune "tolerance", leading to homeostasis (symbiostasis) of the association. This evidence leads us to ask; how can the innate immune system achieve such specificity and how does the symbiont influence the immune response?

In order to better understand the role of host hemocytes in the squid/*Vibrio* symbiosis we have used high-throughput sequencing and liquid chromatography/tandem mass spectrometry (LC-MS/MS) to characterize the transcriptome and proteome in circulating blood cells of adult hosts with fully colonized light

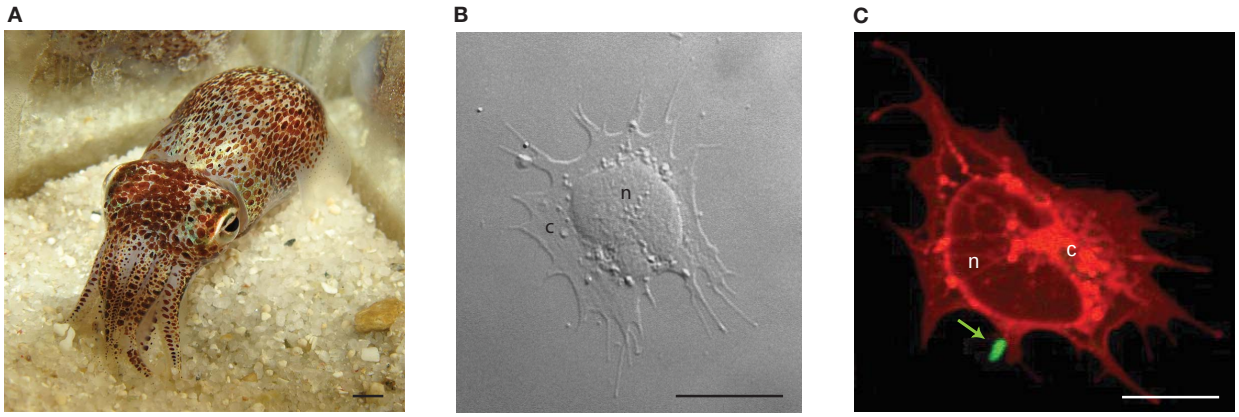


FIGURE 1 | The Hawaiian bobtail squid, *Euprymna scolopes*, and its hemocytes. (A) Adult *E. scolopes*. Scale, 3 mm. **(B)** Hemocyte of the host as viewed by DIC imaging. The macrophage-like hemocyte is the only circulating blood cell type found in the host. These cells can infiltrate the crypt spaces of the light organ where they can interact with the symbiont, *V. fischeri*. Scale, 10 μm. **(C)** Host hemocyte stained with CellTracker (red) and viewed by confocal microscopy. A sole *V. fischeri* cell (green; arrow) can be seen bound to the hemocyte. Scale, 10 μm. n, nucleus; c, cytoplasm.

organs. In addition, we have used quantitative RT-PCR (qRT-PCR) to compare expression of several innate immunity genes from both symbiotic and cured hosts, to understand the molecular mechanisms by which the hemocyte response may change due to colonization.

MATERIALS AND METHODS

ANIMAL COLLECTIONS

Adult animals were collected in shallow sand flats off of Oahu, HI by dip net and were either maintained in the laboratory in re-circulating natural seawater aquaria at the Hawaii Institute of Marine Biology or at the University of Connecticut with artificial seawater (ASW, Instant Ocean, IO) at 23˚C. All animals were acclimated at least 48 h under laboratory conditions and kept on an approximate 12 h light/12 h dark cycle before sample collection (Schleicher and Nyholm, 2011).

HEMOCYTE COLLECTION

Squid hemocytes were isolated from adult *E. scolopes* as previously described (Nyholm et al., 2009; Collins and Nyholm, 2010). Animals were first anesthetized in a 2% solution of ethanol in seawater. Hemolymph was extracted from the cephalic artery, located between the eyes, using a sterile 1-ml syringe with a 28-gauge needle. An average of 50–100 µl of hemolymph (∼5,000 hemocytes/µl) was obtained per animal using this method. Freshly collected hemocytes were washed and resuspended in Squid Ringer's solution and hemocyte concentrations were determined by hemocytometer.

CURING EXPERIMENTS

To generate naïve hemocytes, adult *E. scolopes* were cured of their population of *V. fischeri* symbionts (Nyholm et al., 2009). Briefly, adult squid were maintained individually in 5-gal tanks containing IO ASW (Aquarium Systems). For one set of animals, chloramphenicol and gentamicin was added to the seawater to a final concentration of 20 µg/m1 of each. The concentration of antibiotics used in these experiments effectively eliminates *V. fischeri* symbionts from the light organ without compromising the health of the host in any detectable way (Doino and McFall-Ngai, 1995; Lamarcq and McFall-Ngai, 1998; Foster et al., 2000; Nyholm et al., 2002, 2009; Koropatnick et al., 2007). The animals were transferred daily into fresh IO, either with or without antibiotics, for 5 days, and samples of hemolymph were collected on day 5. The resulting two sets of hemocytes were designated either "normal" (untreated/symbiotic) or "naïve" (treated/cured). The effectiveness of the antibiotics was determined by dissecting and homogenizing the central core of the light organ, and plating an aliquot of the homogenate on seawater tryptone agar (Nyholm et al., 2009).

TRANSCRIPTOME SEQUENCING

RNA was extracted from the hemocytes of 11 adult squid using the RNAqueous-Micro kit (Ambion, Life Technologies) and treated with Turbo DNAse (Ambion, Life Technologies). Ribosomal RNA was removed using a eukaryote RiboMinus kit (Invitrogen) and quantified using an Agilent Bioanalyzer PicoRNA chip. Two-hundred nanograms of treated and pooled RNA (approximately 18 ng from the hemocytes of each of the 11 animals) was used to make a 454 cDNA library according to standard protocol (Roche). The library was titrated with an SVemPCR kit and sequenced on one-half of a 454 picotiter plate.

454 BIOINFORMATICS

454 sequencing reads were assembled using the CLC Genomic Workbench with standard *de novo* assembly parameters (mismatch cost = 2, insertion cost = 3, deletion cost = 3, length fraction = 0.5, similarity = 0.8, conflicts resolved by voting, minimum contig length = 200 bases). Contigs were used to query the UniProt and NCBI nr protein databases using blastx with a maximum *e*-value of 0.01. Significant alignments were further analyzed using SignalP (Petersen et al., 2011), InterProScan (Zdobnov and Apweiler, 2001), and ClustalX (Jeanmougin et al., 1998) using default parameters. The output from the blastx searches were imported into the program BLAST2GO (Conesa et al., 2005) and were used to assign Gene Ontology categories. Level 2 hierarchies were used for comparison between proteomic and transcriptomic sequencing. BLAST results were also used to make annotated protein databases to be searched with MS/MS peptide data.

PROTEIN EXTRACTION

Hemocytes collected from an adult symbiotic squid were homogenized in 64 mM Tris pH 8, 1% SDS, 1× protease inhibitor cocktail (Sigma Aldrich, P2714). After the homogenate was centrifuged (Eppendorf 5810 R, 14,000 rpm, 4˚C, 30 min), the protein concentration of the supernatant was quantified spectrophotometrically using the *RC DC* protein assay (Bio-Rad, Hercules, CA, USA). Alternatively, hemocyte proteins from adult symbiotic squid were extracted using the Ready Prep Protein Extraction Kit (Membrane I) as described by the manufacturer's protocol (Bio-Rad, Hercules, CA, USA). Hemocyte proteins from two adult squid were separated into either a hydrophilic or a hydrophobic layer based on temperature dependent phase partitioning. The protein concentration of each layer was determined with the *RC DC* protein assay as described above (Bio-Rad, Hercules, CA, USA).

PROTEIN DIGESTION

Proteins were digested as previously described (Schleicher and Nyholm, 2011). Hemocyte proteins were precipitated with 10% (w/v) trichloroacetic acid (Fisher Scientific) at 4˚C overnight. The precipitates were collected by centrifugation (Eppendorf 5810 R, 11,000 × *g*, 30 min, 4˚C) and washed twice with ice-cold acetone. Protein pellets were briefly air-dried and then resuspended in 25 µl of 8 M urea, 0.4 M ammonium bicarbonate, pH 8.0. Samples were reduced and alkylated with 5 µl of 45 mM dithiothreitol (DTT; Acros Organics) at 37˚C for 20 min and 5 µl of 100 mM iodoacetamide (Acros Organics) at room temperature in the dark for 20 additional minutes. Sequencing grade trypsin was added 1:15 (w/w enzyme to protein; Promega, V5111). The solutions were diluted in water to 100 µl (2 M urea final concentration). Both samples were digested at 37˚C for 18–24 h and then stored at −80˚C until further analysis.

PROTEIN IDENTIFICATION USING NANOLC-MS/MS

Protein identifications were completed at the W.M. Keck Biotechnology Resource Laboratory at Yale University as previously

described (Schleicher and Nyholm, 2011). Briefly, 5 μl (0.375 μg) of the hydrophilic and hydrophobic fraction of hemocyte protein extraction from the Ready Prep Protein Extraction Kit I were analyzed on a LTQ Orbitrap equipped with a Waters nanoAcquity UPLC system. Samples were concentrated at 15 μl/min with 99% Buffer A (100% water, 0.1% formic acid) for 1 min on a trap column (Waters Symmetry® C18 180 μm × 20 mm). Peptides were separated at 300 nl/min with Buffer A and Buffer B (100% CH_3CN, 0.075% formic acid) using a 1.7-μm, 75 μm × 250 mm nanoAcquity™ UPLC™ column (35°C). A 51-min linear gradient was established using increasing amounts of Buffer B (95% Buffer A, 5% Buffer B at initial conditions, 50% A, 50% B, at 50 min, and 15% A, 85% B at 51 min). MS was acquired in the Orbitrap using one microscan and an inject time of 900 followed by four data dependent MS/MS acquisitions in the ion trap.

Hemocyte proteins from an adult symbiotic squid were also identified using a 5600 TripleTOF (AB SCIEX, Concord, ON, Canada) equipped with a Waters nanoAcquity UPLC system. Briefly, 5 μl (3 μg) of sample were loaded onto a Waters Symmetry® C18 180 μm × 20 mm trap column at 15 μl/min with 99% Buffer A for 1 min. Peptides were separated with a 1.7-μm, 75-μm × 150-mm nanoAcquity™ UPLC™ column (45°C) using a 161-min linear gradient at 500 nl/min. The gradient included 95% Buffer A, 5% Buffer B at initial conditions, 60% A, 40% B at 160 min, and 15% A, 85% B at 161 min. MS was acquired from 400 to 1250 Da for 0.25 s. While in high sensitivity mode, 20 data dependent MS/MS were acquired for 0.05 s each.

PROTEOMIC DATA ANALYSIS

MS/MS spectra were converted to Mascot compatible files using either the Mascot Distiller (Orbitrap MS/MS) or the AB Sciex MGF Converter (5600 TripleTOF MS/MS) and searched with the Mascot algorithm (Version 2.3.01; Hirosawa et al., 1993; Perkins et al., 1999) using an *E. scolopes* protein sequence database (25,745 sequences) generated from available transcriptomic data: juvenile light organ ESTs (Chun et al., 2006) and hemocyte transcripts (this study). Searching parameters included peptide charge states of +2 or +3, partial methionine oxidation, carboxamidomethylated cysteine, a peptide tolerance of ±20 ppm, and MS/MS fragment tolerance of ±0.6 Da. A decoy database search was also performed by Mascot.

Scaffold (version Scaffold_3.2.0, Proteome Software Inc., Portland, OR, USA) was used to validate MS/MS based peptide and protein identifications by Mascot. In addition, Scaffold applied X! Tandem (Craig and Beavis, 2004) [The GPM, thegpm.org; version CYCLONE (2010.12.01.1)] as a supplementary search to further confirm protein and peptide identifications. X! Tandem applied the same search parameters as Mascot. Peptide and protein identifications were accepted at greater than 95.0% probabilities each. Only proteins that met the necessary criteria and contained two or more peptides were considered as positive identifications. For positive identifications, the false discovery rates (FDRs) were 0% at both the protein and peptide level as calculated by Scaffold. Putative protein identifications included proteins with only one matching peptide (protein FDR, 0.4%).

POLYMERASE CHAIN REACTION

Total RNA from adult symbiotic or naïve hemocytes was extracted using TRIzol reagent (Life Technologies) according to the manufacturer's instructions. Approximately 1–2 μg of RNA per animal was isolated. Subsequent PCR and qRT-PCR was performed using cDNA template created from 200 ng of total RNA using the iScript cDNA synthesis kit (Bio-Rad) and using the manufacturer's instructions for random primers. End-point PCR was performed on a MyCycler Thermal Cycler (Bio-Rad, Hercules, CA, USA) using the following program: 3 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at 55°C, and 45 s at 72°C; with a final extension of 10 min at 72°C. Products were separated on a 1% agarose gel, stained with SYBR safe DNA Gel Stain (Life Technologies, Carlsbad, CA, USA) and visualized using a Molecular Image Gel Doc (Bio-Rad, Hercules, CA, USA).

qRT-PCR reactions were performed on a Cycle iQ Multicolor Real-Time Detection System (Bio-Rad, Hercules, CA). qRT-PCR was then performed using gene specific primers, 10 μl Sso Fast Evagreen Supermix (Bio-Rad, Hercules, CA) and distilled water to a total volume of 20 μl. Amplification was performed under the following conditions: 3 min at 95°C, followed by 40 cycles at 95°C for 5 s, 55°C for 15 s, 72°C for 30 s, followed by a melt curve program, performed to confirm the presence of one optimal dissociation temperature for the resulting amplicons. Standard curves were created using a five-fold cDNA dilution series with each primer set. All efficiencies were between 90 and 105%. Each gene was assayed in three technical triplicates for each of three biological replicates, and data was analyzed using the ΔΔCT relative quantification method with the 40S and L21 ribosomal subunits as control genes. Primers for both end-point and qRT-PCR are seen in **Table A1** in Appendix.

RESULTS

We analyzed circulating adult hemocytes by high-throughput 454 transcriptomic and LC-MS/MS proteomic analyses. 454 high-throughput sequencing produced 650,686 reads totaling 279.9 Mb (**Table 1**; Dataset S1 in Supplementary Material) while LC-MS/MS analyses of circulating hemocytes putatively identified 702 unique proteins (**Table 2**; Datasets S2 and S3 in Supplementary Material). Analysis of gene ontologies (GO) from these datasets revealed that both the transcriptome and proteome yielded similar types of genes/proteins involved with biological processes, molecular function, and cell components (**Figure 2**).

Despite a large number of reads associated with ribosomal genes, assembly of the transcriptome resulted in 5,586 contigs having an N50 of 521 bp, of which 1,995 were annotated (**Table 1**). Many of these were predicted to be involved with cytoskeletal structure (**Figure 2**; Dataset S1 in Supplementary Material). Among innate immunity-related genes, several members, and/or putative regulators of the NF-κB signaling pathway were identified including TRAF4, TRAF6, REL/NF-κB, inhibitor of NF-κB, NF-κB activating protein, both calcineurin subunits, a regulator of calcineurin, and MyD88 (**Table 3**). Several PRRs were identified including galectin and a previously described peptidoglycan recognition protein (EsPGRP4) (**Table 3**). An undescribed PGRP, most closely related to PGRP-SC2 from the rotifer *Brachionus manjavacas*, was also found (**Table 3**). Further analysis

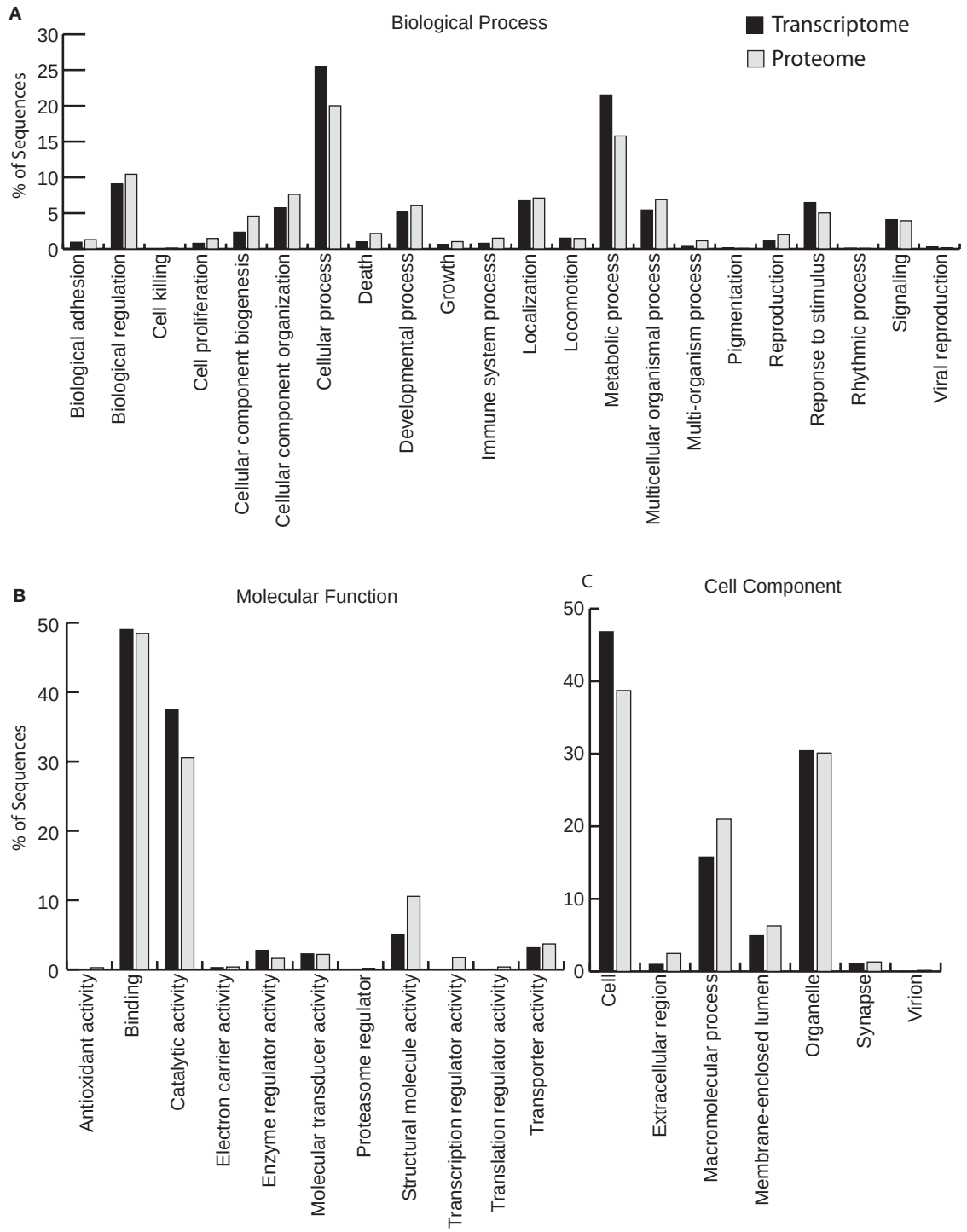


FIGURE 2 | Gene ontology (GO) analysis of the hemocyte transcriptome and proteome. Annotation by GO revealed similar proportions of GO terms in both the transcriptome and proteome (Datasets 1 and 2). Sequences were annotated using previously described BLAST parameters (see Materials and Methods), and level 2 GO terms in **(A)** biological process, **(B)** molecular function and **(C)** cell component were assigned by BLAST2GO. Black bars represent transcriptome sequences; gray bars represent proteome sequences.

revealed a complete open reading frame (ORF) consisting of a 759bp transcript encoding a predicted protein containing 253 amino acids (**Figure 3**). SignalP software predicted that this newly described PGRP (EsPGRP5) has a signal peptide and is therefore likely secreted (**Figure 3**). In addition, alignment with four other PGRPs from *E. scolopes* showed that EsPGRP5 has

Table 1 | Summary of hemocyte 454 transcriptome sequencing.

Number of reads	650, 686
Avg. read length	429.9 bases
Total bases	279.9 Mb
No. of rRNA reads (% of total)	567, 545 (87.2%)
Contigs assembled	5, 586
Contigs annotated by blastx	1, 995
N50	521 bases
Unassembled reads	364, 383

Table 2 | Summary of the hemocyte proteome as analyzed by LC-MS/MS.

Number of squid	3
Number of MS/MS spectra	107, 523
Number of spectra identified	8, 147
Number of proteins identified	630
Number of proteins putatively identified (1 peptide)	72
Peptide false discovery rate	0.0%
Protein false discovery rate	0.4%

conserved residues for amidase activity, suggesting that it is capable of degrading bacterial peptidoglycan and its derivatives (**Figure 3**). The EsPGRP5 gene was one of the most abundant annotated transcripts (95/1995, Dataset S1 in Supplementary Material) and it had significant coverage compared to other highly expressed housekeeping genes (**Table 4**). Peptides from all three PRRs found in the transcriptome were also detected in the proteome (**Figure 4**; **Table 5**; Datasets S1 and S2 in Supplementary Material). Other transcripts found included IK cytokine, CD63 antigen, two putative toxins, plancitoxin and cephalotoxin, and a matrix metalloproteinase that shares similarity with vitronectin, a member of the complement cascade (**Table 5**).

Analyses of the hemocyte proteome revealed a number of proteins involved with innate immune responses, and some of these were also found in the transcriptome (**Figure 4**; **Tables 3** and **5**, Dataset S3 in Supplementary Material). Comparing the MS/MS peptide data solely to the hemocyte transcriptome from this study revealed 596 genes/proteins in common and 145 proteins identified only from the proteome (**Figure 4**; Dataset S3 in Supplementary Material).

Several putative proteins with similarity to members of the complement cascade and other innate immunity functions were identified from the LC-MS/MS analyses, including the lynchpin

Table 3 | Innate immunity-related transcripts from 454 sequencing of *E. scolopes* hemocytes.

Putative ID	Accession no.	Organism	*e*-Value	Putative function	No. of unique peptides mapped[a]
NF-κB CASCADE					
TRAF6	AAY27978.1	*E. scolopes*	4e-48	NF-κB signaling	N/A
TRAF4	ABN04153.1	*B. belcheri*	9e-50	NF-κB signaling	N/A
REL/NF-κB	AAY27981.1	*E. scolopes*	9e-65	NF-κB signaling	N/A
Inhibitor of NF-κB	AAY27980.1	*E. scolopes*	4e-4	NF-κB signaling	N/A
NF-κB activating protein	CAX69433.1	*S. japonicum*	1e-36	NF-κB signaling	N/A
Calcineurin subunit A	ABO26624.1	*H. discus*	5e-54	Suppressor of NF-κB	N/A
Calcineurin subunit B	ACI96107.1	*P. fucata*	2e-56	Suppressor of NF-κB	N/A
Regulator of calcineurin	NP_001011215.1	*X. tropicalis*	4e-11	Regulator of calcineurin	N/A
MyD88	BAC76897.1	*C. porcellus*	1e-20	Toll signaling protein	N/A
MAMP BINDING					
Galectin	ACS72241.1	*A. irradians*	3e-43	Glycan binding	3
PGRP-SC2 precursor (EsPGRP5)	ACV67267.1	*B. manjavacas*	5e-38	PGN binding/cleavage	4
PGRP4	AAY27976.1	*E. scolopes*	1e-52	PGN binding	4[b]
OTHER					
IK cytokine	ABV24915.1	*C. gigas*	7e-47	Anti-interferon cytokine	N/A
CD63 antigen	ABY87409.1	*H. diversicolor*	3e-25	Phagosome protein	3
Plancitoxin	Q75WF2.1	*A. planci*	1e-68	DNA-degrading protein	2
MMP/vitronectin	BAC66058.1	*S. kowalevskii*	7e-7	Immune homeostasis	1
Cephalotoxin	B2DCR8.1	*S. esculenta*	2e-118	Toxin	2

E. scolopes, Euprymna scolopes; B. belcheri, Branchiostoma belcheri; S. japonicum, Schistosoma japonicum; H. discus, Haliotis discus; P. fucata, Pinctada fucata; X. tropicalis, Xenopus tropicalis; C. porcellus, Cavia porcellus; A. irridians, Argopecten irradians; B. manjavacas, Brachionus manjavacas; C. gigas, Crassostrea gigas; H. diversicolor, Haliotis diversicolor; A. planci, Acanthaster planci; S. kowalevskii, Saccoglossus kowalevskii; S. esculenta, Sepia esculenta; PGN, peptidoglycan; MMP, Matrix metalloproteinase.

[a] *Mascot and Scaffold validated peptide matches using translated hemocyte ESTs (Dataset S2 in Supplementary Material).*

[b] *PGRP4 peptide matches identified in an additional search using light organ and hemocyte translated transcripts, which included published full-length E. scolopes protein sequences (Dataset S2 in Supplementary Material).*

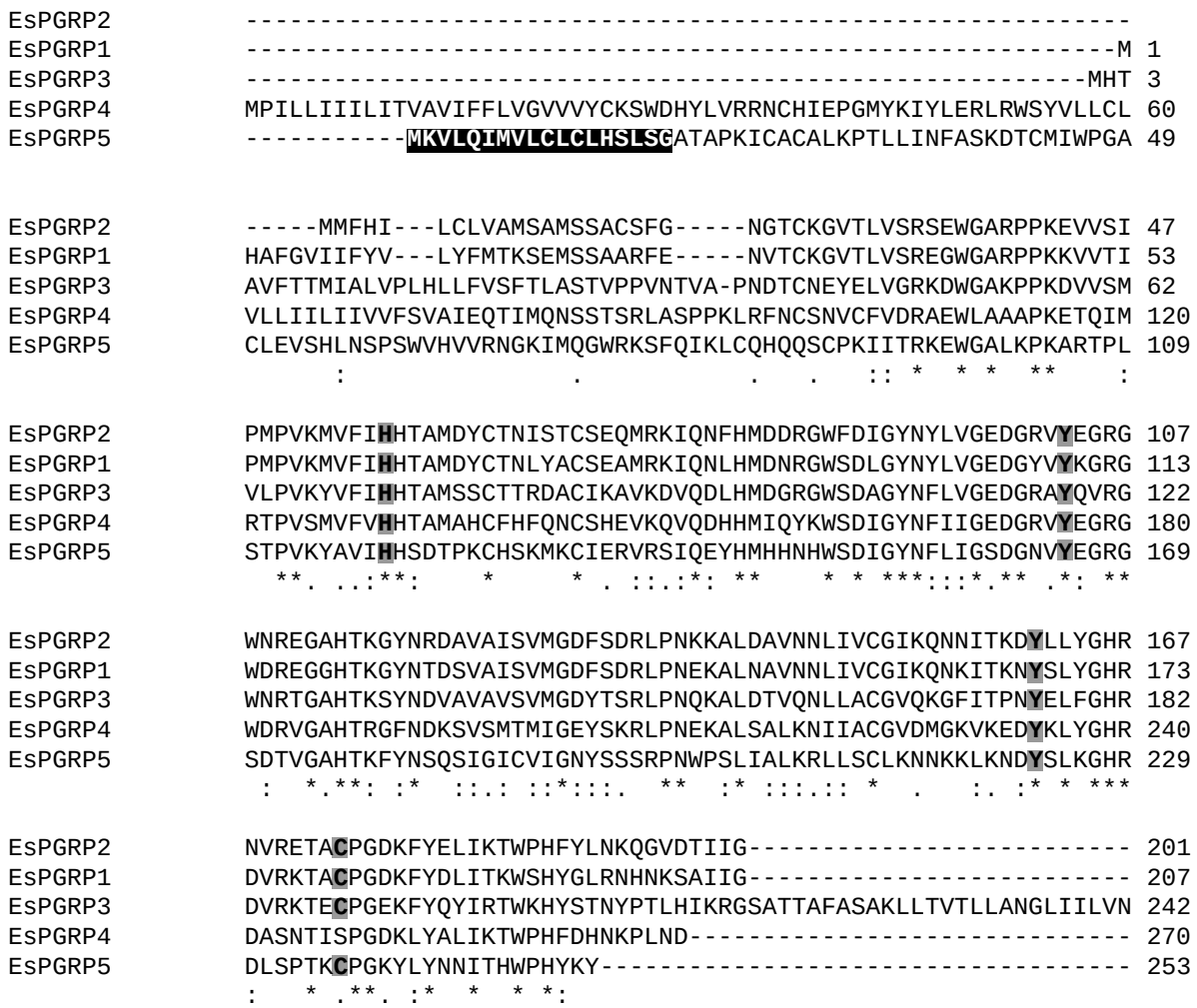


FIGURE 3 | Alignment of *E. scolopes* peptidoglycan recognition proteins. Alignment of EsPGRP5 with the previously described EsPGRPs 1–4, revealed a predicted 253 aa protein with a predicted signal peptide sequence (dark shade) as well as amidase activity (light shade). (∗) Conserved residues among all EsPGRPs; (.) substitutions with weak similarity <0.5 Gonnet PAM 250 matrix; (:) substitutions with strong similarity, >0.5 Gonnet PAM 250 matrix.

Table 4 | Coverage of EsPGRP5 compared to other highly expressed genes.

Transcript	Putative protein ID	*e*-Value	Avg. coverage
Contig1313	Beta actin	0	103.91
Contig5159	Profilin	3e-72	28.50
Contig1197	Translation elongation factor EF-1 alpha	0	17.66
Contig1384	Cytochrome b	2e-42	15.01
Contig1211	Heat shock protein 70	0	11.83
Contig1364	PGRP5	5e-38	11.48
Contig1222	Cytochrome c oxidase subunit I	3e-125	11.39
Contig4119	Ribosomal protein 40S S2	9e-91	2.36

Avg. coverage = sum of read lengths mapped to contig/contig length.

component EsC3, a cephalotoxin protein, two thioester proteins (TEPs), an alpha macroglobulin, and a matrix metalloproteinase similar to vitronectin (**Table 5**). A 398 aa residue protein was detected having six sushi repeat domains that are common to complement and cell adhesion proteins like CR1, CR2, and selectins. In addition there was a putative protein with similarity to the complement member C1qBP.

Both end-point RT- and qRT-PCR showed that members of the complement cascade including EsC3 and a putative matrix metalloproteinase/vitronectin along with cephalotoxin were differentially expressed in hemocytes isolated from hosts with fully colonized light organs compared to those hosts that had been cured (naïve; **Figure 5**). EsC3 transcripts were detected in naïve hemocytes but not in hemocytes from symbiotic hosts. The matrix metalloproteinase/vitronectin and cephalotoxin were both down-regulated in naïve hemocytes (−5.6 and −7.1 fold

respectively). EsPGRP5 decreased in expression in naïve hemocytes compared to those from symbiotic hosts (−4.7 fold). EsPGRP5 also had the lowest CT value of all the genes analyzed, including the 40S and L21 ribosomal controls, adding further evidence that this gene is abundantly expressed in hemocytes (**Table A2** in Appendix). In addition, nitric oxide synthase (NOS), which has previously been shown to be differentially expressed in symbiotic and aposymbiotic (uncolonized) juvenile light organs (Davidson et al., 2004), was up-regulated in hemocytes from cured hosts (+16.8 fold).

DISCUSSION

This study represents the first transcriptomic and proteomic analyses of the circulating hemocytes from *E. scolopes*. As might be expected, in addition to factors involved with general biological processes (**Figure 2**), hemocytes express a suite of innate immunity genes under a normal symbiotic state (**Figure 6**). Our data have revealed a number of genes and proteins that have previously been shown to be important in the squid/*Vibrio* association. In addition we have described a fifth PGRP (EsPGRP5), which was among the most abundant transcripts in hemocytes. We also identified novel genes/proteins, for example a putative cephalotoxin and plancitoxin that have not been previously described in invertebrate hemocytes. Finally, we demonstrate that some of these genes are differentially expressed in hemocytes from symbiotic and cured hosts, suggesting that colonization influences gene expression of the cellular innate immune system of *E. scolopes*. The cellular localization and function of many of these genes and proteins have yet to be characterized in detail but we discuss below the putative function of some of these innate immunity effectors in the context of host hemocytes (**Figure 6**).

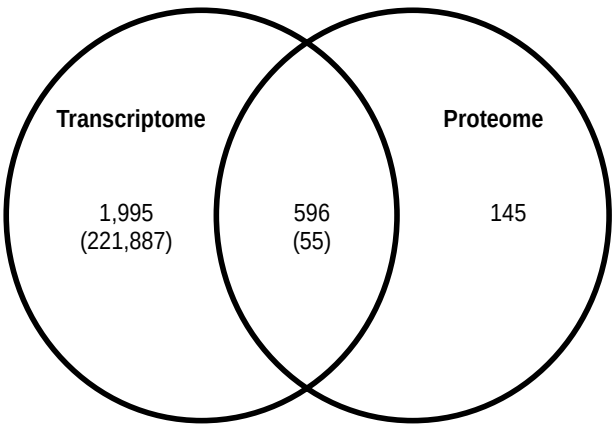


FIGURE 4 | Comparison of the hemocyte transcriptome and proteome. Protein sequences identified by hemocyte peptides were compared to the hemocyte transcriptome using tblastn. 596 sequences were found to be in common between the transcriptome and proteome while 145 protein sequences were identified solely using proteomics. Numbers in parentheses represent total annotated singletons (Transcriptome) and proteome matches to singletons (Shared).

Among the transcripts and identified peptides, many were predicted to be involved in cytoskeletal structure (**Figure 2**; Datasets S1–S3 in Supplementary Material). This is not surprising considering that these macrophage-like hemocytes can switch within minutes from a round spheroid to an activated cell with amoeboid characters. These activated cells are then capable of migrating through tissues and have been shown to enter the crypts of the light organ where they can then interact with *V. fischeri* (Nyholm and McFall-Ngai, 1998; Heath-Heckman and McFall-Ngai, 2011).

Among those genes involved with innate immunity, several putative PRRs were found including a galectin and two PGRPs (EsPGRP4 and EsPGRP5). Galectins are proteins that can bind carbohydrates and have been implicated as PRRs during the immune response (reviewed in Vasta, 2009). In other mollusks, galectins have been shown to bind microorganisms and are up-regulated in the presence of bacteria (Tasumi and Vasta, 2007; Song et al., 2011; Zhang et al., 2011). The role of galectin in this symbiosis has yet to be characterized, but efforts are underway to compare gene expression and protein levels in hemocytes of colonized and uncolonized hosts.

Table 5 | Putative complement pathway and other innate immunity-related proteins identified in *E. scolopes* hemocytes.

Protein	Mascot score[a]	Scaffold probability[b]	Unique peptides[c]	Total spectra	*E. scolopes* protein sequence annotation using the NCBI nr database: organism|gi	*e*-Value
Complement component C3	24	60.20%	1	1	*Euprymna scolopes*|192383355	0
Cephalotoxin	39	99.80%	1	1	*Sepia esculenta*|221271983	2e-118
Thioester protein (TEP) 1	218	99.80%	2	5	*Chlamys farreri*|144952812	2e-11
TEP2	40	98.90%	1	1	*Glossina morsitans morsitans*|85822201	1e-13
Alpha macroglobulin	83	99.80%	2	4	*Chlamys farreri*|40074252	4e-7
Matrix metalloproteinase/vitronectin	45	99.80%	2	2	*Saccoglossus kowalevskii*|291237328	7e-7
Sushi repeat domain-containing protein	403	100.00%	9	17	*Xenopus tropicalis*|301626998	5e-31
Hypothetical protein (similar to C1qBP)	25	Not validated	1	1	*Branchiostoma floridae*|260799758	6e-37

[a] *Protein scores were derived from peptide ion scores and were determined by Mascot. Peptide scores greater than 20 were considered significant ($p < 0.05$).*

[b] *Mascot identifications were validated by Scaffold, which utilizes Mascot scores to calculate protein identification probability.*

[c] *Unique peptides matched to protein sequences generated from light organ and hemocyte ESTs and published E. scolopes protein sequences present in the NCBI nr database (Dataset S2 in Supplementary Material).*

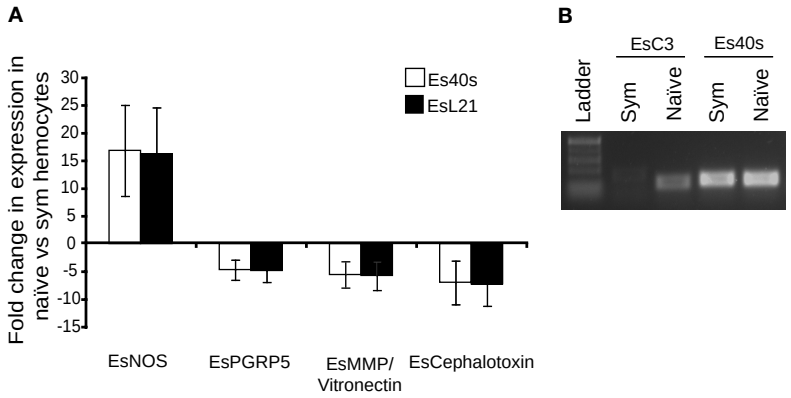


FIGURE 5 | Differential expression of putative innate immunity genes. qRT-PCR **(A)** and end-point PCR **(B)** of select immune genes in adult symbiotic and naïve hemocytes. For qRT-PCR transcription of indicated genes was normalized to the 40S and L21 ribosomal subunit genes and represented as fold difference of the naïve relative to the symbiotic state. Error bars represent SD of three triplicate samples. qRT-PCR and end-point PCR results are representative of three separate RNA extractions.

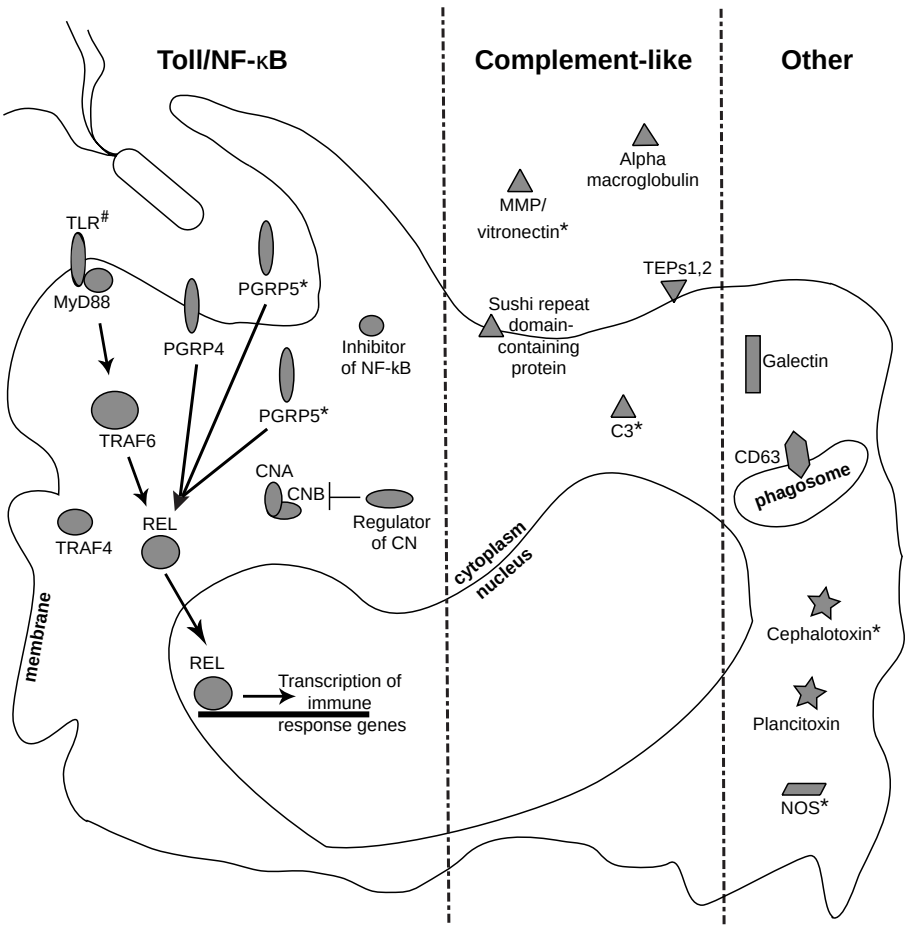


FIGURE 6 | Identification of transcripts and proteins associated with hemocytes. Representation of putative immune components present in circulating squid hemocytes. Interactions with symbiotic and/or non-symbiotic bacteria likely involve recognition by host PRRs, like PGRPs and TLR. Down stream signaling via the NF-κB pathway may modulate immune effector response. Putative members of a complement cascade as well as other immune-related genes were also expressed in circulating hemocytes. Colonization state influenced expression of a subset of these genes (∗; **Figure 4**),suggesting that *V. fischeri* can modulate the host's cellular innate immune response. (#), immune gene not identified in this study (Goodson et al., 2005; Rader et al., unpublished data). Localizations are inferred from protein domain predictions as well as similarity to other systems.

Whether an association is pathogenic or beneficial the immune system must first recognize the microorganism(s). These encounters are mediated by the initial recognition of MAMPs by PRRs (Koropatnick et al., 2004; Medzhitov, 2007; McFall-Ngai et al., 2012). PGRPs are one of a suite of PRRs that recognize MAMPs and are evolutionarily conserved (Royet et al., 2011). Five PGRPs have now been described from *E. scolopes* (Goodson et al., 2005; Troll et al., 2009, 2010; this study) and prior work has demonstrated that some of these are important during the onset of the symbiosis (Troll et al., 2009, 2010). Analyses of EsPGRP1 and EsPGRP2 protein expression in the juvenile light organ have shown that they have different cellular localizations within the crypt epithelia and nascent light organ tissues. *V. fischeri* can influence this expression with specific MAMPS. For example, tracheal cytotoxin (TCT), a derivative of peptidoglycan that along with lipopolysaccharide signals early development of the host light organ (Foster et al., 2000; Koropatnick et al., 2004), influences the nuclear loss of EsPGRP1 in host epithelial cells (Troll et al., 2009). EsPGRP2 which is secreted into the light organ crypts has TCT-degrading amidase activity and may ameliorate the effect of this toxin, especially during times when the symbiont numbers, and presumably TCT concentrations, are high (Troll et al., 2010). The ability to have a diverse repertoire of PGRPs may give the host greater ability to respond to peptidoglycan and TCT depending on the MAMP microenvironment that it encounters. The results of this study describe a fifth PGRP (**Table 3**; **Figure 3**) and suggest that EsPGRP5 is one of the most abundant transcripts expressed in host hemocytes (**Table 4**; **Figure 5**; Dataset S1 in Supplementary Material). Furthermore, qRT-PCR showed that this gene is down-regulated 4.7 fold in hemocytes from cured hosts suggesting that colonization of the light organ may directly influence PRRs in hemocytes. Currently, a polyclonal antibody to EsPGRP5 was generated and is being used to localize this protein in hemocytes at different stages of the association and under challenge by symbiotic and non-symbiotic bacteria.

The Toll/NF-κB pathway is an evolutionarily conserved signaling pathway often involved in regulating innate immune genes (Leulier and Lemaitre, 2008). In *E. scolopes*, homologs of this pathway have been identified (Goodson et al., 2005) and microarray data demonstrated different expression levels in colonized vs. uncolonized juvenile light organs (Chun et al., 2006). Our transcriptomic data revealed several members of this pathway, including MyD88, the universal adaptor protein for the TLRs as well as other IL-1 receptor proteins that plays an important and well-studied role in inflammation and host defense (Kenny and O'Neill, 2008). MyD88 is also involved with interactions of epithelial surfaces with microbial communities. For example in mice, production of certain antimicrobial peptides by paneth cells is dependent on MyD88, and is sufficient to limit bacterial translocation to the mesenteric lymph nodes without altering the number of luminal bacteria (reviewed in Wells et al., 2011). Two TNF receptor-associated factors TRAF4 and TRAF6 were also identified. TRAF4 is a signal transducer of the Toll/IL-1 family. In mouse macrophages, TRAF4 has been shown to negatively regulate the NF-κB signaling cascade in response to the peptidoglycan byproduct muramyl dipeptide (Marinis et al., 2011) while TRAF6 is involved with regulating ROS via the NF-κB pathway in mammalian macrophages (West et al., 2011). Both the A and B subunits of calcineurin were also detected in the transcriptome. Calcineurin is known for modulating immune responses in macrophages and this protein can inhibit both IL-12 and TNFα production. Furthermore, inhibiting calcineurin *in vitro* increases expression of effector genes, including NOS, and NF-κB as much as eight-fold (Conboy et al., 1999). Calcineurin may also be involved in responses to LPS since mice lacking a subunit of calcineurin show a tolerance to LPS, with much lower TNFα production than wild type animals (Jennings et al., 2009). Recent transcriptome profiling of other molluscan hemocytes from an oyster and mussel challenged with either pathogenic vibrios or lipopolysaccharide also revealed a number of expressed pattern recognition and signaling molecules related to the Toll/NF-κB pathway, thus highlighting the importance of this evolutionarily conserved pathway in this cell type (de Lorgeril et al., 2011; Philipp et al., 2012).

Recent analyses of the squid/*Vibrio* system have identified a number of putative members of the complement pathway (Castillo et al., 2009; Schleicher and Nyholm, 2011). C3 is the central protein to all three pathways of the complement system and is involved in opsonization and phagocytosis (Dunkelberger and Song, 2010). A C3 ortholog was recently discovered in *E. scolopes* and shown to be located at the apical surfaces of crypt epithelial cells (Castillo et al., 2009). In this study, C3 was found in the hemocyte proteome (**Table 5**). End-point RT-PCR and qRT-PCR showed C3 expression only in hemocytes from cured hosts. We do not yet know the turnover rate of the EsC3 protein in hemocytes but removing the symbionts led to an increase in gene expression of this main component of the complement pathway. Also identified were two TEPs. TEPs are similar to complement-like components and alpha macroglobulins, which have basic functions as proteases and protease inhibitors respectively (Bou Aoun et al., 2011). Among invertebrates, TEPs play an important role in the innate immune response as members of the complement-like system (Blandin and Levashina, 2004; Bou Aoun et al., 2011). A recent analysis of the light organ proteome found several TEPs (Schleicher and Nyholm, 2011). Also identified was an alpha macroglobulin belonging to a family of protease inhibitors and has been shown to be an important component of the innate immune response of arthropods (Sottrup-Jensen, 1989).

Among the other proteins identified was a matrix metalloproteinase similar to vitronectin. Vitronectin has been described as a serum glycoprotein involved in regulating innate immunity homeostasis (Schvartz et al., 1999). In humans, vitronectin inhibits the membrane attack complex of the complement pathway. Bacteria, including *Haemophilus influenzae* and *Neisseria meningitides* can interact with vitronectin and evade the complement response (Singh et al., 2010). Recent evidence also suggests that the parasite *Plasmodium falciparum* may use vitronectin-like domains to evade the immune system (Ludin et al., 2011). LC-MS/MS analyses identified a putative vitronectin in circulating hemocytes and results of qRT-PCR showed that MMP/vitronectin was down-regulated 5.6 fold in the absence of *V. fischeri* (**Figure 5**). The putative MMP/vitronectin in this study contains a hemopexin-like domain (data not shown) but whether it serves a role similar to vitronectin

in the squid host remains to be determined. Among other proteins identified was one containing six sushi repeat domains in a partial ORF (**Table 5**). Sushi domains are commonly found among complement receptors (e.g., CR1, CR2, C4b binding protein, and factor H) as well as other proteins involved with innate immunity and self/non-self recognition (Kirkitadze and Barlow, 2001; Nyholm et al., 2006). This protein has yet to be characterized in the squid/*Vibrio* association but offers an enticing future target for analysis. A putative peptide was also found that had similarity to C1qBP which is a component of the complement protein cascade but will have to be validated further.

Both the transcriptome and proteome identified a protein with high similarity to cephalotoxin (**Tables 3** and **5**; Ueda et al., 2008). In octopus and cuttlefish, cephalotoxins are normally expressed in the salivary glands and are involved with immobilizing prey. This is the first report of this protein in another cell type. A search of the UniProt database identified a thrombospondin domain similar to ones described in properdin, a proenzyme in the complement cascade (Kemper and Hourcade, 2008; Kemper et al., 2010). Whether cephalotoxin functions as a toxin, a protease, or in another capacity in the hemocytes of *E. scolopes* remains to be determined but qRT-PCR showed a 7.1 fold downturn in expression in naïve hemocytes, suggesting that colonization also influences expression of this gene (**Figure 5**). A second toxin was identified with similarity to plancitoxins found in the crown-of-thorns starfish *Acanthaster planci* (Shiomi et al., 2004). These proteins are predicted to be deoxyribonucleases and future studies should focus on characterizing any potential DNAse activity in squid hemocytes.

Reactive oxygen species are common effectors used by the innate immune system to combat pathogens. In symbiotic associations, ROS have also been shown to be important, often in contributing to microenvironments that may promote a microbial member (Weis et al., 1996; Davidson et al., 2004; Weis, 2008; Ryu et al., 2010). The enzyme nitric oxide synthase (NOS) is required for generating antimicrobial reactive nitrogen species (RNS) such as nitric oxide (NO). Both NOS and NO are active during colonization of the host and NO was detected in vesicles found within mucus secretions that interact with *V. fischeri* during initiation of the association (Davidson et al., 2004). Colonization leads to an attenuation of NOS and NO in the crypts suggesting that the RNS response may change in order to accommodate the symbiont (Davidson et al., 2004). Although not detected in the transcriptome or proteome, we decided to analyze NOS expression in symbiotic and naïve hemocytes because of its likely importance in the host immune response (**Figure 5**). Naïve hemocytes showed a 16.8 fold increase in expression suggesting that colonization leads to a suppression of NOS in hemocytes, similar to what's been described for juvenile light organs.

Another condition that may influence hemocyte gene expression is a changing symbiont population over a 24-h day/night cycle. During the evening when the host is active, there is a full complement of *V. fischeri*, but at dawn, the host expels 95% of these symbionts into the surrounding seawater (Boettcher et al., 1996; Nyholm and McFall-Ngai, 1998). We do not yet know how gene expression of host hemocytes might change during the day/night cycle and in the context of the microenvironment encountered in the crypt spaces of the light organ, but prior studies have reflected changes in host and symbiont transcription in adults light organ tissues over a 24-h period (Wier et al., 2010). Light organ gene expression is also altered by colonization in juvenile hosts (Chun et al., 2006). Our transcriptome was derived from circulating hemocytes from a variety of time points to understand the overall background gene expression in colonized adults. Growing evidence suggests that the immune response of hosts can fluctuate on a daily rhythm (Shirasu-Hiza et al., 2007; Binuramesh and Michael, 2011) and future studies should also consider how innate immunity effectors and pathways might vary over the day/night cycle in the squid/*Vibrio* association.

We still have much to understand concerning the role of host hemocytes in the squid/*Vibrio* symbiosis, but future research into this cell type will likely yield interesting results. Macrophage-like phagocytes serve at the interface between hosts and symbionts. Besides combating pathogens, invertebrate hemocytes have recently been found to play important roles in mediating bacterial symbioses as in the medicinal leech (Silver et al., 2007), linked to a primed immune response to specific bacteria in *Drosophila* (Pham et al., 2007) and promoting symbiont tolerance in the squid/*Vibrio* association (Nyholm et al., 2009). This is the first report to apply both transcriptomics and proteomics to analyze hemocytes in a beneficial symbiosis. One advantage of combining both techniques in future studies is that genes/proteins not identified by one method may be detected by the other. For example, both TEPs and alpha macroglobulin were found in the proteome but not in the transcriptome (**Figure 4**; Dataset S3 in Supplementary Material). Future studies will also use quantitative transcriptomic and proteomic methods to compare gene and protein expression between symbiotic and naïve hemocytes and those challenged with *V. fischeri* and non-symbiotic bacteria. This study lays the groundwork for exploring the molecular mechanisms of hemocyte/symbiont specificity in the squid/*Vibrio* symbiosis and adds further evidence that interactions between beneficial bacteria and the innate immune system are important in mediating these associations.

ACKNOWLEDGMENTS

We thank C. Obergfell and R. O'Neill and the Center for Applied Genetics and Technology at the University of Connecticut for assistance with 454 analyses. We thank T. Abbott, K. Stone, and the Mass spectrometry and Protein Chemistry Facility of the W.M. Keck Biotechnology Resource Laboratory at Yale University for assistance with proteomic data. We thank Kewalo Marine Laboratory and the Hawaii Institute of Marine Biology for assistance with collections. We thank P. Lapierre for assisting us with bioinformatic analyses. The UConn Bioinformatics Facility provided computing resources for BLAST searches and ClustalX alignments. This work was supported by the National Science Foundation IOS-0958006 and the University of Connecticut Research Foundation to Spencer V. Nyholm.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Molecular_Innate_Immunity/10.3389/fimmu.2012.00091/abstract

REFERENCES

Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005). Host-bacterial mutualism in the human intestine. *Science* 307, 1915–1920.

Bauer, E., Williams, B. A., Smidt, H., Verstegen, M. W., and Mosenthin, R. (2006). Influence of the gastrointestinal microbiota on development of the immune system in young animals. *Curr. Issues Intest. Microbiol.* 7, 35–51.

Binuramesh, C., and Michael, R. D. (2011). Diel variations in the selected serum immune parameters in *Oreochromis mossambicus*. *Fish Shellfish Immunol.* 3, 824–829.

Blandin, S., and Levashina, E. A. (2004). Thioester-containing proteins and insect immunity. *Mol. Immunol.* 40, 903–908.

Boettcher, K. J., Ruby, E. G., and McFall-Ngai, M. J. (1996). Bioluminescence in the symbiotic squid *Euprymna scolopes* is controlled by a daily biological rhythm. *J. Comp. Physiol. A* 179, 65–73.

Bou Aoun, R., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D., and Matt, N. (2011). Analysis of thioester-containing proteins during the innate immune response of *Drosophila melanogaster*. *J. Innate Immun.* 3, 52–64.

Castillo, M. G., Goodson, M. S., and McFall-Ngai, M. J. (2009). Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes*. *Dev. Comp. Immunol.* 33, 69–76.

Chun, C. K., Scheetz, T. E., Bonaldo, Mde. F., Brown, B., Clemens, A., Crookes-Goodson, W. J., Crouch, K., DeMartini, T., Eyestone, M., Goodson, M. S., Janssens, B., Kimbell, J. L., Koropatnick, T. A., Kucaba, T., Smith, C., Stewart, J. J., Tong, D., Troll, J. V., Webster, S., Winhall-Rice, J., Yap, C., Casavant, T. L., McFall-Ngai, M. J., and Soares, M. B. (2006). An annotated cDNA library of juvenile *Euprymna scolopes* with and without colonization by the symbiont *Vibrio fischeri*. *BMC Genomics* 7, 154. doi:10.1186/1471-2164-7-154

Collins, A. J., and Nyholm, S. V. (2010). Obtaining hemocytes from the Hawaiian bobtail squid *Euprymna scolopes* and observing their adherence to symbiotic and non-symbiotic bacteria. *J. Vis. Exp.* 36, 1714.

Conboy, I. M., Manoli, D., Mhaiskar, V., and Jones, P. (1999). Calcineurin and vacuolar-type H+-ATPase modulate macrophage effector functions. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6324–6329.

Conesa, A., Gotz, S., Garcia-Gomez, J. M., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676.

Craig, R., and Beavis, R. C. (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467.

Davidson, S. K., Koropatnick, T. A., Kossmehl, R., Sycuro, L., and McFall-Ngai, M. J. (2004). NO means 'yes' in the squid-*Vibrio* symbiosis: nitric oxide (NO) during the initial stages of a beneficial association. *Cell. Microbiol.* 6, 1139–1151.

de Lorgeril, J., Zenagui, R., Rosa, R. D., Piquemal, D., and Bachere, E. (2011). Whole transcriptome profiling of successful immune response to *Vibrio* infections in the oyster *Crassostrea gigas* by digital gene expression analysis. *PLoS ONE* 6, e23142. doi:10.1371/journal.pone.0023142

Doino, J. A., and McFall-Ngai, M. J. (1995). A transient exposure to symbiosis-competent bacteria induces light organ morphogenesis in the host squid. *Biol. Bull.* 189, 347–355.

Duan, J., and Kasper, D. L. (2011). Regulation of T cells by gut commensal microbiota. *Curr. Opin. Rheumatol.* 23, 372–376.

Dunkelberger, J. R., and Song, W. (2010). Complement and its role in innate and adaptive immune responses. *Cell Res.* 20, 34–50.

Feldhaar, H., and Gross, R. (2009). Insects as hosts for mutualistic bacteria. *Int. J. Med. Microbiol.* 299, 1–8.

Foster, J. S., Apicella, M. A., and McFall-Ngai, M. J. (2000). *Vibrio fischeri* lipopolysaccharide induces developmental apoptosis, but not complete morphogenesis, of the *Euprymna scolopes* symbiotic light organ. *Dev. Biol.* 226, 242–254.

Goodson, M. S., Kojadinovic, M., Troll, J. V., Scheetz, T. E., Casavant, T. L., Soares, M. B., and McFall-Ngai, M. J. (2005). Identifying components of the NF-κB pathway in the beneficial *Euprymna scolopes*-*Vibrio fischeri* light organ symbiosis. *Appl. Environ. Microbiol.* 71, 6934–6946.

Gross, R., Vavre, F., Heddi, A., Hurst, G. D., Zchori-Fein, E., and Bourtzis, K. (2009). Immunity and symbiosis. *Mol. Microbiol.* 73, 751–759.

Heath-Heckman, E. A., and McFall-Ngai, M. J. (2011). The occurrence of chitin in the hemocytes of invertebrates. *Zoology* 114, 191–198.

Hirosawa, M., Hoshida, M., Ishikawa, M., and Toya, T. (1993). MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming. *Comput. Appl. Biosci.* 9, 161–167.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G., and Gibson, T. J. (1998). Multiple sequence alignment with ClustalX. *Trends Biochem. Sci.* 23, 403–405.

Jennings, C., Kusler, B., and Jones, P. (2009). Calcineurin inactivation leads to decreased responsiveness to LPS in macrophages and dendritic cells and protects against LPS-induced toxicity in vivo. *Innate Immun.* 15, 109–200.

Kemper, C., Atkinson, J. P., and Hourcade, D. E. (2010). Properdin: emerging roles of a pattern-recognition molecule. *Annu. Rev. Immunol.* 28, 131–155.

Kemper, C., and Hourcade, D. E. (2008). Properdin: new roles in pattern recognition and target clearance. *Mol. Immunol.* 45, 4048–4056.

Kenny, E. F., and O'Neill, L. A. (2008). Signalling adaptors used by Toll-like receptors: an update. *Cytokine* 43, 342–349.

Kirkitadze, M. D., and Barlow, P. N. (2001). Structure and flexibility of multiple domain proteins that regulate complement activation. *Immunol. Rev.* 180, 146–161.

Koropatnick, T. A., Engle, J. T., Apicella, M. A., Stabb, E. V., Goldman, W. E., and McFall-Ngai, M. J. (2004). Microbial factor-mediated development in a host-bacterial mutualism. *Science* 306, 1186–1188.

Koropatnick, T. A., Kimbell, J. R., and McFall-Ngai, M. J. (2007). Responses of host hemocytes during the initiation of the squid-*Vibrio* symbiosis. *Biol. Bull.* 212, 29–39.

Krasity, B. C, Troll, J. V., Weiss, J. P., and McFall-Ngai, M. J. (2011). LBP/BPI proteins and their relatives: conservation over evolution and roles in mutualism. *Biochem. Soc. Trans.* 39, 1039–1044.

Lamarcq, L. H., and McFall-Ngai, M. J. (1998). Induction of a gradual, reversible morphogenesis of its host's epithelial brush border by *Vibrio fischeri*. *Infect. Immun.* 66, 777–785.

Lee, Y. K., and Mazmanian, S. K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* 330, 1768–1773.

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

Ludin, P., Nilsson, D., and Mäser, P. (2011). Genome-wide identification of molecular mimicry candidates in parasites. *PLoS ONE* 6, e17546. doi:10.1371/journal.pone.0017546

Marinis, J. M., Homer, C. R., McDonald, C., and Abbot, D. W. (2011). A novel motif in the Crohn's disease susceptibility protein, NOD2, Allows TRAF4 to down-regulate innate immune responses. *J. Biol. Chem.* 286, 1938–1950.

McFall-Ngai, M. J. (2002). Unseen forces: the influence of bacteria on animal development. *Dev. Biol.* 242, 1–14.

McFall-Ngai, M. J. (2007). Adaptive immunity: care for community. *Nature* 445, 153.

McFall-Ngai, M. J., Heath-Heckman, E. A., Gillette, A. A., Peyer, S. M., and Harvie, E. A. (2012). The secret languages of coevolved symbioses: insights from the *Euprymna scolopes*-*Vibrio fischeri* symbiosis. *Semin. Immunol.* 24, 3–8.

McFall-Ngai, M. J., Nyholm, S. V., and Castillo, M. G. (2010). The role of the immune system in the initiation and persistence of the *Euprymna scolopes* – *Vibrio fischeri* symbiosis. *Semin. Immunol.* 22, 48–53.

Medzhitov, R. (2007). Recognition of microorganisms and activation of the immune response. *Nature* 449, 819–826.

Moran, N. A., McCutcheon, J. P., and Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190.

Nyholm, S. V., Deplancke, B., Gaskins, H. R., Apicella, M. A., and McFall-Ngai, M. J. (2002). Roles of *Vibrio fischeri* and nonsymbiotic bacteria in the dynamics of mucus secretion during symbiont colonization of the *Euprymna scolopes* light organ. *Appl. Environ. Microbiol.* 68, 5113–5122.

Nyholm, S. V., and McFall-Ngai, M. J. (1998). Sampling the light-organ microenvironment of *Euprymna scolopes*: description of a population of host cells in association with the bacterial symbiont *Vibrio fischeri*. *Biol. Bull.* 195, 89–97.

Nyholm, S. V., and McFall-Ngai, M. J. (2004). The winnowing: establishing the squid-Vibrio symbiosis. *Nat. Rev. Microbiol.* 2, 632–642.

Nyholm, S. V., Passegue, E., Ludington, W. B., Voskoboynik, A., Mitchel, K., Weissman, I. L., and DeTomaso, A. W. (2006). Fester, a

candidate allorecognition receptor from a primitive chordate. *Immunity* 25, 163–173.

Nyholm, S. V., Stewart, J. J., Ruby, E. G., and McFall-Ngai, M. J. (2009). Recognition between symbiotic *Vibrio fischeri* and the haemocytes of *Euprymna scolopes*. *Environ. Microbiol.* 11, 483–493.

Olszak, T., An, D., Zeissig, S., Vera, M. P., Richter, J., Fanke, A., Glickman, J. N., Siebert, R., Baron, R. M., Kasper, D. L., and Blumberg, R. S. (2012). Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* (in press).

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551–3567.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786.

Pham, L. N., Dionne, M. S., Shirasu-Hiza, M., and Schneider, D. S. (2007). A specific primed immune response in *Drosophila* is dependent on phagocytes. *PLoS Pathog.* 3, e26. doi:10.1371/journal.ppat.0030026

Philipp, E. E. R., Kraemer, L., Melzner, F., Poustka, A. J., Thieme, S., Findeisen, U., Schreiber, S., and Rosentiel, P. (2012). Massive parallel RNA sequencing identifies a complex immune gene repertoire in the lophotrochozoan *Mytilus edulis*. *PLoS ONE* 7, e33091. doi:10.1371/journal.pone.0033091

Royet, J., Gupta, D., and Dziarski, R. (2011). Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat. Rev. Immunol.* 11, 837–851.

Ryu, J. H., Ha, E. M., and Lee, W. J. (2010). Innate immunity and gut-microbe mutualism in *Drosophila*. *Dev. Comp. Immunol.* 34, 369–376.

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.* 14, 255–274.

Schleicher, T. R., and Nyholm, S. V. (2011). Characterizing the host and symbiont proteomes in the association between the bobtail squid, *Euprymna scolopes*, and the bacterium, *Vibrio fischeri*. *PLoS ONE* 6, e25649. doi:10.1371/journal.pone.0025649

Schvartz, I., Seger, D., and Shaltiel, S. (1999). Vitronectin. *Int. J. Biochem. Cell Biol.* 31, 539–544.

Shiomi, K., Midorikawa, S., Ishida, M., Nagashima, Y., and Nagai, H. (2004). Plancitoxins, lethal factors from the crown-of-thorns starfish *Acanthaster planci*, are deoxyribonucleases II. *Toxicon* 44, 499–506.

Shirasu-Hiza, M. M., Dionne, M. S., Pham, L. N., Ayres, J. S., and Schneider, D. S. (2007) Interactions between circadian rhythm and immunity in *Drosophila melanogaster*. *Curr. Biol.* 17, R353–R355.

Silver, A. C., Kikuchi, Y., Fadl, A. A., Sha, J., Chopra, A. K., and Graf, J. (2007). Interaction between innate immune cells and a bacterial type III secretion system in a mutualistic and pathogenic associations. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9481–9486.

Singh, B., Su, Y., and Riesbeck, K. (2010). Vitronectin in bacterial pathogenesis: a host protein used in complement escape and cellular invasion. *Mol. Microbiol.* 78, 545–560.

Song, X., Zhang, H., Wang, L., Zhao, J., Mu, C., Song, L., Qiu, L., and Liu, X. (2011). A galectin with quadruple-domain from bay scallop *Argopecten irradians* is involved in the innate immune response. *Dev. Comp. Immunol.* 35, 592–602.

Sottrup-Jensen, L. (1989). α-Macroglobulins: structure, shape, and mechanism of proteinase complex formation. *J. Biol. Chem.* 264, 11539–11542.

Stewart, F. J., and Cavanaugh, C. M. (2006). Symbiosis of thioautotrophic bacteria with *Riftia pachyptila*. *Prog. Mol. Subcell. Biol.* 41, 197–225.

Stoll, S., Feldhaar, H., Fraunholz, M. J., and Gross, R. (2010). Bacteriocyte dynamics during development of a holometabolous insect, the carpenter ant *Camponotus floridanus*. *BMC Microbiol.* 10, 308. doi:10.1186/1471-2180-10-308

Tasumi, S., and Vasta, G. R. (2007). A galectin of unique domain organization from hemocytes of the eastern oyster (*Crassostrea virginica*) is a receptor for the protistan parasite *Perkinsus marinus*. *J. Immunol.* 179, 3086–3098.

Troll, J. V., Adin, D. M., Wier, A. M., Paquette, N., Silverman, N., Goldman, W. E., Stadermann, F. J., Stabb, E. V., and McFall-Ngai, M. J. (2009). Peptidoglycan induces loss of a nuclear peptidoglycan recognition protein during host tissue development in a beneficial animal-bacterial symbiosis. *Cell. Microbiol.* 11, 1114–1127.

Troll, J. V., Bent, E. H., Pacquette, N., Wier, A. M., Goldman, W. E., Silverman, N., and McFall-Ngai, M. J. (2010). Taming the symbiont for coexistence: a host PGRP neutralizes a bacterial symbiont toxin. *Environ. Microbiol.* 12, 2190–2203.

Ueda, A., Nagai, H., Ishida, M., Nagashima, Y., and Shiomi, K. (2008). Purification and molecular cloning of SE-cephalotoxin, a novel proteinaceous toxin from the posterior salivary gland of the cuttlefish *Sepia esculenta*. *Toxicon* 52, 574–581.

Vasta, G. R. (2009). Roles of galectins in infection. *Nat. Rev. Microbiol.* 7, 424–438.

Visick, K. L., and Ruby, E. G. (2006). *Vibrio fischeri* and its host: it takes two to tango. *Curr. Opin. Microbiol.* 9, 632–638.

Watanabe, H., and Tokuda, G. (2010). Cellulolytic systems in insects. *Annu. Rev. Entomol.* 55, 609–632.

Weis, V. M. (2008). Cellular mechanisms of cnidarian bleaching: stress causes the collapse of symbiosis. *J. Exp. Biol.* 211(Pt 19), 3059–3066.

Weis, V. M., Small, A. L., and McFall-Ngai, M. J. (1996). A peroxidase related to the mammalian antimicrobial protein myeloperoxidase in the *Euprymna-Vibrio* mutualism. *Proc. Natl. Acad. Sci. U.S.A.* 93, 13683–13688.

Weiss, B. L., Wang, J., and Aksoy, S. (2011). Tsetse immune system maturation requires the presence of obligate symbionts in larvae. *PLoS Biol.* 9, e1000619. doi:10.1371/journal.pbio.1000619

Wells, J. M., Rossi, O., Meijerink, M., and van Baarlen, P. (2011). Epithelial crosstalk at the microbiota-mucosal interface. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4607–4614.

West, A. P., Brodsky, I. E., Rahner, C., Woo, D. K., Bromage, H., Tempst, P., Walsh, M. C., Choi, Y., Shadel, G. S., and Ghosh, S. (2011). TLR signaling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* 472, 476–480.

Wier, A. M., Nyholm, S. V., Mandel, M. J., Massengo-Tiassé, R. P., Schaefer, A. L., Koroleva, I., Splinter-Bondurant, S., Brown, B., Manzella, L., Snir, E., Almabrazi, H., Scheetz, T. E., Bonaldo, Mde. F., Casavant, T. L., Soares, M. B., Cronan, J. E., Reed, J. L., Ruby, E. G., and McFall-Ngai, M. J. (2010). Transcriptional patterns in both host and bacterium underlie a daily rhythmof anatomical and metabolic change in a beneficial symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2259–2264.

Zdobnov, E. M., and Apweiler, R. (2001). InterProScan- an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17, 847–848.

Zhang, D., Jiang, S., Hu, Y., Cui, S., Guo, H., Wu, K., Li, Y., and Su, T. (2011). A multidomain galectin involved in innate immune response of pearl oyster *Pinctada fucata*. *Dev. Comp. Immunol.* 35, 1–6.

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.

Received: 19 January 2012; accepted: 08 April 2012; published online: 10 May 2012.

Citation: Collins AJ, Schleicher TR, Rader BA and Nyholm SV (2012) Understanding the role of host hemocytes in a squid/Vibrio symbiosis using transcriptomics and proteomics. Front. Immun. ***3****:91. doi: 10.3389/fimmu.2012.00091*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

APPENDIX

Table A1 | Primer sequences used in end-point and qRT-PCR.

Primer name	Primer sequence 5′–3′
Es40sF	AATCTCGGCGTCCTTGAGAA
Es40sR	GCATCAATTGCACGACGAGT
EsL21F	GCCTTGGCTTGAGCCTTCAACTTT
EsL21R	GGTGATCGTCAACAAACGCGTGAA
EsC3F	TGCTGTTCCGTTCTGTGAGCACTA
EsC3R	GCAACACACTCTCTTTGAGCGCAT
EsNosF	TGTTCGAGAGTGTTGCGGACATCA
EsNosR	TCACTTGAATGTTGCCGTCTCCAG
EsPGRP5F	TTGGAGCCCACACCAAGTTCTACA
EsPGRP5R	CAAGACAGCAGGCGTTTCAATGCT
EsCephalotoxinF1	TACCACGCGAATTGACACAT
EsCephalotoxinR1	TACTGGAACGCACAAGTTGC
EsVitronectinF1	TGCGCCAAAATACATCAGAG
EsVitronectinR1	TGGGAATATCCCACGGTAAA

Table A2 | CT values for qRT-PCR of innate immunity genes in hemocytes from symbiotic and cured hosts.

Gene	Cured/Naïve ct* value (average ± SE)	Sym ct* value (average ± SE)
EsC3	34.04 ± 0.76	NA[#]
EsNOS	33.09 ± 0.74	36.59 ± 0.70
EsPGRP5	24.08 ± 0.51	21.26 ± 0.61
EsVitronectin/matrix metalloproteinase	27.65 ± 0.54	24.60 ± 0.40
EsCephalotoxin	31.97 ± 1.03	28.58 ± 1.04
40 s control	26.31 ± 0.59	25.75 ± 0.67
L21 control	28.89 ± 0.61	28.39 ± 0.75

**Cycle threshold (ct) is the cycle number at which PCR amplification as measured by fluorescence is exponential and therefore above background.*
#EsC3 is not detectably expressed in the symbiotic hemocyte samples by end-point or qRT-PCR.

frontiers in
IMMUNOLOGY

MINI REVIEW ARTICLE
published: 05 March 2012
doi: 10.3389/fimmu.2012.00033

The commensal microbiota drives immune homeostasis

Marie-Claire Arrieta[1] and Barton Brett Finlay[1,2,3]*

[1] *Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada*
[2] *Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC, Canada*
[3] *Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC, Canada*

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Michael Bailey, The Ohio State University, USA
John Rawls, University of North Carolina at Chapel Hill, USA

***Correspondence:**
Barton Brett Finlay, Michael Smith Laboratories, University of British Columbia, #301 – 2185 East Mall, Vancouver, BC, Canada V6T 1Z4.
e-mail: bfinlay@interchange.ubc.ca

For millions of years, microbes have coexisted with eukaryotic cells at the mucosal surfaces of vertebrates in a complex, yet usually harmonious symbiosis. An ever-expanding number of reports describe how eliminating or shifting the intestinal microbiota has profound effects on the development and functionality of the mucosal and systemic immune systems. Here, we examine some of the mechanisms by which bacterial signals affect immune homeostasis. Focusing on the strategies that microbes use to keep our immune system healthy, as opposed to trying to correct the immune imbalances caused by dysbiosis, may prove to be a more astute and efficient way of treating immune-mediated disease.

Keywords: intestinal microbiota, immune homeostasis

INTRODUCTION

Higher organisms benefit from their microbiota, which offers nutritional and physiological advantages to the host in exchange for a nutrient-rich habitat. Microbes (bacteria, archaea, viruses, fungi, protozoans, etc.) reside on all mucosal surfaces of the body, with the gut harboring by far the most numerous population. Separating the microbial cells from the host is the intestinal barrier, which serves as an interface for many complex interactions. Two major components of the intestinal barrier are the intestinal epithelium and the gut associated lymphoid tissue (GALT). This tissue has the challenging dual task of selectively absorbing nutrients from the intestinal lumen, while preventing microbial entry and infection. The majority of microbes that reside in the intestine are commensals and do not pose a threat to the host under normal conditions. Thus, the host's immune system must learn to tolerate this enormous commensal antigenic load. One of the strategies that the host utilizes to avoid an inflammatory response against the microbiota is to remain "ignorant" to a portion of the microbial antigens, by using the intestinal barrier, including the mucous layer and immunoglobulin (Ig) A, to maintain these antigens at a distance to the mucosal surface (Hansson and Johansson, 2010; Hooper and Macpherson, 2010). However, it is now clear that there is also constant communication between microbial cells or signals and host cell receptors, and that many of these interactions lead to the development of immune tolerance toward bacterial antigens (Hooper and Macpherson, 2010; Geuking et al., 2011).

Traditionally, it has been accepted that the evolution of the mammalian immune system has been driven by selective pressures imposed by pathogenic bacteria (Matsunaga and Rahman, 1998). For example, the emergence of jaws in cartilaginous fish matches the appearance of antigen receptor specificity and immunological memory in B and T cells, as well as the major histocompatibility complexes (MHC) I and II, all characteristic of the adaptive immune system (Agrawal et al., 1998). Immunologists have speculated that the jaw gave organisms the ability to ingest harder food sources that could damage the gut mucosa, likely increasing their risk for infections. The nutritional advantages of having a jaw allowed organisms to augment their body size and life expectancy, increasing their risk for repeated microbial attack and the necessity to develop immunological memory (Matsunaga and Rahman, 1998). While this may be true for the emergence of some features of the adaptive immune system, this view may not apply to many other components of the immune system. First, it does not take into account that the vast majority of interactions that occur between the microbiota and the mammalian immune system involve non-pathogenic bacteria. More importantly, commensals, as opposed to pathogens, do not limit the host's lifespan, reproducibility, and consequent vertical gene transfer, and actually probably enhance it. Thus, it is more likely that commensal microorganisms pose a stronger evolutionary pressure on the host's immune system than pathogens, a view that we share with others (McFall-Ngai, 2007; Lee and Mazmanian, 2010).

The goal of this minireview is to describe some of the diverse mechanisms of commensal–host communication in the mammalian intestine and to explore the effect that these have in immune homeostasis within the intestine and systemically.

IMMUNE CONSEQUENCES OF AN AXENIC LIFE

The generation of germ-free mice has given us great insight into the immune effects of a microbe-free life, the development of the immune system during early microbial colonization, and the effects of specific bacterial groups and species on the local and systemic immune system.

The immune system is poorly developed in the absence of microbes. Structurally, the aggregated lymphoid structures of the GALT: mesenteric lymph nodes (MLNs), isolated lymphoid follicles, cryptopatches, and Peyer's patches are all underdeveloped in germ-free mice (Mazmanian et al., 2005; Bouskra et al., 2008). These mice exhibit decreased amounts of most cytokines, CD4+ T helper cells, FoxP3+ Tregs, B cells, Th17 cells, IgA, antimicrobial

peptides, MHC class II, etc. (Moreau et al., 1978; Dobber et al., 1992; Strauch et al., 2005; Ivanov et al., 2008; Niess et al., 2008; Gaboriau-Routhiau et al., 2009; Round and Mazmanian, 2009), and inoculating them with a single species of bacteria can revert many of these defects (Umesaki et al., 1995; Mazmanian et al., 2005). Gut bacteria also have critical effects on the development of intestinal epithelial cells (IECs). IECs are the first cells to encounter the microbiota and are able to integrate most microbial signals and dictate the tone of the immune response of the underlying lamina propria immune cells (Rescigno, 2011). IECs express microbial pattern recognition receptors (PRR) that bind to microbial signals and upregulate the expression of mucus, cytokines, and other immune components (Rescigno, 2011). In the absence of bacteria, not only do these cells have a low expression of PRRs (Willing and Van Kessel, 2007), they also fail to develop their typical microvilli pattern and show a slow cell turnover rate (Abrams et al., 1963).

Not surprisingly, the functionality of the immune system in axenic mice is very poor as well. Germ-free and even mono-colonized mice do not develop oral tolerance (Rask et al., 2005). Germ-free mice and antibiotic-treated mice are significantly more susceptible to infectious disease from several intestinal pathogens, including *Shigella flexneri* (Sprinz et al., 1961), *Citrobacter rodentium* (Wlodarska et al., 2011), *Listeria monocytogenes* (Zachar and Savage, 1979), and *Salmonella typhimurium* (Nardi et al., 1989; Sekirov et al., 2008; Ferreira et al., 2011). The microbiota can prevent or ameliorate infection by direct microbial antagonism or by indirectly promoting appropriate host immune defenses. Direct microbial effects, also defined as "colonization resistance" (van der Waaij et al., 1971), include host receptor and nutrient competition, and secretion of antimicrobial substances (Sekirov and Finlay, 2009). The importance of these effects were demonstrated in a study where transfer of a normal microflora, in the absence of T and B cells, cleared infection with *S. typhimurium* (Endt et al., 2010). This study highlights the importance of the competitive pressures exerted by the microbiota as well as the role of the innate immune response in enteric infections. In another study, a fecal transplant from a mouse strain resistant to *C. rodentium* was sufficient to delay pathogen colonization and reduce mortality in a mouse strain lethally susceptible to infection. In addition, antibiotic therapy increased disease severity in resistance mice, suggesting that the difference in infection susceptibility between the two mice strains can be explained, at least in part, by differences in microbiota composition (Willing et al., 2011).

A normal microbiota has also been shown to help clear pathogens by stimulating several host immune effectors. In one study, oral treatment with peptidoglycan from the intestinal microbiota was effective in preventing pneumococcal sepsis in mice. Levels of this bacterial membrane component in serum correlated with systemic neutrophil killing capacity and activation of the innate immune receptor NOD1, indicating that signals from the microbiota translocate across the intestinal barrier and interact with systemic immune cells (Clarke et al., 2010). Another example of how microbial-derived signals regulate the systemic immune response comes from a study that showed that administration of the TLR5 ligand, flagellin, to antibiotic-treated mice inhibited colonization with vancomycin resistant *Enterococcus* (VRE). Intraperitoneal injection of flagellin upregulated the expression of the bactericidal lectin REGIIIγ in IECs and Paneth cells in the small intestine via activation of TLR5 in hematopoietic cells and secretion of IL-22 (Kinnebrew et al., 2010). Flagellin stimulation has also been shown to induce differentiation of non-specific IgA+ plasma cells via activation of TLR5 in intestinal dendritic cells (Uematsu et al., 2008). Colonization with the microorganism *Bacteroides fragilis* is sufficient to prevent murine experimental colitis induced by *Helicobacter hepaticus*. *B. fragilis* is coated with polysaccharide A (PSA) and this capsular component was shown to suppress IL-17, increase IL-10 production and ameliorate disease in this model (Mazmanian et al., 2008).

THE COMMENSAL MICROBIOTA BALANCES PRO AND ANTI-INFLAMMATORY IMMUNE MECHANISMS

The Th17 immune response provides a good example of how the microbiota mediates the balance between immune homeostasis and uncontrolled inflammation. Overproduction of Th17 cytokines IL-17 and IL-23 is associated with colitis (Atarashi et al., 2008; Buonocore et al., 2010) and autoimmunity (Wen et al., 2008; Lee et al., 2010; Wu et al., 2010); however, Th17 cells are critical in controlling extracellular bacterial and fungal infections (Smith et al., 2007; Wu et al., 2010). Induction of Th17 immunity is crucial to prevent infection with *C. rodentium*, a murine pathogen that induces similar inflammation to enteropathogenic *E. coli* in humans (Wu et al., 2010). Thus, the right intensity of Th17 response must be achieved in order to prevent microbial attack and avoid uncontrolled inflammation. Segmented filamentous bacteria (SFB) are an unclassified species that resides in the murine ileum and has been shown to be a potent stimulator of the Th17 response (Gaboriau-Routhiau et al., 2009; Wu et al., 2010). Colonization with SFB induced the upregulation of cytokines, antimicrobial peptides, and serum amyloid A (SAA), an acute phase protein secreted during inflammation. SAA is believed to promote the Th17 differentiation of CD4+ T cells (Ivanov et al., 2009; Ather et al., 2011). Another mechanism that leads to Th17 differentiation is the production of commensal-derived adenosine 5′ triphosphate (ATP). ATP in the lamina propria stimulates dendritic cells to produce IL-6 and TGF-β, two cytokines necessary for Th17 cell expansion (Atarashi et al., 2008). Not surprisingly, microbial signals are also involved in limiting the Th17 immune response. IECs produce IL-25, which inhibits IL-23 production and prevents Th17 differentiation in response to the commensal flora (Zaph et al., 2008). In addition, oral administration of purified PSA from *B. fragilis* downregulated the Th17 response and induced propagation of Tregs in mice that develop autoimmune encephalomyelitis. This treatment was sufficient to prevent disease in this animal model (Ochoa-Reparaz et al., 2010). Collectively, these findings show how the microbiota produces inflammatory and tolerogenic signals in order to achieve immune homeostasis, a state that is favorable for both the microbiota and the host.

It is still unknown how certain microbes induce a pro-inflammatory response and others a tolerogenic one, but due to the central role that Treg cells have in regulating inflammatory responses (**Figure 1**), one could speculate that microbes manipulate the Treg cell response as a strategy to regulate immune responses towards them. But how do Treg cells perform such task? These cells are selected in the thymus for their ability to suppress

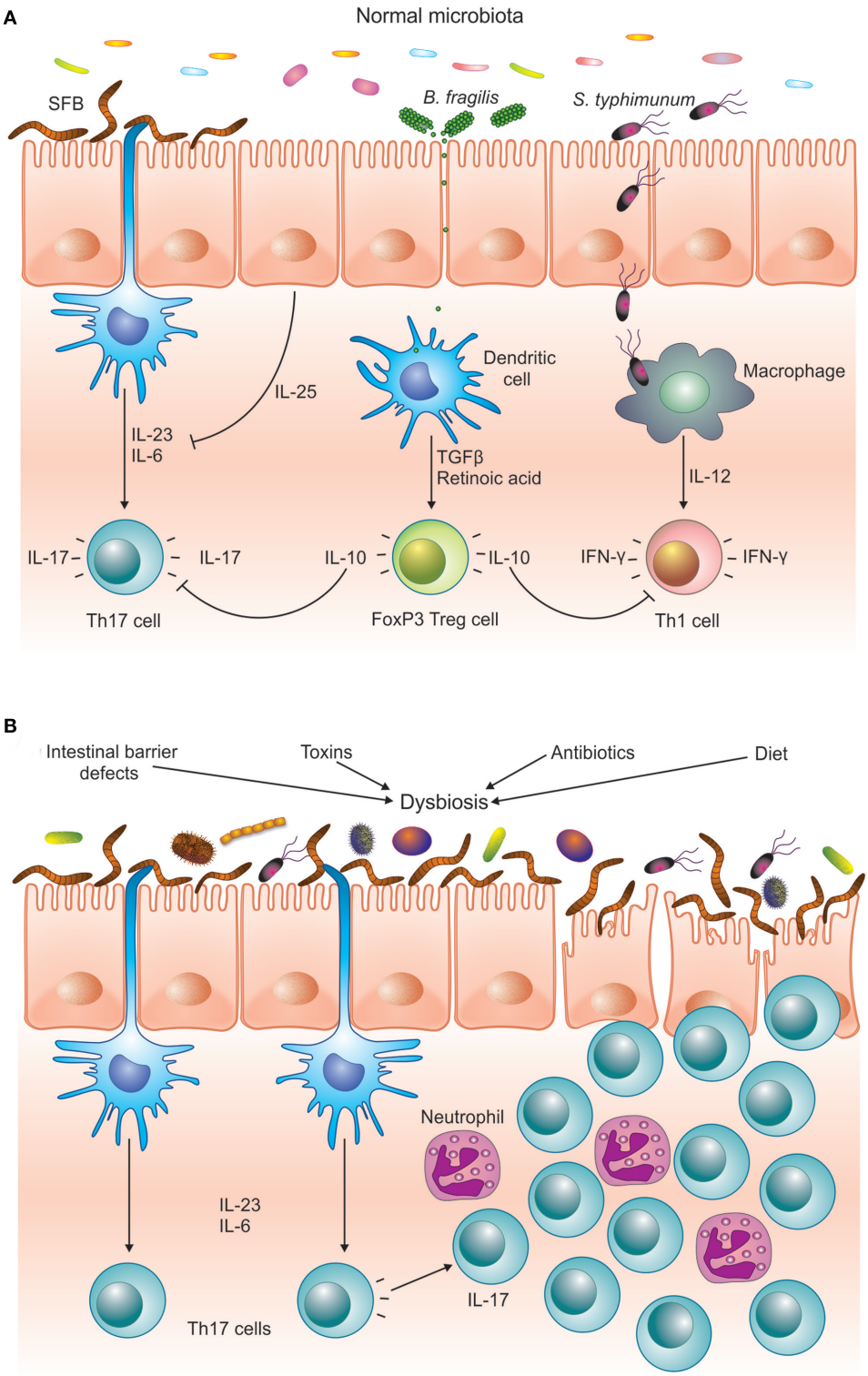


FIGURE 1 | The microbiota helps to maintain immune homeostasis by stimulating different arms of the T cell response. **(A)** SFB is a potent inducer of Th17 cells, whereas pathogens like *S. typhimurium* induce an effector Th1 response. PSA from *B. fragilis* stimulates the differentiation of FoxP3+ T cells, which downregulate the pro-inflammatory Th1 and Th17 responses. **(B)** Different environmental and host-derived factors are known to cause dysbiosis. Imbalances in the intestinal microbiota can favor Th17 pro-inflammatory T cells over regulatory FoxP3+ T cells, inducing inflammatory response that can damage the intestinal epithelium. In a genetically predisposed host this can lead to chronic inflammatory disease, such as IBD or autoimmunity.

T cells with a high affinity to self-MHC molecules, thus preventing autoimmune responses, but it is still unknown how these cells differentiate between pathogens or commensal pathogen signals. One explanation may be that the antigens from the commensal microbiota are used to train naïve T cells to differentiate into Treg cells. A recent study provides evidence that this may in fact occur. Analysis of several repertoires of T cell receptor (TCR) α chain revealed that the TCRs from FoxP3+ Treg cells within the colonic lamina propria are highly heterogeneous compared to Treg cells from effector/memory T cells in the lamina propria and from Treg cells from secondary lymphoid organs. Moreover, the study showed that the mouse microbiota was essential for the induction of this particular population of colonic Treg cells from naïve T cells, implying that there may be post-thymic mechanisms of immune cell education that occur peripherally via interactions with the commensal microbiota (Lathrop et al., 2011). This implies that the immune system may have evolved to rely on the microbiota to complete the training of a specific population of immune cells.

DISSECTING THE GUT MICROBIOTA

Although our knowledge of the immune effects caused by the microbiota is growing, much less is known about what constitutes a normal microbiota and which groups of microorganisms are beneficial or detrimental to immune homeostasis. Sequencing-based studies have substantially increased our knowledge of the ecology of the mammalian intestine, although it is safe to assume that we are likely at the early stages of comprehending this complex population.

Analyses of the human gut microbiota have yielded several important findings. First, although there is great inter-individual variability at the species level, there appears to be low variability between individuals at the phylum level (Eckburg et al., 2005; Turnbaugh et al., 2009). The human gut is colonized by mainly three phyla: Firmicutes, Bacteroidetes, and Actinobacteria. The most predominant bacterial families are Ruminococcaceae, Lachnospiraceae, Eubacteriaceae, and Bacteroidaceae, in what has been defined as the "core members of the microbiota" (Tap et al., 2009). Genetics and environmental factors, such as birth delivery method, diet, and antibiotic use have an important role in establishing the composition of the microbiota (Ley et al., 2006; Penders et al., 2006; Adlerberth, 2008; Benson et al., 2010; Willing et al., 2010). Compositional studies have also shown that certain members of the flora may be indicative of disease states. For example, the bacterium *Faecalibacterium prausnitzii*, a member of the Ruminococcaceae family is absent in samples from Crohn's disease (CD) patients (Sokol et al., 2008; Willing et al., 2009). Moreover, amelioration of disease with the anti TNF antibody, infliximab, allows *F. prausnitzii* to recolonize (Swidsinski et al., 2008). Furthermore, *F. prausnitzii* levels do not change in patients with ulcerative colitis, a disease characterized by Th2-type inflammation, suggesting that this microorganism is particularly susceptible to Th1 (predominant in CD) and not Th2-type inflammation.

A recent study analyzed the human fecal microbiota of 39 individuals from 6 different developed countries and found that there are three distinct ecosystems in the human gut, or "enterotypes." The genus *Bacteroides* is highly abundant in enterotype 1, *Prevotella* is very common for enterotype type 2 and *Ruminococcus* in enterotype 3. Although the functional implications of all enterotypes are still under study, it seems that each enterotype favors the synthesis of certain metabolic products. Enterotype 1 ecosystem favors the production of enzymes involved in biotin synthesis, whereas enterotype 2 species produce more enzymes for thiamine synthesis (Arumugam et al., 2011). Another study examined the fecal microbiota of 98 individuals and showed that samples clustered into two of the previously described enterotypes: *Bacteroides*-rich and *Prevotella*-rich ecosystems. Interestingly, the authors were able to correlate these enterotypes with long-term diets. Individuals with high animal fat in their diet clustered in the *Bacteroides* enterotype, whereas diets high in carbohydrate clustered in the *Prevotella* enterotype. A short-term intervention with a carbohydrate-rich diet did not cause a switch from *Bacteroides* into *Prevotella* enterotypes, suggesting that they reflect long-term diet conditions (Wu et al., 2011). There are still many unanswered questions regarding the correlation between intestinal enterotypes and the immune status of the host, as well as the possibility that there are more than three enterotypes or subenterotypes. However, the compositional classification of the human gut microbiota into functional ecosystems has enormous potential as a marker of disease and as a treatment.

FINAL REMARKS

The immune environment of the gut is characterized by tolerance. To achieve this state, host immune cells must learn to exhibit a tolerogenic behavior toward most microbial antigens. Thus, besides protecting the host from microbial attack, post-natal bacterial colonization likely serves the immune system with another important function: to be used as the second "training ground" for some naïve immune cells. Although this view poses a caveat to the traditional view of the main function of immune system – to distinguish between self and non-self – it may explain why the majority of the bacterial antigens in the gut are perceived as self-antigens.

More mechanisms on how and what components of the commensal microbiota affect mucosal and systemic immunity will undoubtedly continue to be found. This will not only expand our view of our microbiota as an integral part of our bodies (Gill et al., 2010), but it also opens the real possibility of manipulating specific functions of the microbiota to tune immune defects that cause infectious, inflammatory, or autoimmune disease. Learning from the strategies that the microbiota use to balance the different arms of the immune response and using them to overcome disease may prove to be more efficient than modifying the immune imbalances that originate from defective genes and microbial dysbiosis.

ACKNOWLEDGMENTS

The authors would like to thank Marta Wlodarska, Dr. Benjamin Willing, and Dr. Jerôme Tourret for the critical revision of this manuscript, and Fern Ness for the artwork for **Figure 1**. The Finlay laboratory is supported by operating grants from Canadian Institutes of Heath Research (CIHR), the Crohn and Colitis Foundation of Canada (CCFC), and the Howard Hughes Medical Institute (HHMI).

REFERENCES

Abrams, G. D., Bauer, H., and Sprinz, H. (1963). Influence of the normal flora on mucosal morphology and cellular renewal in the ileum. A comparison of germ-free and conventional mice. *Lab. Invest.* 12, 355–364.

Adlerberth, I. (2008). Factors influencing the establishment of the intestinal microbiota in infancy. *Nestle Nutr. Workshop Ser. Pediatr. Program.* 62, 13–29; discussion 29–33.

Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394, 744–751.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D. R., Fernandes, G. R., Tap, J., Bruls, T., Batto, J. M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H. B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E. G., Wang, J., Guarner, F., Pedersen, O., De Vos, W. M., Brunak, S., Dore, J., Antolin, M., Artiguenave, F., Blottiere, H. M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K. U., Friss, C., Van De Guchte, M., Guedon, E., Haimet, F., Huber, W., Van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S. D., and Bork, P. (2011). Enterotypes of the human gut microbiome. *Nature* 473, 174–180.

Atarashi, K., Nishimura, J., Shima, T., Umesaki, Y., Yamamoto, M., Onoue, M., Yagita, H., Ishii, N., Evans, R., Honda, K., and Takeda, K. (2008). ATP drives lamina propria T(H)17 cell differentiation. *Nature* 455, 808–812.

Ather, J. L., Ckless, K., Martin, R., Foley, K. L., Suratt, B. T., Boyson, J. E., Fitzgerald, K. A., Flavell, R. A., Eisenbarth, S. C., and Poynter, M. E. (2011). Serum amyloid A activates the NLRP3 inflammasome and promotes Th17 allergic asthma in mice. *J. Immunol.* 187, 64–73.

Benson, A. K., Kelly, S. A., Legge, R., Ma, F., Low, S. J., Kim, J., Zhang, M., Oh, P. L., Nehrenberg, D., Hua, K., Kachman, S. D., Moriyama, E. N., Walter, J., Peterson, D. A., and Pomp, D. (2010). 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.* 107, 18933–18938.

Bouskra, D., Brezillon, C., Berard, M., Werts, C., Varona, R., Boneca, I. G., and Eberl, G. (2008). Lymphoid tissue genesis induced by commensals through NOD1 regulates intestinal homeostasis. *Nature* 456, 507–510.

Buonocore, S., Ahern, P. P., Uhlig, H. H., Ivanov, Ii, Littman, D. R., Maloy, K. J., and Powrie, F. (2010). Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* 464, 1371–1375.

Clarke, T. B., Davis, K. M., Lysenko, E. S., Zhou, A. Y., Yu, Y., and Weiser, J. N. (2010). Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. *Nat. Med.* 16, 228–231.

Dobber, R., Hertogh-Huijbregts, A., Rozing, J., Bottomly, K., and Nagelkerken, L. (1992). The involvement of the intestinal microflora in the expansion of CD4+ T cells with a naive phenotype in the periphery. *Dev. Immunol.* 2, 141–150.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science* 308, 1635–1638.

Endt, K., Stecher, B., Chaffron, S., Slack, E., Tchitchek, N., Benecke, A., Van Maele, L., Sirard, J. C., Mueller, A. J., Heikenwalder, M., Macpherson, A. J., Strugnell, R., Von Mering, C., and Hardt, W. D. (2010). The microbiota mediates pathogen clearance from the gut lumen after non-typhoidal *Salmonella* diarrhea. *PLoS Pathog.* 6, e1001097. doi:10.1371/journal.ppat.1001097

Ferreira, R. B., Gill, N., Willing, B. P., Antunes, L. C., Russell, S. L., Croxen, M. A., and Finlay, B. B. (2011). The intestinal microbiota plays a role in *Salmonella*-induced colitis independent of pathogen colonization. *PLoS ONE* 6, e20338. doi:10.1371/journal.pone.0020338

Gaboriau-Routhiau, V., Rakotobe, S., Lecuyer, E., Mulder, I., Lan, A., Bridonneau, C., Rochet, V., Pisi, A., De Paepe, M., Brandi, G., Eberl, G., Snel, J., Kelly, D., and Cerf-Bensussan, N. (2009). The key role of segmented filamentous bacteria in the coordinated maturation of gut helper T cell responses. *Immunity* 31, 677–689.

Geuking, M. B., Cahenzli, J., Lawson, M. A., Ng, D. C., Slack, E., Hapfelmeier, S., Mccoy, K. D., and Macpherson, A. J. (2011). Intestinal bacterial colonization induces mutualistic regulatory T cell responses. *Immunity* 34, 794–806.

Gill, N., Wlodarska, M., and Finlay, B. B. (2010). The future of mucosal immunology: studying an integrated system-wide organ. *Nat. Immunol.* 11, 558–560.

Hansson, G. C., and Johansson, M. E. (2010). The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Gut Microbes* 1, 51–54.

Hooper, L. V., and Macpherson, A. J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat. Rev. Immunol.* 10, 159–169.

Ivanov, I. I., Atarashi, K., Manel, N., Brodie, E. L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K. C., Santee, C. A., Lynch, S. V., Tanoue, T., Imaoka, A., Itoh, K., Takeda, K., Umesaki, Y., Honda, K., and Littman, D. R. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* 139, 485–498.

Ivanov, I. I., Frutos Rde, L., Manel, N., Yoshinaga, K., Rifkin, D. B., Sartor, R. B., Finlay, B. B., and Littman, D. R. (2008). Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. *Cell Host Microbe* 4, 337–349.

Kinnebrew, M. A., Ubeda, C., Zenewicz, L. A., Smith, N., Flavell, R. A., and Pamer, E. G. (2010). Bacterial flagellin stimulates Toll-like receptor 5-dependent defense against vancomycin-resistant *Enterococcus* infection. *J. Infect. Dis.* 201, 534–543.

Lathrop, S. K., Bloom, S. M., Rao, S. M., Nutsch, K., Lio, C. W., Santacruz, N., Peterson, D. A., Stappenbeck, T. S., and Hsieh, C. S. (2011). Peripheral education of the immune system by colonic commensal microbiota. *Nature* 478, 250–254.

Lee, Y. K., and Mazmanian, S. K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* 330, 1768–1773.

Lee, Y. K., Menezes, J. S., Umesaki, Y., and Mazmanian, S. K. (2010). Microbes and health Sackler colloquium: proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl. Acad. Sci. U.S.A.* 108(Suppl. 1), 4615–4622.

Ley, R. E., Peterson, D. A., and Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124, 837–848.

Matsunaga, T., and Rahman, A. (1998). What brought the adaptive immune system to vertebrates? The jaw hypothesis and the seahorse. *Immunol. Rev.* 166, 177–186.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., and Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122, 107–118.

Mazmanian, S. K., Round, J. L., and Kasper, D. L. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* 453, 620–625.

McFall-Ngai, M. (2007). Adaptive immunity: care for the community. *Nature* 445, 153.

Moreau, M. C., Ducluzeau, R., Guy-Grand, D., and Muller, M. C. (1978). Increase in the population of duodenal immunoglobulin A plasmocytes in axenic mice associated with different living or dead bacterial strains of intestinal origin. *Infect. Immun.* 21, 532–539.

Nardi, R. M., Silva, M. E., Vieira, E. C., Bambirra, E. A., and Nicoli, J. R. (1989). Intragastric infection of germfree and conventional mice with *Salmonella typhimurium*. *Braz. J. Med. Biol. Res.* 22, 1389–1392.

Niess, J. H., Leithauser, F., Adler, G., and Reimann, J. (2008). Commensal gut flora drives the expansion of proinflammatory CD4 T cells in the colonic lamina propria under normal and inflammatory conditions. *J. Immunol.* 180, 559–568.

Ochoa-Reparaz, J., Mielcarz, D. W., Wang, Y., Begum-Haque, S., Dasgupta, S., Kasper, D. L., and Kasper, L. H. (2010). A polysaccharide from the human commensal *Bacteroides fragilis* protects against CNS demyelinating disease. *Mucosal Immunol.* 3, 487–495.

Penders, J., Thijs, C., Vink, C., Stelma, F. F., Snijders, B., Kummeling, I., Van Den Brandt, P. A., and Stobberingh, E. E. (2006). Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 118, 511–521.

Rask, C., Evertsson, S., Telemo, E., and Wold, A. E. (2005). A full flora, but not monocolonization by *Escherichia coli* or lactobacilli, supports tolerogenic processing of a

fed antigen. *Scand. J. Immunol.* 61, 529–535.

Rescigno, M. (2011). The intestinal epithelial barrier in the control of homeostasis and immunity. *Trends Immunol.* 32, 256–264.

Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323.

Sekirov, I., and Finlay, B. B. (2009). The role of the intestinal microbiota in enteric infection. *J. Physiol. (Lond.)* 587, 4159–4167.

Sekirov, I., Tam, N. M., Jogova, M., Robertson, M. L., Li, Y. L., Lupp, C., and Finlay, B. B. (2008). Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infect. Immun.* 76, 4726–4736.

Smith, K., Mccoy, K. D., and Macpherson, A. J. (2007). Use of axenic animals in studying the adaptation of mammals to their commensal intestinal microbiota. *Semin. Immunol.* 19, 59–69.

Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J. J., Blugeon, S., Bridonneau, C., Furet, J. P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottiere, H. M., Dore, J., Marteau, P., Seksik, P., and Langella, P. (2008). *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16731–16736.

Sprinz, H., Kundel, D. W., Dammin, G. J., Horowitz, R. E., Schneider, H., and Formal, S. B. (1961). The response of the germfree guinea pig to oral bacterial challenge with *Escherichia coli* and *Shigella flexneri*. *Am. J. Pathol.* 39, 681–695.

Strauch, U. G., Obermeier, F., Grunwald, N., Gurster, S., Dunger, N., Schultz, M., Griese, D. P., Mahler, M., Scholmerich, J., and Rath, H. C. (2005). Influence of intestinal bacteria on induction of regulatory T cells: lessons from a transfer model of colitis. *Gut* 54, 1546–1552.

Swidsinski, A., Loening-Baucke, V., Vaneechoutte, M., and Doerffel, Y. (2008). Active Crohn's disease and ulcerative colitis can be specifically diagnosed and monitored based on the biostructure of the fecal flora. *Inflamm. Bowel Dis.* 14, 147–161.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J. P., Ugarte, E., Munoz-Tamayo, R., Paslier, D. L., Nalin, R., Dore, J., and Leclerc, M. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–2584.

Turnbaugh, P. J., Hamady, M., Yatsunenko, T., Cantarel, B. L., Duncan, A., Ley, R. E., Sogin, M. L., Jones, W. J., Roe, B. A., Affourtit, J. P., Egholm, M., Henrissat, B., Heath, A. C., Knight, R., and Gordon, J. I. (2009). A core gut microbiome in obese and lean twins. *Nature* 457, 480–484.

Uematsu, S., Fujimoto, K., Jang, M. H., Yang, B. G., Jung, Y. J., Nishiyama, M., Sato, S., Tsujimura, T., Yamamoto, M., Yokota, Y., Kiyono, H., Miyasaka, M., Ishii, K. J., and Akira, S. (2008). Regulation of humoral and cellular gut immunity by lamina propria dendritic cells expressing Toll-like receptor 5. *Nat. Immunol.* 9, 769–776.

Umesaki, Y., Okada, Y., Matsumoto, S., Imaoka, A., and Setoyama, H. (1995). Segmented filamentous bacteria are indigenous intestinal bacteria that activate intraepithelial lymphocytes and induce MHC class-II molecules and fucosyl asialo Gm1 glycolipids on the small-intestinal epithelial-cells in the ex-germ-free mouse. *Microbiol. Immunol.* 39, 555–562.

van der Waaij, D., Berghuis-De Vries, J. M., and Lekkerkerk, L.-V. (1971). Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *J. Hyg. (Lond.)* 69, 405–411.

Wen, L., Ley, R. E., Volchkov, P. Y., Stranges, P. B., Avanesyan, L., Stonebraker, A. C., Hu, C. Y., Wong, F. S., Szot, G. L., Bluestone, J. A., Gordon, J. I., and Chervonsky, A. V. (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* 455, 1109–1113.

Willing, B., Halfvarson, J., Dicksved, J., Rosenquist, M., Jarnerot, G., Engstrand, L., Tysk, C., and Jansson, J. K. (2009). Twin studies reveal specific imbalances in the mucosa-associated microbiota of patients with ileal Crohn's disease. *Inflamm. Bowel Dis.* 15, 653–660.

Willing, B. P., Gill, N., and Finlay, B. B. (2010). The role of the immune system in regulating the microbiota. *Gut Microbes* 1, 213–223.

Willing, B. P., Vacharaksa, A., Croxen, M., Thanachayanont, T., and Finlay, B. B. (2011). Altering host resistance to infections through microbial transplantation. *PLoS ONE* 6, e26988. doi:10.1371/journal.pone.0026988

Willing, B. P., and Van Kessel, A. G. (2007). Enterocyte proliferation and apoptosis in the caudal small intestine is influenced by the composition of colonizing commensal bacteria in the neonatal gnotobiotic pig. *J. Anim. Sci.* 85, 3256–3266.

Wlodarska, M., Willing, B., Keeney, K. M., Menendez, A., Bergstrom, K. S., Gill, N., Russell, S. L., Vallance, B. A., and Finlay, B. B. (2011). Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated *Citrobacter rodentium*-induced colitis. *Infect. Immun.* 79, 1536–1545.

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., Bewtra, M., Knights, D., Walters, W. A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F. D., and Lewis, J. D. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108.

Wu, H. J., Ivanov, Ii, Darce, J., Hattori, K., Shima, T., Umesaki, Y., Littman, D. R., Benoist, C., and Mathis, D. (2010). Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* 32, 815–827.

Zachar, Z., and Savage, D. C. (1979). Microbial interference and colonization of the murine gastrointestinal tract by *Listeria monocytogenes*. *Infect. Immun.* 23, 168–174.

Zaph, C., Du, Y., Saenz, S. A., Nair, M. G., Perrigoue, J. G., Taylor, B. C., Troy, A. E., Kobuley, D. E., Kastelein, R. A., Cua, D. J., Yu, Y., and Artis, D. (2008). Commensal-dependent expression of IL-25 regulates the IL-23-IL-17 axis in the intestine. *J. Exp. Med.* 205, 2191–2198.

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.

Received: 01 December 2011; paper pending published: 22 December 2011; accepted: 14 February 2012; published online: 05 March 2012.

*Citation: Arrieta M-C and Finlay BB (2012) The commensal microbiota drives immune homeostasis. Front. Immun. **3**:33. doi: 10.3389/fimmu.2012.00033*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in IMMUNOLOGY

HYPOTHESIS AND THEORY ARTICLE
published: 03 May 2012
doi: 10.3389/fimmu.2012.00096

A basal chordate model for studies of gut microbial immune interactions

Larry J. Dishaw[1]*, Jaime A. Flores-Torres[2], M. Gail Mueller[1,3], Charlotte R. Karrer[3], Diana P. Skapura[3], Daniela Melillo[4], Ivana Zucchetti[4], Rosaria De Santis[4], Maria Rosaria Pinto[4] and Gary W. Litman[1,3]

[1] Department of Pediatrics, USF/ACH Children's Research Institute, University of South Florida College of Medicine, St. Petersburg, FL, USA
[2] Division of Neonatology, Department of Pediatrics, University of South Florida College of Medicine, St. Petersburg, FL, USA
[3] Department of Molecular Genetics, All Children's Hospital/Johns Hopkins Medicine, St. Petersburg, FL, USA
[4] Department of Animal Physiology and Evolution, Stazione Zoologica Anton Dohrn, Napoli, Italy

Edited by:
Maria Yazdanbakhsh, Leiden University Medical Center, Netherlands

Reviewed by:
Robert Adam Harris, Karolinska Institutet, Sweden
John P. Vasilakos, 3M Company, USA

***Correspondence:**
Larry J. Dishaw, Department of Pediatrics, University of South Florida, 140 Seventh Avenue South, CRI 3008, St. Petersburg, FL 33701, USA.
e-mail: ldishaw@health.usf.edu

Complex symbiotic interactions at the surface of host epithelia govern most encounters between host and microbe. The epithelium of the gut is a physiologically ancient structure that is comprised of a single layer of cells and is thought to possess fully developed immunological capabilities. *Ciona intestinalis* (sea squirt), which is a descendant of the last common ancestor of all vertebrates, is a potentially valuable model for studying barrier defenses and gut microbial immune interactions. A variety of innate immunological phenomena have been well characterized in *Ciona*, of which many are active in the gut tissues. Interactions with gut microbiota likely involve surface epithelium, secreted immune molecules including variable region-containing chitin-binding proteins, and hemocytes from a densely populated laminar tissue space. The microbial composition of representative gut luminal contents has been characterized by molecular screening and a potentially relevant, reproducible, dysbiosis can be induced via starvation. The dialog between host and microbe in the gut can be investigated in *Ciona* against the background of a competent innate immune system and in the absence of the integral elements and processes that are characteristic of vertebrate adaptive immunity.

Keywords: protochordate, innate immunity, *Ciona*, microbiota, gut epithelium

INTRODUCTION

Because of their phylogenetic position relative to vertebrates, deuterostome invertebrates (**Figure 1**) are compelling models for studies of immunity. Of the representative deuterostome invertebrates, echinoderms, which include sea urchins and starfish, diverged prior to chordates and have proven to be highly informative models in which to examine innate immunity (Rast and Messier-Solek, 2008; Messier-Solek et al., 2010). Protochordates are comprised of invertebrate chordates such as sea squirts (Urochordata or Tunicata) and amphioxus (Cephalochordata). These species, which share certain developmental features with vertebrates and possess competent innate immunity, diverged prior to the origins of the vertebrate adaptive immune system.

Ciona intestinalis (sea squirt), which has been the focus of our recent efforts, is relatively easy to maintain and propagate (**Figure 2**) at room temperature and continues to serve as a highly informative model for studies of animal development (Katz, 1983; Meinertzhagen and Okamura, 2001; Canestro et al., 2003; Shi et al., 2005; Baghdiguian et al., 2007; Davidson, 2007; Sasakura et al., 2009) and immune defense (Fujita et al., 2004; Melillo et al., 2006; Parrinello, 2009; Sasaki et al., 2009; Zucchetti et al., 2009). As a model of animal development (Satoh and Levine, 2005; Lemaire et al., 2008; Christiaen et al., 2009) *Ciona* has proven invaluable in: (1) unraveling details of Hox-gene influences on development (Ikuta et al., 2010), (2) mapping pathways in cardiac development (Davidson, 2007), (3) defining the roles of cis-regulatory networks (Kubo et al., 2010), (4) modeling the effects of maternally derived epigenetic silencing (Sasakura et al., 2010), and (5) defining the evolution of the cell death machinery (Terajima et al., 2003; Baghdiguian et al., 2007). Many of the involved processes utilize signaling pathways that are relevant to studies concerning immunity and immune homeostasis.

Deuterostome invertebrates possess homologs of a large number of vertebrate innate immune receptors, effectors, and their corresponding regulatory elements (Rast and Messier-Solek, 2008; Messier-Solek et al., 2010). The most surprising finding regarding immunity in sea urchin and amphioxus is the expansion of multigene families encoding homologs of different innate immune pattern recognition receptors (PRRs), including: Toll-like receptors (TLRs), scavenger receptors, and leucine-rich receptor (LRR)-containing intracellular sensors such as nucleotide-binding oligomerization domain-like receptors (NLRs; Rast et al., 2006; Holland et al., 2008). Novel evolutionary constraints most likely led to the lineage-specific expansions and functional divergence of the various gene families encoding these molecules in amphioxus and sea urchin. Detailed sequence analyses have revealed examples of parallel or convergent evolution relating to the expansions (Leulier and Lemaitre, 2008). Very little sequence similarity among presumed homologs (i.e., lack of one-to-one orthology) is evident between these receptors in invertebrate deuterostomes and their vertebrate counterparts. Innate immune genes in *Ciona* have not undergone the expansions seen in amphioxus and sea urchin (Dehal et al., 2002; Hughes and Friedman, 2005).

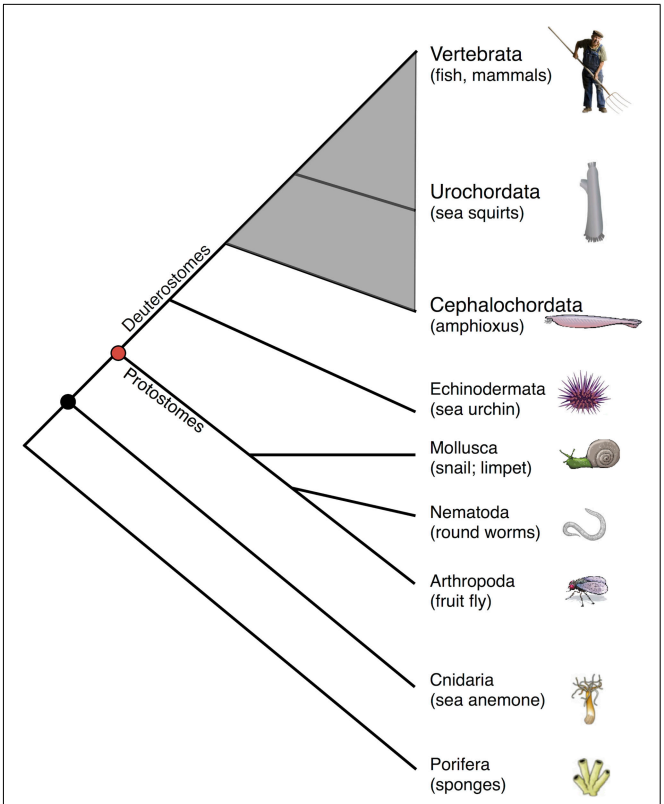


FIGURE 1 | Simplified cladogram of animal phylogeny emphasizing chordates (shaded gray) relative to other animal forms. Note that only two protochordate taxa are shown, and all are exclusively marine animals.

INNATE IMMUNITY AND INFLAMMATION IN *CIONA*

Comparative biological studies have revealed that most multicellular life is protected by conserved mechanisms of innate immunity which include: barrier defenses [i.e., host (mucosal) epithelium] and the associated non-specific secretory components (e.g., antibacterial peptides), PRRs, which are present on the surface and within the cytoplasm of phagocytes and other host cells, various phagocyte effector mechanisms (e.g., reactive oxygen species, induced cell death), and different enzymatically catalyzed cascades involved in clotting, melanization, and complement activation (**Figure 3**; Lavelle et al., 2010; Maldonado-Contreras and McCormick, 2011). A variety of innate immune genes and mechanisms have been identified in *Ciona* (Parrinello, 2009), which based on histological and molecular characterizations have evolved complex barrier defense strategies. Challenge with microbe-associated molecular patterns (MAMPs) such as the Gram-negative lipopolysaccharide (LPS) induces inflammatory-like reactions that are characterized, in part, by tumor-necrosis factor (TNF)-like gene expression and cell recruitment to various body compartments (Di Bella and De Leo, 2000; Melillo et al., 2006; Parrinello et al., 2007, 2008, 2010; Cammarata et al., 2008; Bonura et al., 2009). These responses can recruit a variety of cell types (Parrinello et al., 1996; Arizza and Parrinello, 2009; Arizza et al., 2011) and induce a number of immunological phenomena (Smith and Peddie, 1992; Melillo et al., 2006; Parrinello et al., 2007), including the expression of characteristic innate immune receptors (Shida et al., 2003; Parrinello, 2009; Sasaki et al., 2009; Dishaw et al., 2011). Although a variety of different MAMPs induce *Ciona* TLRs, LPS does not activate TLR expression directly, as with TLR4 in vertebrates, suggesting the presence of alternative LPS sensors (Sasaki et al., 2009). A repertoire of innate effectors [e.g., TLRs, TNF, complement components, and the protochordate-specific variable region-containing chitin-binding proteins (VCBPs)] are expressed in the gut of *Ciona* (Marino et al., 2002; Sasaki et al., 2009; Skjoedt et al., 2010; Dishaw et al., 2011) and could play essential roles in the stable maintenance of host–microbial interactions (see below).

Documentation of these types of functional effects is critical to the design of experimental approaches for characterizing the interactions of immune receptors with gut microbiota. However, enumerating gene homologs and defining their expression patterns is one thing, understanding their dynamic interactions in a cellular and molecular context is a daunting challenge in invertebrate systems, particularly those that inhabit marine and aquatic environments. Despite a number of potential shortcomings, *Ciona* has the potential to reveal conserved mechanisms sustaining the evolution of host–microbial interactions (see below). Experimental manipulation of host–microbial interactions at the gut epithelial interface is critical to such studies and in order to approach these questions, it is essential to first understand the *Ciona* gut as both a physical barrier and as an immunological organ.

AN INVERTEBRATE CHORDATE GUT MODEL

Ciona is a highly successful, cosmopolitan solitary tunicate that has adapted to living in diverse marine environments (Caputi et al., 2007). Other tunicates include both colonial and solitary forms and spend their adult lives as sessile, filter-feeding, organisms. In addition, a few pelagic forms have been identified (Denoeud et al., 2010; Nishida et al., 2010). Because the tunicate feeding strategy involves siphoning seawater, the gut in *Ciona* is in continuous contact with both dietary and seawater microbiota where in addition to its essential physiological role, it acts as both a physical barrier and site of continuous immunological interaction. An immunocompetent gut, which includes mucosal immunity mediated by surface epithelium, is present even in the simplest metazoans (Bosch et al., 2009). In mammals, proper development of gut mucosal immune tissues is dependent on the proper timing and colonization of the gut by microbial communities (see below); however, details governing these events remain to be defined (Cebra et al., 1998; Cebra, 1999; Hooper and Gordon, 2001; Mazmanian et al., 2005; Edelman and Kasper, 2008; Kanther and Rawls, 2010). Intensive studies in comparative models, such as *Ciona*, have the potential for shedding light on the basic biology of gut immune homeostasis and in turn, may reveal basic mechanisms of dysfunctional gut immunity [e.g., inflammatory bowel diseases (IBD) in mammals]. As an initial step in adapting a urochordate as a new model system for understanding gut immune homeostasis and mucosal immunity, we have combined cellular, molecular, and microbiological approaches to characterize the *Ciona* gut and its microbiota. At this preliminary stage of investigation, basic aspects of complex microbial community dynamics can be identified that mirror many of the

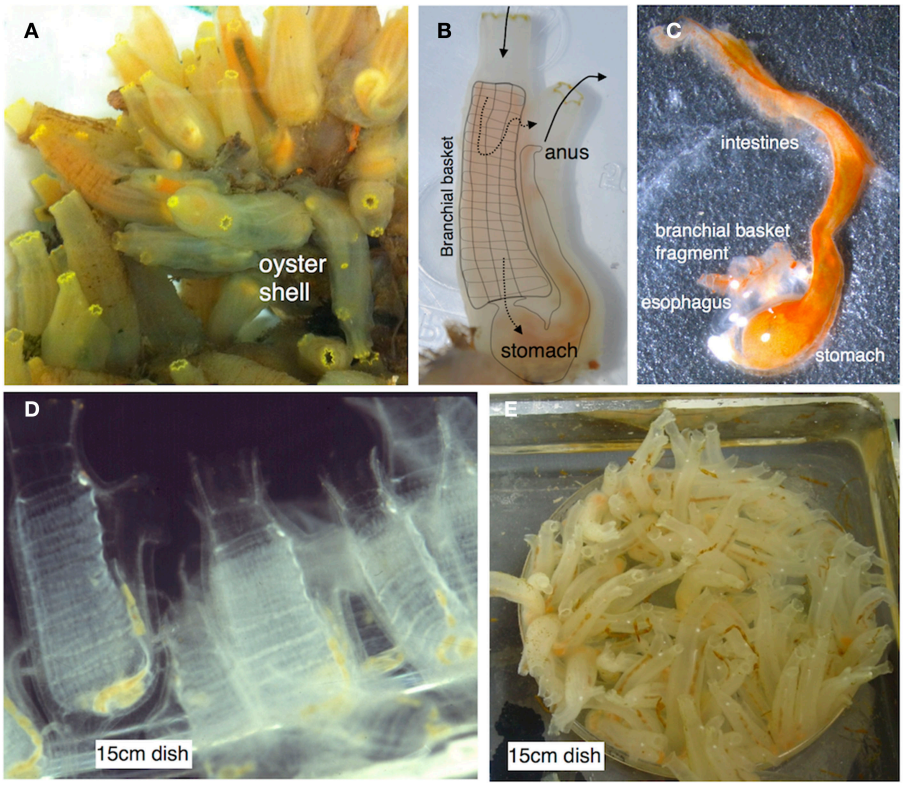


FIGURE 2 | (A) *Ciona intestinalis* is a solitary urochordate that typically grows in close proximity on suitable substrates. **(B)** *Ciona* is relatively translucent and the gut can be visualized through the tunic (overlay depicts the branchial basket and gut; solid arrows depict water flow; dotted arrows indicate two routes of flow). **(C)** The entire gut can be readily dissected. **(D)** Juveniles are attracted and attach to standard tissue culture plastic plates. **(E)** *Ciona* can be reared to adulthood in a laboratory environment (Cirino et al., 2002; Joly et al., 2007). *Ciona* adults are hermaphrodites and release sperm and egg into the water column. Typically, fertilization leads to rapid development and hatching of swimming tadpole larva. After settlement, the animal undergoes metamorphosis into a permanent, sessile, filter-feeding lifestyle; colonization of the gut by microbes is likely immediate. Images **(D,E)** are courtesy Dr. Paola Cirino.

core features and symbiotic intricacies of host–microbe interactions that are recognized in the mammalian gut ecosystem (Savage, 1977; Hooper and Gordon, 2001; Willing et al., 2011). The interactions between host and gut microflora are not simple, but instead involve complex mutualisms (Bischoff, 2011) that ultimately help govern immune development, and the establishment and maintenance of immune homeostasis (Artis, 2008; Chung and Kasper, 2010; Hooper and Macpherson, 2010).

COMPLEX BIOLOGY OF THE *CIONA* GUT

Ciona filters microbe-rich seawater through a modified pharynx where ciliated cells push food particles into a gut, which is divided into esophagus, stomach, mid-gut, and hind-gut; the latter two are referred to collectively as intestines. The esophagus connects the branchial basket to the stomach in which both cilia and mucous glands are highly developed for the efficient transfer of foods. The stomach epithelium forms many cilia-rich ridges and grooves and is composed of at least two major epithelial cell types, as well as an undifferentiated cell population (Burighel and Cloney, 1977). The stomach epithelium, which is presumed to be the site of most digestive enzyme secretion, contains secretory cells, and is coated by a thin layer of mucus. The mid-gut is distinguished by an interior typhlosole that runs the entire length and is rich in testicular acini. Three types of largely granular epithelial cells, absorptive, mucous, and large round (or elliptical), define the mid-gut. Glycogen stores are concentrated in the mid-gut and to a lesser degree in the stomach. Energy is stored as both fat and glycogen (Yonge, 1925). The sexual ducts exit the atrial siphon adjacent to the hind-gut. Although absorption is most prominent in the mid-gut, diffusion of dissolved substances occurs throughout the alimentary track. The *Ciona* gut demonstrates complex epithelial cell renewal traits (Ermak, 1981), which are of particular importance to several aspects of mucosal physiology, including immune function. The highly developed and compartmentalized stomach and distinct intestinal region in *Ciona* morphologically resemble that of more recently diverged chordates (Millar, 1953; Burighel and Cloney, 1977).

COMPLEX MICROBIAL COMMUNITY DYNAMICS DEFINE THE *CIONA* GUT

Details surrounding the relationships between filter-feeding invertebrates and the microbial communities colonizing their guts are lacking but may be broadly relevant to gut immunity in vertebrates for determining: (1) the role of diet in gut microbial ecology, (2)

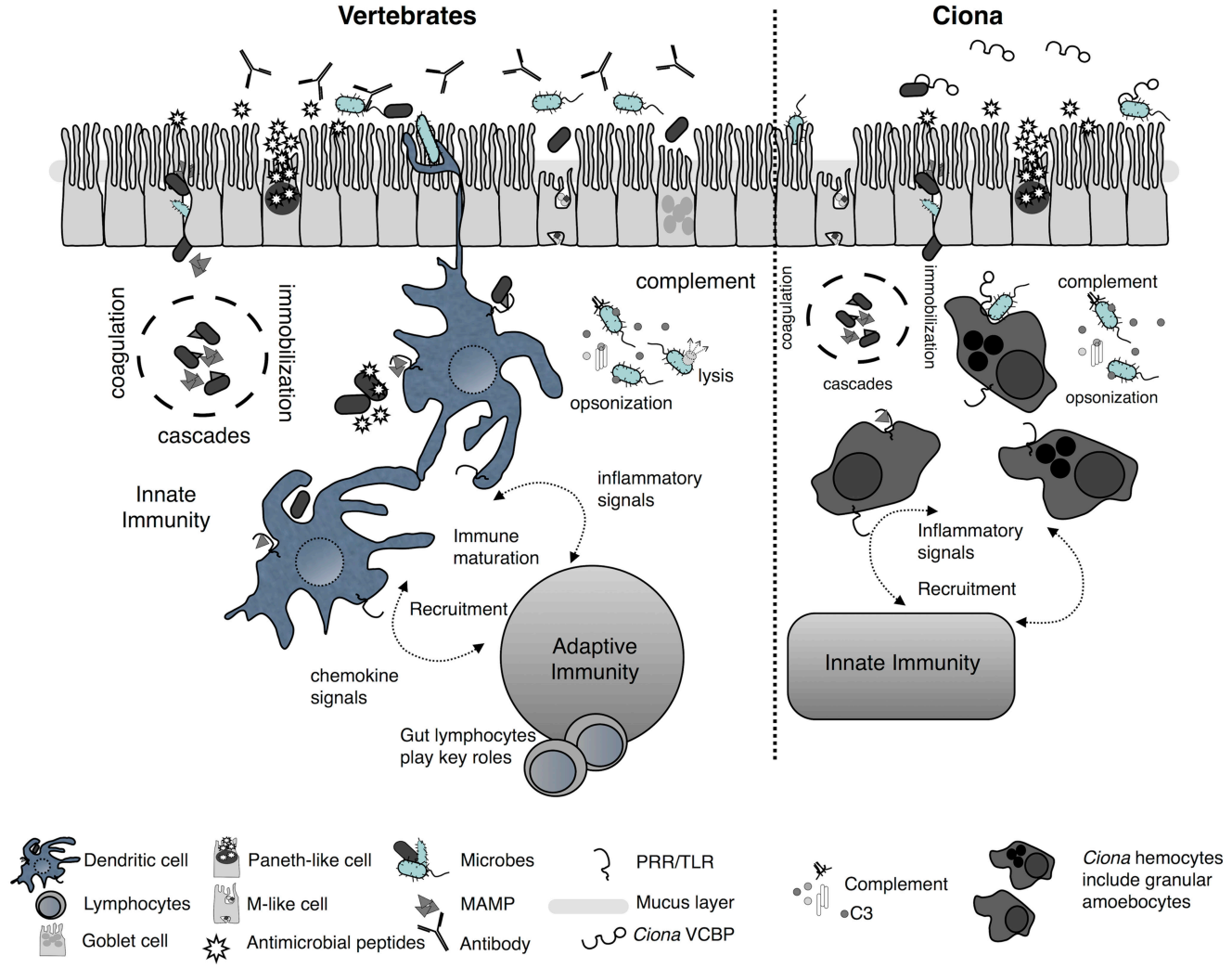


FIGURE 3 | Simplified schematic of mucosal immunity emphasizing barrier defense strategies of vertebrates and *Ciona*. Barrier defenses (across the epithelium) are governed by innate immune phenomena characterized by the secretion of mucus, antimicrobials, and soluble immune molecules. On the basolateral surface and deeper, host innate immunity consists of various proteolytic-coagulation cascades for wound healing and microbe trapping, as well as complement defense pathways. Distinct phagocytes populate this area as well other cell types [e.g., dendritic cells (DC) in vertebrates]. Gut DCs sample luminal antigens and present them to the adaptive immune system which includes gut-specific lymphocytes of both T and B cell lineages. This results in the maturation of immunity and the recruitment of additional cell types. A parallel, more simplified, system in *Ciona* has evolved to also include barrier defenses across distinct epithelial lineages in the gut and the secretion of immune mediators, including antimicrobial peptides, into the lumen. In the basolateral side, distinct populations of hemocytes, including granular amoebocytes, populate the laminar connective tissues. Immunological competence in this area is mediated by coagulation/immobilization cascades and microbe trapping, complement defenses, distinct antigenic sampling via PRRs, secretion of pro-inflammatory signals, and recruitment of specialized hemocytes. However, as opposed to the vertebrate gut, the *Ciona* gut generates immune amplification within an innate immune context and does not couple to a distinct, more specialized, system such as adaptive immunity.

the nature of host selection of gut microbiota, and (3) the role of microbiota in gut immune homeostasis. These features of gut physiology are essential to our defining of homeostasis, even if a model organism such as *Ciona* does not share all the innate receptor orthologs found in mammals. Because the composition of gut microbiota is relevant to host physiology for a variety of reasons not limited to their metabolic output, we suspect that the gut microbial communities in *Ciona* will not simply reflect diet and/or environmental availability, but will reveal species-specific communities that determine or influence various aspects of overall immune competence.

Gut bacteria from four *Ciona* populations (Woods Hole, MA, USA; San Diego and Monterey Bay, CA, USA; and Naples, Italy) have been partially sampled using PCR-amplified 16S ribosomal genes recovered in small clone libraries derived from whole gut homogenates (and/or recovered fecal matter). The 16S products were then characterized by sequencing of individual clones as well as by screening of restriction fragment length polymorphisms

(**Figure 4**). A fraction of the recovered microflora (∼25 bacterial species) was cultured successfully. The *Ciona* gut revealed distinct communities of bacteria that were affected by both diet and environment, including species of metabolic significance (e.g., *Chitinophaga*). Starvation induced reproducible dysbiosis (i.e., disruption or displacement of microbial communities) and revealed bacterial families and genera that were conserved across populations and between two *Ciona* species (e.g., Gammaproteobacteria such as *Vibrio* sp. and *Shewanella* sp., as well as various *Oceanospirillales* genera; Dishaw, unpublished observations).

The successful recovery, identification, and growth of native gut bacteria from wild *Ciona* adults provides the basic background to inoculate the developing gut of juveniles grown in controlled laboratory environments (e.g., under semi- or sterile conditions or colonized by complex mixtures of microbes). Changes in community structure subsequent to experimental manipulation can be monitored using real time quantitative PCR. These experiments, which currently are in progress, have the potential to define the onset and normal timing of microbial colonization in the development of the *Ciona* gut in general, and more specifically, characterize how this interaction affects the maturation of the hemocyte-rich laminar spaces (i.e., immune tissues). Determining how interactions with microbiota affect maturation of the gut immune tissues will utilize dysbiosis techniques including antibiotic treatments and/or development of juveniles under germ-free or semi-germ-free conditions. In this regard, two features of the *Ciona* model are particularly attractive: (1) it is relatively easy to produce and maintain hundreds of *Ciona* juveniles and (2) transparency of tissues makes it feasible to visually characterize gut development and gauge luminal content (**Figure 2**) and make possible the tracking of host–microbe interactions (e.g., the use of labeled bacteria) from early in development through adulthood.

GUT MICROBIAL IMMUNE INTERACTIONS

The circulatory system of *Ciona* is open and continuous with histologically defined blood lacunae. Gut-associated lacunae, which

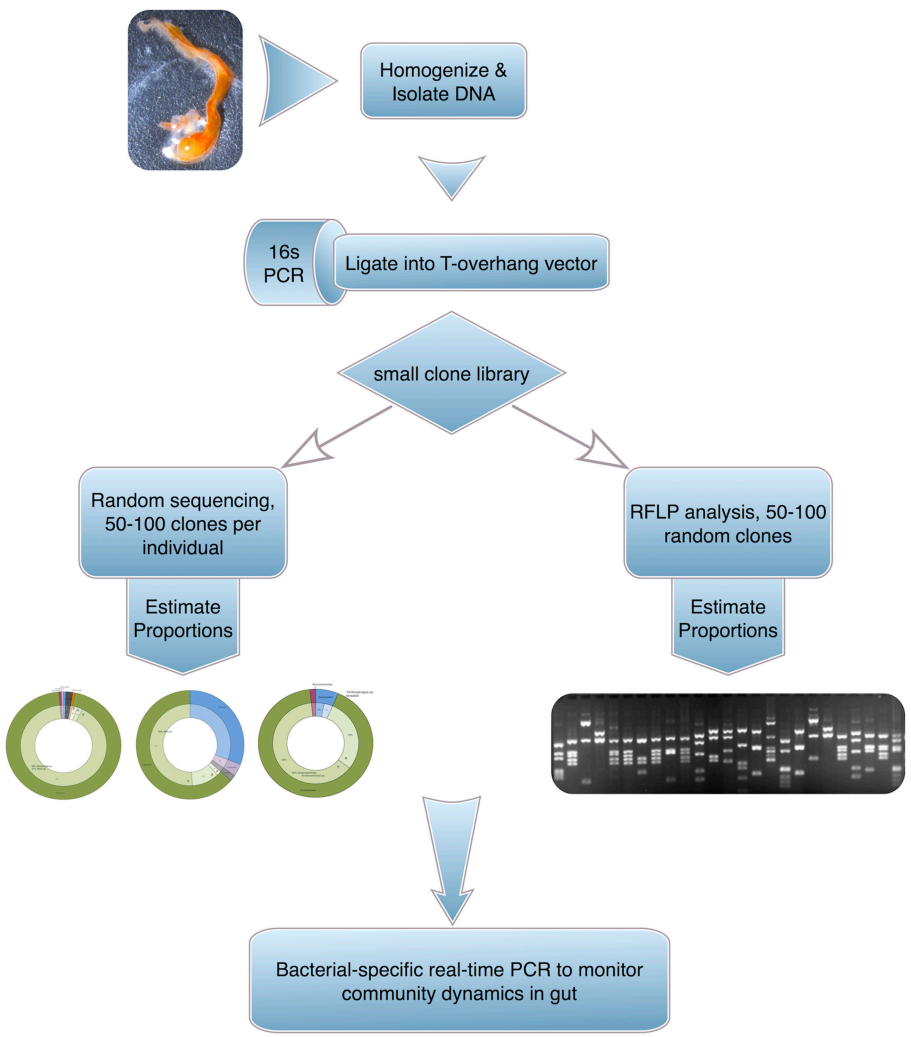


FIGURE 4 | A general approach to assess gut microbial communities in *Ciona*. The technique, albeit labor intensive, provides high-grade sequence-based identifications.

we have termed the *Ciona* gut lamina propria (Dishaw et al., 2011), share blood cells between the various tissue spaces and are richly infiltrated by a variety of hemocyte types. A complex developmental maturation of this subepithelial lamina that likely is influenced by luminal antigen exposure is indicated. Notably, the hemocytes are not restricted to the gut as morphologically indistinguishable cell types can be detected in other parts of the body. Adult *Ciona* injected with MAMPs (e.g., in the tunic) can be induced to generate localized inflammatory responses, which include an active recruitment of these and other immunocompetent cells (Di Bella and De Leo, 2000; Pinto et al., 2003; Bonura et al., 2009; Parrinello et al., 2010). In this regard, variable (V) region-containing chitin-binding proteins (VCBPs), which are expressed by distinct gut epithelial cells in amphioxus (Cannon et al., 2002), also have been identified in *Ciona* (**Figure 5**). VCBPs have been shown to be secreted by discrete cells of the stomach and intestinal epithelium into gut lumen where they interact with bacteria via their V-type immunoglobulin domains (Dishaw et al., 2011). The functional relevance of the VCBP chitin-binding domain remains unclear. *Ciona* granular amoebocytes, which also express VCBPs, are present in both blood and the laminar spaces of the gut. *In vitro* experiments have demonstrated that granular amoebocytes, recovered from blood, recognize (phagocytose) bacteria coated with VCBPs (Dishaw et al., 2011) *in vitro*. We have hypothesized that the morphologically indistinguishable cells found in the lamina propria function in an equivalent manner and play a major role in the dynamics of gut immunity in both *Ciona* and amphioxus. This hypothesis is supported further by the finding that native VCBPs bind luminal bacteria (Dishaw et al., 2011) and it is entirely likely that VCBPs enhance phagocytic recognition of gut bacteria coated with VCBPs that traverse epithelial barriers (e.g., in instances of epithelial damage). An emerging functional role for VCBPs places them in the broader context of our hypothesis by suggesting that some innate receptors, secreted into the gut lumen, may be serving still undefined roles as symbiosis factors. Penetration of the mucosal barrier would trigger an immunological event by the tagged (i.e., opsonized) microbe to protect the integrity of self (i.e., phagocytosis, inflammation, and cell recruitment). Broad analogies can be drawn between this process and immune recognition in the vertebrate mucosal environment, albeit involving different effectors. MAMP challenge across the gut mucosal barriers also may recruit cells from distant tissue spaces; however, this may be unnecessary since the subepithelial lamina is densely populated with many hemocyte types along the gut length (**Figure 5**). Several preliminary observations (Dishaw, De Stantis, and Pinto, unpublished) suggest that lamina-associated gut hemocytes from *Ciona* may be exposed to luminal contents (e.g., through injury of epithelium or exposure to factors that affect epithelial tight junctions). Electron microscopic analysis will be critical to demonstrate if hemocytes actively cross the epithelium and interact with gut microbiota and if luminal antigens enter the laminar spaces.

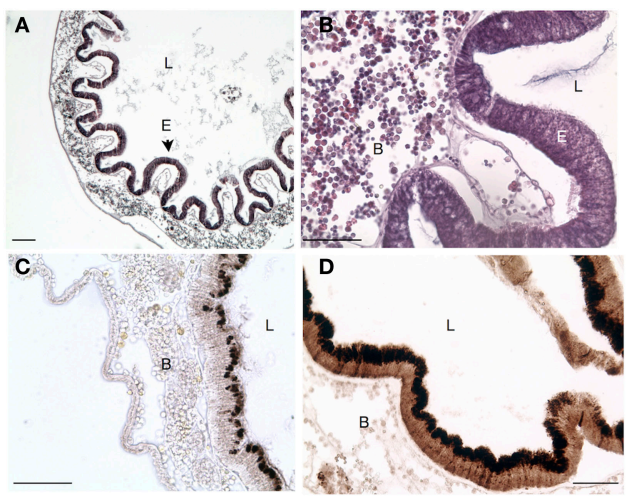


FIGURE 5 | Transverse section of stomach in *Ciona* depicting. **(A)** Characteristic folding of epithelium that is not seen in the intestines. **(B)** Higher magnification view [of **(A)**] emphasizing the abundance of laminar tissue blood hemocytes. **(C)** Antibody staining for VCBP-C in untreated *Ciona*. **(D)** Extensive upregulation of VCBP expression in animals challenged with *E. coli* lysates. Scale bar: 100 μm in **(A)** and 50 μm in **(B–D)**; Hematoxylin and eosin staining in **(A,B)** and 3,3′ diaminobenzidine (DAB) substrate detection in **(C,D)**; L, lumen; B, blood; and E, epithelium.

CONSERVED SENSORS IN HOST–MICROBE INTERACTIONS

The extent to which innate receptors are expressed by gut epithelium and the functional implications of their interaction with symbiotic or pathogenic gut microbes, as well as virulence factors, are of central interest. Innate immune receptors in vertebrates, primarily TLRs, are expressed selectively in a polarized fashion on intestinal epithelial cells (IECs) of the small and large intestines (Abreu, 2010). TLRs provide indirect signals to the adaptive immune system by first providing innate immune signals to the underlying (basal lamina) tissues. In *Ciona*, the expression, function, and regulation of a variety of immune molecules (Azumi et al., 2003; Shida et al., 2003), including antimicrobial peptides (Fedders and Leippe, 2008), TLRs, Gram-negative binding proteins (GNBPs), lipopolysaccharide binding protein (LBP), TNF, MBL, complement protein C3, as well as VCBPs, are of potential interest. *Ciona* responds to bacterial ligands in a manner consistent with the patterns of expression of TLRs and other PRRs in gut tissues, resulting in the induction of pro-inflammatory molecules, e.g., TNF (Sasaki et al., 2009). Whereas *Ciona* only has two TLRs, both are expressed in distinct locations along the gut and along with other innate immune molecules likely play a significant role in discriminating among gut commensals and sensing pathogens (Sasaki et al., 2009; Abreu, 2010). The presence of only two TLRs, which interact with more than one ligand, could be seen as a disadvantage of the *Ciona* system over mammals, in which multiple TLRs discriminate among distinct ligands. Multi-functional PRRs, such as the TLRs, may be coupled to downstream pathways which may serve to better discriminate ligands; such pathways could differ significantly in *Ciona* and investigations in this system will help reveal how various organisms discriminate among TLR ligands. Under normal conditions, the gut in *Ciona* likely maintains a state of balance (i.e., homeostasis) between tolerance and protection through host epithelial–microbe interactions, as has been seen in vertebrates.

However, most such interactions are designed to sustain ancient symbiotic relationships and are not necessarily immune-restricted (Hooper and Macpherson, 2010). Innate receptors play significant roles in these basic processes that govern homeostasis and studies that define conserved mechanisms governing host–microbial interactions in the gut are of fundamental biomedical interest.

DISCUSSION

The animal gut is host to a massively populated dynamic ecosystem of microbes (Savage, 1977) with enormously complex antigenic diversity. Such gut microbial communities in mammals are linked intimately to the maturation and development of mucosal immunity and represent an important determinant of health and disease (Mazmanian et al., 2005; Fujimura et al., 2010; Hooper and Macpherson, 2010). A particularly complex physiological challenge is posed to the mucosal immune system of the host, which must differentiate distinct populations of commensal (i.e., possibly useful) microorganisms from pathogenic communities. Specifically, some symbionts are recognized and tolerated, and subsequently form a cooperative system in the gut, whereas pathogens, which in many cases are invasive, are not well tolerated and are cleared. Traditional views of gut immunity are complicated further by numerous commensals that although beneficial, can induce a pathogenic state in the host (i.e., pathobionts, Round and Mazmanian, 2009). Breakdown of commensal-immune suppression and tolerance mechanisms can lead to disruptions of homeostasis and in turn to inflammation, resulting in a range of distinct clinical phenotypes that define acute and/or chronic IBD (Chung and Kasper, 2010; Nell et al., 2010). Details of how and why such changes in the gut ecosystem can lead to disease manifestations remain largely unknown (Bischoff, 2011). Although chronic inflammation in mammals involves a variety of gut-specific lymphocytes (i.e., adaptive immunity), gut microbial recognition, and immune homeostasis largely represent an innate immune phenomenon, for which the epithelium is primarily responsible (Artis, 2008). Details of the microbial and innate immune dialog within the gut (e.g., across the epithelium), defined exclusively in terms of innate immunity (i.e., in the absence of the adaptive immune system), remain largely unknown and underscore the unique potential of *Ciona* as a model system for investigating gut microbial–immune interactions.

Gut microbiota also appear to play a major role beyond the gut immune environment extending into a range of host physiological responses (Bischoff, 2011; Maslowski and Mackay, 2011), which include but are not limited to metabolomics and behavior (Sekirov et al., 2010). In addition to their roles in host nutritional physiology, disease protection, and pathogenesis, symbiotic microbial interactions in the gut may be linked to the phylogenetic origins of vertebrate adaptive immunity, which evolved through the recruitment of pre-existing genetic elements (Flajnik and Kasahara, 2010; Litman et al., 2010) and the spontaneous acquisition of specialized components of the recombination machinery (Dreyfus, 1992; Agrawal et al., 1998). The selective pressures that drive these events appear to have led to more than one form of molecular innovation (Cooper and Alder, 2006) and remain elusive (Matsunaga and Rahman, 1998). It has been proposed that selective pressures were maintained by an increasing need to reduce collateral damage caused by persistent or chronic infections (i.e., the toxic index hypothesis; Usharauli, 2010), as well as by innovations of strategies to manage increasingly complex symbiotic communities of the gut (McFall-Ngai, 2007; Lee and Mazmanian, 2010), which in turn also drove innovation in highly specialized cell types (Rescigno and Di, 2009; Chow and Mazmanian, 2010). The two theories may not be mutually exclusive if both phenomena intersect in the gut, which evolved as a major factor in adaptive immune maturation (Cebra, 1999; Ivanov et al., 2009; Atarashi et al., 2011). A variety of gut-specific adaptive immune mechanisms help maintain microbial–immune homeostasis (Hooper and Macpherson, 2010; Feng and Elson, 2011), yet the effects on host physiology and immunity in distant parts of the body by the resident gut microbiota (Clarke et al., 2010; Lathrop et al., 2011; Maslowski and Mackay, 2011) support the existence of an ancient mutualism.

The animal gut has evolved to tolerate the presence of dense communities of commensal residents; however, the majority of this microbiota remain invisible to the majority of the gut immune system on the basal laminar surface, in part due to physical barriers such as mucus (**Figure 3**). The epithelium has evolved to become a fully functional immune system that expresses various innate receptors and secretes immune-related molecules into the gut lumen (Muller et al., 2005; Duerkop et al., 2009; Marchiando et al., 2010). However, details regarding epithelial recognition and tolerance of luminal bacteria remain largely unknown (Rautava and Walker, 2007; Marchiando et al., 2010), e.g., tolerance to endotoxin is likely governed at the earliest exposure in development (Lotz et al., 2006). It has been suggested that immune systems, as defined currently, evolved first to manage complex symbiotic relationships, while the preservation of "self" became an inevitable adaptation (Bosch and McFall-Ngai, 2011). By this account, mechanisms (later acquired by innate immunity) that govern host–microbial interactions are of ancient phylogenetic origins.

Investigations in *Ciona*, which are focused on the interaction of microbes at the epithelial surface, may help reveal: (1) if certain bacterial species effect barrier function (Lyczak, 2003; Ohland and Macnaughton, 2010); (2) if the degree of microbiome complexity influences epithelial response to infections (Mans et al., 2009); (3) how polarized expression of PRRs help maintain microbial recognition, immune integrity, and homeostasis; (4) how bacteria modulate intercellular tight junctions, which are key to barrier integrity (Turner, 2009); and (5) how secretory pro-inflammatory molecules, like TNF, affect barrier integrity by increased intercellular permeability. Much like mammals, rich microbial communities of distinct phyla and genera are selected in *Ciona* from an exceptionally vast range of choices of dietary and seawater microbes. Various commensal bacteria likely have co-evolved with the species and may have become integral to nutritional acquisition and gut homeostasis. As such, *Ciona* offers a unique opportunity to study and characterize not only the molecular events surrounding gut microbial engagement with mucosal immunity but to define the symbiotic ecosystem required for gut homeostasis and thereby host well-being in a controlled, systematic manner using a novel, tractable model system.

ACKNOWLEDGMENTS

The authors would like to also acknowledge Mr. Alberto Macina (Aqua-Culture of Marine Organisms facility) and Rita Graziano and Franco Iamunno (Electron Microscopy facility) of the Stazione Zoologica Anton Dohrn for providing technical assistance. This work was supported by NIH grant AI23338 (Gary W. Litman) and startup funds from the All Children's Hospital Foundation and USF Sponsored Research (Larry J. Dishaw).

REFERENCES

Abreu, M. T. (2010). Toll-like receptor signalling in the intestinal epithelium: how bacterial recognition shapes intestinal function. *Nat. Rev. Immunol.* 10, 131–144.

Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998). Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system *Nature* 394, 744–751.

Arizza, V., and Parrinello, D. (2009). Inflammatory hemocytes in *Ciona intestinalis* innate immune response. *Invertebrate Surviv. J.* 6, S58–S66.

Arizza, V., Parrinello, D., Cammarata, M., Vazzana, M., Vizzini, A., Giaramita, F. T., and Parrinello, N. (2011). A lytic mechanism based on soluble phospholipases A2 (sPLA2) and beta-galactoside specific lectins is exerted by *Ciona intestinalis* (ascidian) unilocular refractile hemocytes against K562 cell line and mammalian erythrocytes. *Fish Shellfish Immunol.* 30, 1014–1023.

Artis, D. (2008). Epithelial-cell recognition of commensal bacteria and maintenance of immune homeostasis in the gut. *Nat. Rev. Immunol.* 8, 411–420.

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., Cheng, G., Yamasaki, S., Saito, T., Ohba, Y., Taniguchi, T., Takeda, K., Hori, S., Ivanov, I. I., Umesaki, Y., Itoh, K., and Honda, K. (2011). Induction of colonic regulatory T cells by indigenous *Clostridium* species. *Science* 331, 337–341.

Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., Arai, M., Inoue, Y., Shimizu, T., Satoh, N., Rokhsar, D. S., Du Pasquier, L., Kasahara, M., Satake, M., and Nonaka, M. (2003). Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot" *Immunogenetics* 55, 570–581.

Baghdiguian, S., Martinand-Mari, C., and Mangeat, P. (2007). Using Ciona to study developmental programmed cell death. *Semin. Cancer Biol.* 17, 147–153.

Bischoff, S. C. (2011). "Gut health": a new objective in medicine? *BMC Med.* 9, 24. doi:10.1186/1741-7015-9-24

Bonura, A., Vizzini, A., Salerno, G., Parrinello, N., Longo, V., and Colombo, P. (2009). 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.* 46, 2389–2394.

Bosch, T. C., Augustin, R., Anton-Erxleben, F., Fraune, S., Hemmrich, G., Zill, H., Rosenstiel, P., Jacobs, G., Schreiber, S., Leippe, M., Stanisak, M., Grotzinger, J., Jung, S., Podschun, R., Bartels, J., Harder, J., and Schroder, J. M. (2009). Uncovering the evolutionary history of innate immunity: the simple metazoan Hydra uses epithelial cells for host defence. *Dev. Comp. Immunol.* 33, 559–569.

Bosch, T. C., and McFall-Ngai, M. J. (2011). Metaorganisms as the new frontier. *Zoology (Jena)* 114, 185–190.

Burighel, P., and Cloney, R. A. (1977). "Urochordata: Ascidiacea," in *Microscopic Anatomy of Invertebrates, Vol. 15: Hemichordata, Chaetognatha and the Invertebrate Chordates*, eds F. W. Harrison and E. E. Ruppert, 221–347.

Cammarata, M., Arizza, V., Cianciolo, C., Parrinello, D., Vazzana, M., Vizzini, A., Salerno, G., and Parrinello, N. (2008). The prophenoloxidase system is activated during the tunic inflammatory reaction of *Ciona intestinalis*. *Cell Tissue Res.* 333, 481–492.

Canestro, C., Bassham, S., and Postlethwait, J. H. (2003). Seeing chordate evolution through the Ciona genome sequence. *Genome Biol.* 4, 208.

Cannon, J. P., Haire, R. N., and Litman, G. W. (2002). Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. *Nat. Immunol.* 3, 1200–1207.

Caputi, L., Andreakis, N., Mastrototaro, F., Cirino, P., Vassillo, M., and Sordino, P. (2007). Cryptic speciation in a model invertebrate chordate. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9364–9369.

Cebra, J. J. (1999). Influences of microbiota on intestinal immune system development. *Am. J. Clin. Nutr.* 69, 1046S–1051S.

Cebra, J. J., Periwal, S. B., Lee, G., Lee, F., and Shroff, K. E. (1998). Development and maintenance of the gut-associated lymphoid tissue (GALT): the roles of enteric bacteria and viruses. *Dev. Immunol.* 6, 13–18.

Chow, J., and Mazmanian, S. K. (2010). A pathobiont of the microbiota balances host colonization and intestinal inflammation. *Cell Host Microbe* 7, 265–276.

Christiaen, L., Wagner, E., Shi, W., and Levine, M. (2009). Isolation of individual cells and tissues from electroporated sea squirt (*Ciona*) embryos by fluorescence-activated cell sorting (FACS). *Cold Spring Harb. Protoc.* 2009, pdb.prot5349.

Chung, H., and Kasper, D. L. (2010). Microbiota-stimulated immune mechanisms to maintain gut homeostasis. *Curr. Opin. Immunol.* 22, 455–460.

Cirino, P., Toscano, A., Caramiello, D., Macina, A., Miraglia, V., and Monte, A. (2002). Laboratory culture of the ascidian *Ciona intestinalis* (L.): a model system for molecular developmental biology research. *Mar. Mod. Elec. Rec.* [serial online]. Available at: http://www.mbl.edu/html/BB/MMER/CIR/CirTit. html

Clarke, T. B., Davis, K. M., Lysenko, E. S., Zhou, A. Y., Yu, Y., and Weiser, J. N. (2010). Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. *Nat. Med.* 16, 228–231.

Cooper, M. D., and Alder, M. N. (2006). The evolution of adaptive immune systems. *Cell* 124, 815–822.

Davidson, B. (2007). *Ciona intestinalis* as a model for cardiac development. *Semin. Cell Dev. Biol.* 18, 16–26.

Dehal, P., Satou, Y., Campbell, R. K., Chapman, J., Degnan, B., De Tomaso, A., Davidson, B., Di Gregorio, A., Gelpke, M., Goodstein, D. M., Harafuji, N., Hastings, K. E. M., Ho, I., Hotta, K., Huang, W., Kawashima, T., Lamaire, P., Martinez, D., Meinertzhagan, I. A., Necula, S., Nonaka, M., Putnam, N., Rash, S., Salga, H., Sataka, M., Terry, A., Yamada, L., Wang, H.-G., Awazu, S., Azumi, K., Boore, J., Branno, M., Chin-bow, S., DeSantis, R., Doyle, S., Francino, P., Keys, D. N., Haga, S., Hayashi, H., Hino, K., Imai, K. S., Inaba, K., Kano, S., Kobayashi, K., Kobayashi, M., Lee, B.-I., Makabe, K. W., Manohar, C., Matassi, G., Medina, M., Mochizuki, Y., Mount, S., Morishita, T., Miura, S., Nakayama, A., Nishizaka, S., Nomoto, H., Ohta, F., Oishi, K., Rigoutsos, I., Sano, M., Sasaki, A., Sasakura, Y., Shoguchi, E., Shin-i, T., Spagnuolo, A., Stainier, D., Suzuki, M. M., Tassy, O., Takatori, N., Tokuoka, M., Yagi, K., Yoshizaki, F., Wada, S., Zhang, C., Hyatt, P. D., Larimer, F., Detter, C., Doggett, N., Glavina, T., Hawkins, T., Richardson, P., Lucas, S., Kohara, Y., Levine, M., Satoh, N., and Rokhsar, D. S. (2002). The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298, 2157–2160.

Denoeud, F., Henriet, S., Mungpakdee, S., Aury, J. M., Da Silva, C., Brinkmann, H., Mikhaleva, J., Olsen, L. C., Jubin, C., Canestro, C., Bouquet, J. M., Danks, G., Poulain, J., Campsteijn, C., Adamski, M., Cross, I., Yadetie, F., Muffato, M., Louis, A., Butcher, S., Tsagkogeorga, G., Konrad, A., Singh, S., Jensen, M. F., Cong, E. H., Eikeseth-Otteraa, H., Noel, B., Anthouard, V., Porcel, B. M., Kachouri-Lafond, R., Nishino, A., Ugolini, M., Chourrout, P., Nishida, H., Aasland, R., Huzurbazar, S., Westhof, E., Delsuc, F., Lehrach, H., Reinhardt, R., Weissenbach, J., Roy, S. W., Artiguenave, F., Postlethwait, J. H., Manak, J. R., Thompson, E. M., Jaillon, O., Du Pasquier, L., Boudinot, P., Liberles, D. A., Volff, J. N., Philippe, H., Lenhard, B., Crollius, H. R., Wincker, P., and Chourrout, D. (2010). Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 330, 1381–1385.

Di Bella, M. A., and De Leo, G. (2000). Hemocyte migration during inflammatory-like reaction of *Ciona intestinalis* (Tunicata, Ascidiacea). *J Invertebr. Pathol.* 76, 105–111.

Dishaw, L. J., Giacomelli, S., Melillo, D., Zucchetti, I., Haire, R. N., Natale, L., Russo, N. A., De Santis, R., Litman, G. W., and Pinto, M. R. (2011). A role for variable region-containing chitin-binding proteins (VCBPs) in host gut-bacteria interactions. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16747–16752.

Dreyfus, D. H. (1992). Evidence suggesting an evolutionary relationship between transposable elements

and immune system recombination sequences. *Mol. Immunol.* 29, 807–810.

Duerkop, B. A., Vaishnava, S., and Hooper, L. V. (2009). Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* 31, 368–376.

Edelman, S. M., and Kasper, D. L. (2008). Symbiotic commensal bacteria direct maturation of the host immune system. *Curr. Opin. Gastroenterol.* 24, 720–724.

Ermak, T. H. (1981). A comparison of cell proliferation patterns in the digestive track of ascidians. *J. Exp. Zool.* 217, 325–339.

Fedders, H., and Leippe, M. (2008). A reverse search for antimicrobial peptides in *Ciona intestinalis*: identification of a gene family expressed in hemocytes and evaluation of activity. *Dev. Comp. Immunol.* 32, 286–298.

Feng, T., and Elson, C. O. (2011). Adaptive immunity in the host-microbiota dialog. *Mucosal Immunol.* 4, 15–21.

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

Fujimura, K. E., Slusher, N. A., Cabana, M. D., and Lynch, S. V. (2010). Role of the gut microbiota in defining human health. *Expert Rev. Anti Infect. Ther.* 8, 435–454.

Fujita, T., Endo, Y., and Nonaka, M. (2004). Primitive complement system–recognition and activation. *Mol. Immunol.* 41, 103–111.

Holland, L. Z., Albalat, R., Azumi, K., Benito-Gutierrez, E., Blow, M. J., Bronner-Fraser, M., Brunet, F., Butts, T., Candiani, S., Dishaw, L. J., Ferrier, D. E., Garcia-Fernandez, J., Gibson-Brown, J. J., Gissi, C., Godzik, A., Hallbook, F., Hirose, D., Hosomichi, K., Ikuta, T., Inoko, H., Kasahara, M., Kasamatsu, J., Kawashima, T., Kimura, A., Kobayashi, M., Kozmik, Z., Kubokawa, K., Laudet, V., Litman, G. W., McHardy, A. C., Meulemans, D., Nonaka, M., Olinski, R. P., Pancer, Z., Pennacchio, L. A., Pestarino, M., Rast, J. P., Rigoutsos, I., Robinson-Rechavi, M., Roch, G., Saiga, H., Sasakura, Y., Satake, M., Satou, Y., Schubert, M., Sherwood, N., Shiina, T., Takatori, N., Tello, J., Vopalensky, P., Wada, S., Xu, A., Ye, Y., Yoshida, K., Yoshizaki, F., Yu, J. K., Zhang, Q., Zmasek, C. M., de Jong, P. J., Osoegawa, K., Putnam, N. H., Rokhsar, D. S., Satoh, N., and Holland, P. W. (2008). The amphioxus genome illuminates vertebrate origins and cephalochordate biology. *Genome Res.* 18, 1100–1111.

Hooper, L. V., and Gordon, J. I. (2001). Commensal host-bacterial relationships in the gut. *Science* 292, 1115–1118.

Hooper, L. V., and Macpherson, A. J. (2010). Immune adaptations that maintain homeostasis with the intestinal microbiota. *Nat. Rev. Immunol.* 10, 159–169.

Hughes, A. L., and Friedman, R. (2005). Loss of ancestral genes in the genomic evolution of *Ciona intestinalis*. *Evol. Dev.* 7, 196–200.

Ikuta, T., Satoh, N., and Saiga, H. (2010). Limited functions of Hox genes in the larval development of the ascidian *Ciona intestinalis*. *Development* 137, 1505–1513.

Ivanov, I. I., Atarashi, K., Manel, N., Brodie, E. L., Shima, T., Karaoz, U., Wei, D., Goldfarb, K. C., Santee, C. A., Lynch, S. V., Tanoue, T., Imaoka, A., Itoh, K., Takeda, K., Umesaki, Y., Honda, K., and Littman, D. R. (2009). Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* 139, 485–498.

Joly, J., Kano, S., Matsuoka, T., Auger, H., Hirayama, K., Satoh, N., Awazu, S., Legendre, L., and Sasakura, Y. (2007). Culture of *Ciona intestinalis* in closed systems. *Dev. Dyn.* 236, 1832–1840.

Kanther, M., and Rawls, J. F. (2010). Host-microbe interactions in the developing zebrafish. *Curr. Opin. Immunol.* 22, 10–19.

Katz, M. J. (1983). Comparative anatomy of the tunicate tadpole, *Ciona intestinalis*. *Biol. Bull.* 164, 1–27.

Kubo, A., Suzuki, N., Yuan, X., Nakai, K., Satoh, N., Imai, K. S., and Satou, Y. (2010). Genomic cis-regulatory networks in the early *Ciona intestinalis* embryo. *Development* 137, 1613–1623.

Lathrop, S. K., Bloom, S. M., Rao, S. M., Nutsch, K., Lio, C. W., Santacruz, N., Peterson, D. A., Stappenbeck, T. S., and Hsieh, C. S. (2011). Peripheral education of the immune system by colonic commensal microbiota. *Nature* 478, 250–254.

Lavelle, E. C., Murphy, C., O'Neill, L. A., and Creagh, E. M. (2010). The role of TLRs, NLRs, and RLRs in mucosal innate immunity and homeostasis. *Mucosal Immunol.* 3, 17–28.

Lee, Y. K., and Mazmanian, S. K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system? *Science* 330, 1768–1773.

Lemaire, P., Smith, W. C., and Nishida, H. (2008). Ascidians and the plasticity of the chordate developmental program. *Curr. Biol.* 18, R620–R631.

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

Litman, G. W., Rast, J. P., and Fugmann, S. D. (2010). Origins of vertebrate adaptive immunity. *Nat. Rev. Immunol.* 10, 543–553.

Lotz, M., Gutle, D., Walther, S., Menard, S., Bogdan, C., and Hornef, M. W. (2006). Postnatal acquisition of endotoxin tolerance in intestinal epithelial cells. *J. Exp. Med.* 203, 973–984.

Lyczak, J. B. (2003). Commensal bacteria increase invasion of intestinal epithelium by *Salmonella enterica* serovar *Typhi*. *Infect. Immun.* 71, 6610–6614.

Maldonado-Contreras, A. L., and McCormick, B. A. (2011). Intestinal epithelial cells and their role in innate mucosal immunity. *Cell Tissue Res.* 343, 5–12.

Mans, J. J., von Lackum, K., Dorsey, C., Willis, S., Wallet, S. M., Baker, H. V., Lamont, R. J., and Handfield, M. (2009). The degree of microbiome complexity influences the epithelial response to infection. *BMC Genomics* 10, 380. doi:10.1186/1471-2164-10-380

Marchiando, A. M., Graham, W. V., and Turner, J. R. (2010). Epithelial barriers in homeostasis and disease. *Annu. Rev. Pathol.* 5, 119–144.

Marino, R., Kimura, Y., De Santis, R., Lambris, J. D., and Pinto, M. R. (2002). Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian *Ciona intestinalis*. *Immunogenetics* 53, 1055–1064.

Maslowski, K. M., and Mackay, C. R. (2011). Diet, gut microbiota and immune responses. *Nat. Immunol.* 12, 5–9.

Matsunaga, T., and Rahman, A. (1998). What brought the adaptive immune system to vertebrates? – The jaw hypothesis and the seahorse. *Immunol. Rev.* 166, 177–186.

Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., and Kasper, D. L. (2005). An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell* 122, 107–118.

McFall-Ngai, M. (2007). Adaptive immunity: care for the community. *Nature* 445, 153–153.

Meinertzhagen, I. A., and Okamura, Y. (2001). The larval ascidian nervous system: the chordate brain from its small beginnings. *Trends Neurosci.* 24, 401–410.

Melillo, D., Sfyroera, G., De Santis, R., Graziano, R., Marino, R., Lambris, J. D., and Pinto, M. R. (2006). First identification of a chemotactic receptor in an invertebrate species: structural and functional characterization of *Ciona intestinalis* C3a receptor. *J. Immunol.* 177, 4132–4140.

Messier-Solek, C., Buckley, K. M., and Rast, J. P. (2010). Highly diversified innate receptor systems and new forms of animal immunity. *Semin. Immunol.* 22, 39–47.

Millar, R. H. (1953). *Ciona*, Liverpool: University Press of Liverpool, 1–122.

Muller, C. A., Autenrieth, I. B., and Peschel, A. (2005). Innate defenses of the intestinal epithelial barrier. *Cell. Mol. Life Sci.* 62, 1297–1307.

Nell, S., Suerbaum, S., and Josenhans, C. (2010). The impact of the microbiota on the pathogenesis of IBD: lessons from mouse infection models. *Nat. Rev. Microbiol.* 8, 564–577.

Nishida, H., Satoh, N., and Hirose, E. (2010). More diversity and more convergence in tunicate biology. *Zool. Sci.* 27, 67–68.

Ohland, C. L., and Macnaughton, W. K. (2010). Probiotic bacteria and intestinal epithelial barrier function. *Am. J. Physiol. Gastrointest. Liver Physiol.* 298, G807–G819.

Parrinello, N. (2009). Focusing on *Ciona intestinalis* (tunicate) innate immune system. Evolutionary implications. *Invertebrate Surviv. J.* 6, S46–S57.

Parrinello, N., Arizza, V., Cammarata, M., Giaramita, F. T., Pergolizzi, M., Vazzana, M., Vizzini, A., and Parrinello, D. (2007). Inducible lectins with galectin properties and human IL1alpha epitopes opsonize yeast during the inflammatory response of the ascidian *Ciona intestinalis*. *Cell Tissue Res.* 329, 379–390.

Parrinello, N., Cammarata, M., and Arizza, V. (1996). Univacuolar refractile hemocytes from the tunicate *Ciona intestinalis* are cytotoxic for mammalian erythrocytes in vitro. *Biol. Bull.* 190, 418–425.

Parrinello, N., Vizzini, A., Arizza, V., Salerno, G., Parrinello, D., Cammarata, M., Giaramita, F. T., and Vazzana, M. (2008). Enhanced expression of a cloned and sequenced *Ciona intestinalis* TNFalpha-like (CiTNF alpha) gene during the LPS-induced inflammatory response. *Cell Tissue Res.* 334, 305–317.

Parrinello, N., Vizzini, A., Salerno, G., Sanfratello, M. A., Cammarata, M.,

Arizza, V., Vazzana, M., and Parrinello, D. (2010). Inflamed adult pharynx tissues and swimming larva of *Ciona intestinalis* share CiTNFalpha-producing cells. *Cell Tissue Res.* 341, 299–311.

Pinto, M. R., Chinnici, C. M., Kimura, Y., Melillo, D., Marino, R., Spruce, L. A., De Santis, R., Parrinello, N., and Lambris, J. D. (2003). CiC3-1a-mediated chemotaxis in the deuterostome invertebrate *Ciona intestinalis* (Urochordata). *J. Immunol.* 171, 5521–5528.

Rast, J. P., and Messier-Solek, C. (2008). Marine invertebrate genome sequences and our evolving understanding of animal immunity. *Biol. Bull.* 214, 274–283.

Rast, J. P., Smith, L. C., Loza-Coll, M., Hibino, T., and Litman, G. W. (2006). Genomic insights into the immune system of the sea urchin. *Science* 314, 952–956.

Rautava, S., and Walker, W. A. (2007). Commensal bacteria and epithelial cross talk in the developing intestine. *Curr. Gastroenterol. Rep.* 9, 385–392.

Rescigno, M., and Di, Sabatino, A. (2009). Dendritic cells in intestinal homeostasis and disease. *J. Clin. Invest.* 119, 2441–2450.

Round, J. L., and Mazmanian, S. K. (2009). The gut microbiota shapes intestinal immune responses during health and disease. *Nat. Rev. Immunol.* 9, 313–323.

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

Sasakura, Y., Inaba, K., Satoh, N., Kondo, M., and Akasaka, K. (2009). *Ciona intestinalis* and *Oxycomanthus japonicus*, representatives of marine invertebrates. *Exp. Anim.* 58, 459–469.

Sasakura, Y., Suzuki, M. M., Hozumi, A., Inaba, K., and Satoh, N. (2010). Maternal factor-mediated epigenetic gene silencing in the ascidian *Ciona intestinalis*. *Mol. Genet. Genomics* 283, 99–110.

Satoh, N., and Levine, M. (2005). Surfing with the tunicates into the post-genome era. *Genes Dev.* 19, 2407–2411.

Savage, D. C. (1977). Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* 31, 107–133.

Sekirov, I., Russell, S. L., Antunes, L. C., and Finlay, B. B. (2010). Gut microbiota in health and disease. *Physiol. Rev.* 90, 859–904.

Shi, W., Levine, M., and Davidson, B. (2005). Unraveling genomic regulatory networks in the simple chordate, *Ciona intestinalis*. *Genome Res.* 15, 1668–1674.

Shida, K., Terajima, D., Uchino, R., Ikawa, S., Ikeda, M., Asano, K., Watanabe, T., Azumi, K., Nonaka, M., Satou, Y., Satoh, N., Satake, M., Kawazoe, Y., and Kasuya, A. (2003). Hemocytes of *Ciona intestinalis* express multiple genes involved in innate immune host defense. *Biochem. Biophys. Res. Commun.* 302, 207–218.

Skjoedt, M. O., Palarasah, Y., Rasmussen, K., Vitved, L., Salomonsen, J., Kliem, A., Hansen, S., Koch, C., and Skjodt, K. (2010). 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.* 34, 59–68.

Smith, V. J., and Peddie, C. M. (1992). Cell cooperation during host defense in the solitary tunicate *Ciona intestinalis* (L). *Biol. Bull.* 183, 211–219.

Terajima, D., Shida, K., Takada, N., Kasuya, A., Rokhsar, D., Satoh, N., Satake, M., and Wang, H. G. (2003). Identification of candidate genes encoding the core components of the cell death machinery in the *Ciona intestinalis* genome. *Cell Death Differ.* 10, 749–753.

Turner, J. R. (2009). Intestinal mucosal barrier function in health and disease. *Nat. Rev. Immunol.* 9, 799–809.

Usharauli, D. (2010). Chronic infection and the origin of adaptive immune system. *Med. Hypotheses* 75, 241–243.

Willing, B. P., Antunes, L. C., Keeney, K. M., Ferreira, R. B., and Finlay, B. B. (2011). Harvesting the biological potential of the human gut microbiome. *Bioessays* 33, 414–418.

Yonge, C. M. (1925). Studies on the comparative physiology of digestion. III. Secretion, digestion, and assimilation in the gut of *Ciona intestinalis*. *J. Exp. Biol.* 2, 373–388.

Zucchetti, I., De Santis, R., Grusea, S., Pontarotti, P., and Du Pasquier, L. (2009). Origin and evolution of the vertebrate leukocyte receptors: the lesson from tunicates. *Immunogenetics* 61, 463–481.

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.

Received: 17 January 2012; accepted: 12 April 2012; published online: 03 May 2012.

Citation: Dishaw LJ, Flores-Torres JA, Mueller MG, Karrer CR, Skapura DP, Melillo D, Zucchetti I, De Santis R, Pinto MR and Litman GW (2012) A basal chordate model for studies of gut microbial immune interactions. Front. Immun. ***3****:96. doi: 10.3389/fimmu.2012.00096*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

ORIGINAL RESEARCH ARTICLE
published: 12 October 2012
doi: 10.3389/fimmu.2012.00314

Dynamic evolution of the LPS-detoxifying enzyme intestinal alkaline phosphatase in zebrafish and other vertebrates

Ye Yang[1], Anica M. Wandler[1†], John H. Postlethwait[2]* and Karen Guillemin[1]*

[1] Institute of Molecular Biology, University of Oregon, Eugene, OR, USA
[2] Institute of Neuroscience, University of Oregon, Eugene, OR, USA

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Tatsuya Ota, The Graduate University for Advanced Studies, Japan
Jeffrey A. Yoder, North Carolina State University, USA

***Correspondence:**
John H. Postlethwait, Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA.
e-mail: jpostle@uoneuro.uoregon.edu;
Karen Guillemin, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403, USA.
e-mail: guillemin@molbio.uoregon.edu

[†]***Current Address:***
Anica M. Wandler, Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA.

Alkaline phosphatases (Alps) are well-studied enzymes that remove phosphates from a variety of substrates. Alps function in diverse biological processes, including modulating host-bacterial interactions by dephosphorylating the Gram-negative bacterial cell wall component lipopolysaccharide (LPS). In animals, Alps are encoded by multiple genes characterized by either ubiquitous expression (named *Alpl*s for their liver expression, but a key to proper bone mineralization), or their tissue-specific expression, for example in the intestine (*Alpi*). We previously characterized a zebrafish *alpi* gene (renamed here *alpi.1*) that is regulated by Myd88-dependent innate immune signaling and that is required to prevent a host's excessive inflammatory reactions to its resident microbiota. Here we report the characterization of two new *alp* genes in zebrafish, *alpi.2* and *alp3*. To understand their origins, we investigated the phylogenetic history of *Alp* genes in animals. We find that vertebrate *Alp* genes are organized in three clades with one of these clades missing from the mammals. We present evidence that these three clades originated during the two vertebrate genome duplications. We show that *alpl* is ubiquitously expressed in zebrafish, as it is in mammals, whereas the other three *alps* are specific to the intestine. Our phylogenetic analysis reveals that in contrast to *Alpl*, which has been stably maintained as a single gene throughout the vertebrates, the *Alpi*s have been lost and duplicated multiple times independently in vertebrate lineages, likely reflecting the rapid and dynamic evolution of vertebrate gut morphologies, driven by changes in bacterial associations and diet.

Keywords: zebrafish, intestinal alkaline phosphatase, vertebrate, evolution, microbiota

INTRODUCTION

Alkaline phosphatases (Alps) are a superfamily of metalloenzymes that catalyze the hydrolytic removal of phosphate from a variety of molecules (Millán, 2006). Alps have been extensively studied biochemically, but the full spectrum of their biological functions is not known. In animals, Alps are encoded by multiple genes that can be classified by their expression patterns into two general groups, the tissue-non-specific *Alp*s (known as *Alpl*, for their liver expression, also known as *TNAP* for tissue non-specific alkaline phosphatase), and the tissue-specific *Alp*s [placental *Alp* (*Alpp*), intestinal *Alp* (*Alpi*), etc.]. The best studied biological function of Alps is the role of mammalian Alpl in osteogenesis by promoting bone mineralization, as demonstrated by the hypophosphatasia that results from *ALPL* deficiency in humans and mice (Whyte, 2010). More recently, the *Alpi*s have been implicated in mediating host-bacterial interactions through their ability to dephosphorylate lipid A of the Gram-negative bacterial cell wall component lipopolysaccharide (LPS; Lalles, 2010).

In mammals, Alpi is expressed by intestinal epithelial cells (IECs) and is enriched in vesicles that are actively released from IEC microvillar tips into the intestinal lumen (McConnell et al., 2009; Shifrin et al., 2012). Thus Alpi is located at the interface between the intestinal tissue, the ingesta and the vast microbiota, which suggests its involvement in a variety of biological processes. Recent studies have discovered that Alpi regulates metabolism by controlling the uptake of nutrients such as lipids (Narisawa et al., 2003; Lynes et al., 2011) and calcium (Brun et al., 2012), affects gut physiology by maintaining protective surface microclimate pH in the duodenum (Akiba et al., 2007; Mizumori et al., 2009), and impacts innate immunity by modulating bacterial LPS-induced inflammation (Poelstra et al., 1997b; Bates et al., 2007; Campbell et al., 2010).

Lipopolysaccharide, also commonly referred to as endotoxin, is a component of Gram-negative bacterial cell walls and is abundantly present in the intestinal lumen of animals. LPS is a classic microbial associated molecular pattern (MAMP) and potent inducer of innate immune signaling in both vertebrates and invertebrates (Beutler and Rietschel, 2003). In mammals, LPS binds specifically to a complex consisting of Toll-like receptor 4 (TLR4) and MD-2 through two phosphate groups of its lipid A moiety (Kim et al., 2007), and induces innate immune responses by activating two distinct pathways, namely NF-kB (through MyD88-dependent and independent pathways) and IRF-3 (through TRIF/TRAM). Although the specifics of LPS binding do not seem to be conserved between mammals and teleosts (Sullivan et al., 2009), this MAMP elicits similar pro-inflammatory

responses through a Myd88-dependent mechanism in zebrafish (Bates et al., 2007) as in mammals.

Alps have been shown to remove the lipid A phosphates of LPS at physiological pH levels (Poelstra et al., 1997a,b), thereby reducing its affinity for TLR4 and, correspondingly, its endotoxic properties. Our studies in zebrafish larvae (Bates et al., 2006, 2007) demonstrated the functional significance of Alpi's LPS dephosphorylating activity in the intestine, and showed that this gene plays an integral role in modulating innate immune responses in the gut through a Myd88-dependent negative feedback loop. We found that LPS incubation as well as Gram-negative bacterium inoculation upregulated zebrafish Alpi, a process that required MyD88. We showed that Alpi functioned in the detoxification of LPS because treatment with the Alpi-specific inhibitor L-phenylalanine or *alpi.1*-specific morpholino rendered zebrafish more sensitive in an LPS killing assay, whereas fish were resistant to LPS pretreated with calf Alpi. Furthermore, zebrafish with reduced Alpi activity exhibited elevated levels of pro-inflammatory cytokines and intestinal neutrophil influx, both Myd88-dependent processes. However, when Alpi deficient zebrafish were derived germ-free, removing microbiota-associated LPS, no excess neutrophil influx was observed. Collectively these results show that zebrafish intestinal colonization by Gram-negative bacteria upregulates the host enzyme Alpi, which functions to reduce host inflammatory responses to resident microbiota.

The anti-inflammatory function of Alpi is supported by many other observations from mammalian systems. Cell culture studies showed that in IECs (i.e., HT-29, T84, and IEC-6) overexpressing Alpi, LPS-activated NF-κB nuclear translocation was significantly inhibited (Goldberg et al., 2008). At the whole animal level, administration of bovine Alpi proved to reduce local/systemic inflammation and improve tissue morphology in the mouse polymicrobial sepsis model induced by cecal ligation and puncture (Van Veen et al., 2005), in the rat liver ischemia–reperfusion model (Van Veen et al., 2006), in the murine chronic colitis model induced by dextran sulfate sodium (DSS; Tuin et al., 2009; Campbell et al., 2010; Ramasamy et al., 2011) or TNBS (Martinez-Moya et al., 2012), and in the neonatal necrotizing enterocolitis rat model (Rentea et al., 2012). In clinical trials in humans, exogenous Alpi exerted protective anti-inflammatory effects on patients after cardiopulmonary surgery (Kats et al., 2009), patients with moderate to severe ulcerative colitis (Lukas et al., 2010), and patients with severe sepsis or septic shock (Heemskerk et al., 2009; Pickkers et al., 2012). Collectively, these findings confirm the importance of Alpi as an innate immune regulator, locally and systemically. LPS-detoxification by Alpi is also confirmed in cells (Goldberg et al., 2008) and animals (Beumer et al., 2003). The anti-inflammatory role of Alps is not restricted to the intestinal type, since Alps from other sources (e.g., placental Alp) protected mice against *Escherichia coli*-induced sepsis (Verweij et al., 2004).

Alp's role in lipid A dephosphorylation and modulation of LPS recognition appears to be an ancient function for this family of enzymes, as demonstrated by recent work in the Hawaiian bobtail squid *Euprymna scolopes* (Rader et al., 2012). The squid acquires its Gram-negative bacterial symbiont *Vibrio fischeri* from the environment at the juvenile stage and thereafter enters a lifelong partnership with the luminous marine microbe. Rader et al. characterized two *E. scolopes* Alps (EsAlps), which are closely related to other mollusk Alps. EsAlp is highly active at the lumina of crypt spaces where the bacteria reside. Interestingly, EsAlp activity remained at low levels before and during the lipid A-induced tissue regression at the initial animal-bacteria contact. This enzyme regulation at the early stage proved important for the formation of the symbiotic relationship as demonstrated by two observations: (i) that inhibition of Alp by levamisole compromised the normal colonization of the symbiont and (ii) that *V. fischeri* lipid A pretreated by Alp failed to cause early stage apoptosis that is necessary for persistent colonization of *V. fischeri*. After colonization, however, the squid continuously adjusts EsAlp activity in accordance with the diel rhythm of bacterial population density, i.e., high at dusk and low at dawn, a pattern the authors suggest is strategically governed to render the animal insensitive to lipid A signaling by Alp dephosphorylation of LPS, and therefore protect the animal from excessive inflammation and tissue damage. Taken together, the data from *E. scolopes* presented an elegant example of the conserved role of Alps in tuning host immune recognition of LPS.

All animals live in close associations with microbial communities. Most frequently, the vast majority of these microbes reside in the digestive tract, where they assist the host in the breakdown of ingested food. Gut microbiota are highly species-specific, based on host phylogeny, diet, and digestive tract morphology (Ley et al., 2008a,b). Further microbial community specialization occurs along the length of the gut. Considering the conserved role of Alps in host-bacteria interactions, we imagine that intestinal Alps have been under continuous selective pressure to accommodate changing host-microbe interactions. For example, evolution of host digestive tract physiologies, driven by dietary changes, could spur *Alpi* gene duplication and diversification to buffer host inflammatory responses during the acquisition and compartmentalization of novel bacterial communities that facilitate food digestion.

In this report, we characterized two new *alp* genes in zebrafish and investigated their evolutionary history through the lens of the animal *Alp* gene phylogenies. We find that unlike the *Alpl* clade, the other *Alp* genes, which are frequently intestinally expressed, have been dynamically lost and duplicated throughout animal lineages, consistent with dynamically changing host-microbe interactions. These results suggest that *Alp* gene evolution has played an important role in shaping innate immune response to the intestinal microbiota.

RESULTS

THE FOUR ZEBRAFISH *alp* GENES ARE DISTRIBUTED AMONG THREE VERTEBRATE *Alp* GENE CLADES

We had previously characterized two zebrafish *alp* genes, which we called *alp* (accession number NM_201007.1) on chromosome 11 and *alpi* (accession number NM_001014353.1) on chromosome 22, and had shown that the former gene is ubiquitously expressed and the latter is intestinal-specific (Bates et al., 2007). BLAST searches against the subsequent refinement of the zebrafish genome sequence revealed two additional *alp*-related genes: *zgc:110409* (accession number NM_001025188.1) on chromosome 22, and the most recently described *alp*-related gene (accession number XM_003201677.1) on unassembled Scaffold

Zv9 NA903 that Ensembl annotated as "*alpl*." All four genes likely encode enzymatically active Alps based on their conserved Alp motifs. Protein sequence alignment against human ALPL and ALPI showed that active site residues and proposed metal-binding residues are invariant in the proteins encoded by the two human genes and four zebrafish genes (**Figure 1**).

To better understand the identity and origin of the four zebrafish *alp*-related genes, we investigated their evolutionary history. Phylogenetic analysis of vertebrate Alp protein sequences rooted on non-vertebrate chordate sequences revealed three distinct clades of *Alp* genes (**Figure 2**). The first of these clades, which we call Alp1, contains the human tissue non-specific gene *ALPL* and gene sequences from both lineages of bony vertebrates, the Sarcopterygii (lobe fin fish, including the basally diverging coelacanth and lungfish as well as tetrapods) and the Actinopterygii (ray fin fish, including the basally diverging gar as well as teleosts). The second clade, Alp2, contains the human intestinal-specific gene *ALPI*, the two human placental Alps, ALPP and ALPP2, as well as representative sequences from both lobefins (e.g., mammals) and ray fins (e.g., teleosts). The third clade, Alp3, contains genes exclusive to the fishes, including ray fins and the basally diverging lobe fin, the coelacanth.

Examination of the vertebrate *Alp* gene phylogeny suggests the following hypothesis for the origin of this gene family. A single *Alp* gene in a pre-vertebrate chordate duplicated to form four copies after the first and second vertebrate genome duplications (VGD1 and VGD2); we can call these genes *Alp1*, *Alp2*, *Alp3*, all of which persisted in some lineages, and *Alp4*, which was subsequently lost before the divergence of rayfins and lobefins. Rayfins and lobefins both retained *Alp1*, which became *Alpl*, and preserved *Alp2*, which experienced several lineage-specific tandem duplication events to become *Alpi-like* genes. *Alp3* persisted in rayfins and a basally diverging lobefin, the coelacanth, but was lost from crown lobefins, the tetrapods.

Analysis of the four zebrafish *alp* genes within this phylogeny revealed that (i) the chromosome 11 "*alp*" is a genuine ortholog of the human tissue non-specific gene *ALPL*; (ii) the chromosome 22 "*alpi*" and the neighboring "*zgc: 110409*" are tandem duplicates derived from the ancestral *alpi* gene and represent coorthologs of human *ALPI*; and (iii) the new Scaffold Zv9 NA903 *alp* annotated in Ensembl as "*alpl*" belongs to the Alp3 clade maintained in teleosts but lost in tetrapods. Based on these findings, we developed a new nomenclature for the four genes, i.e., the current "*alp*" is renamed *alpl*, "*alpi*" is *alpi.1*, "*zgc:110409*" becomes *alpi.2*, and "*alpl*" is *alp3*. We cloned and sequenced the complete coding DNA sequences of the four zebrafish *alp* genes and submitted to GenBank: *alpl* (accession number JX847415), *alpi.1* (accession number JX847416), *alpi.2* (accession number JX847417), and *alp3* (accession number JX847418).

THE VERTEBRATE *Alpl* AND *Alpi* GENES AROSE DURING THE VERTEBRATE GENOME DUPLICATIONS

At least two alternative models can explain the origin of multiple Alp paralogs in vertebrates. Under one hypothesis, the *Alpi* and *Alpl* genes arose from gene duplication in pre-vertebrate ancestors and were inherited by the vertebrates, but under an alternative hypothesis, vertebrate *Alp* paralogs arose as ohnologs (paralogs derived from genome duplication) in the two rounds of whole genome duplication VGD1 and VGD2 that occurred at the base of the vertebrate radiation (Dehal and Boore, 2005). These two hypotheses and more complicated alternatives can be winnowed down by examining phylogenetic trees that include both invertebrate and vertebrate taxa. **Figure 3A** shows that, while non-vertebrate chordates (sea squirts *Ciona intestinalis* and *Ciona savignyi* and amphioxus) have multiple *Alp* genes, these diverge basal to the vertebrates, as would be expected if the vertebrate *Alp* genes arose after the divergence of vertebrates from non-vertebrate chordates. Likewise, except for one fly (*Drosophila melanogaster*) sequence, *Alp* genes of insects diverge basal to the chordates, while the non-bilaterian Cnidarian sequences (the sea anemone *Nematostella vectensis* and the hydra *Hydra magnipapillata*) root the tree. The preponderance of evidence leads to the conclusion that the *Alp* paralogs arose in the vertebrate genome duplication events.

Analysis of conserved syntenies comparing the human genome to the genomes of coelacanth and gar supports this model for *Alp* gene history. Located near *ALP3* in the coelacanth genome are the genes *EIF4G1* and *ECE2*, which are paralogs of genes located near the human *ALPL* and *ALPI* genes, namely *EIF4G3 – ECE1* and *EIF4E2 – ECEL1*, respectively (**Figure 3B**). At least five genes immediately adjacent to coelacanth *ALP3* are adjacent in the human genome, although inversions have altered gene order. These three paralogons containing *ALP – ECE – EIF4* genes likely resulted from two rounds of duplication, most parsimoniously explained as happening in the first and second vertebrate genome duplication events VGD1 and VGD2 (Dehal and Boore, 2005). The Synteny Database (Catchen et al., 2009) identifies four chromosome segments containing *ECE* paralogs, two with *ALP* gene neighbors and two without, including regions on Hsa3 and Hsa7 (**Figure 3C**). These data are consistent with the hypothesis that an ancestral chordate chromosome segment contained *ALP*, *ECE*, and *EIF4* genes became four paralogons after VGD2, followed in the human lineage by the loss of two of the *ALP* paralogs from the chromosome segments that eventually became the relevant part of Hsa3 and Hsa7 and the diversification of the eventual human chromosome 1 (Hsa1) and Hsa2 genes as the *ALPI-like* (*Alp2*) and *ALPL-like* (*Alp1*) genes of today's vertebrates.

THREE OF THE ZEBRAFISH *alp* GENES SHOW ENRICHED EXPRESSION IN THE INTESTINE

We further explored the four zebrafish *alp* genes by investigating their tissue expression patterns. We used semi-quantitative reverse transcription PCR to estimate the abundance of the transcripts in intestinal tissue ("I") versus the rest of the body (referred to as carcass, "C"). Transcript levels of *alpl* were abundant in the carcass as well as the intestine (**Figure 4A**). In contrast, transcripts of the other three genes were enriched (*alpi.1* and *alpi.2*) or exclusively expressed (*alp3*) in intestinal tissue (**Figure 4A**). We next performed *in situ* hybridization with gene-specific RNA probes to further examine the expression patterns of the four *alp* transcripts. Consistent with our previous analysis (Bates et al., 2007), we observed that *alpl* was diffusely expressed in many tissues (**Figure 4B**). Also, as we showed previously (Bates et al., 2007), a*lpi.1* was highly expressed in the intestine (**Figure 4C**).

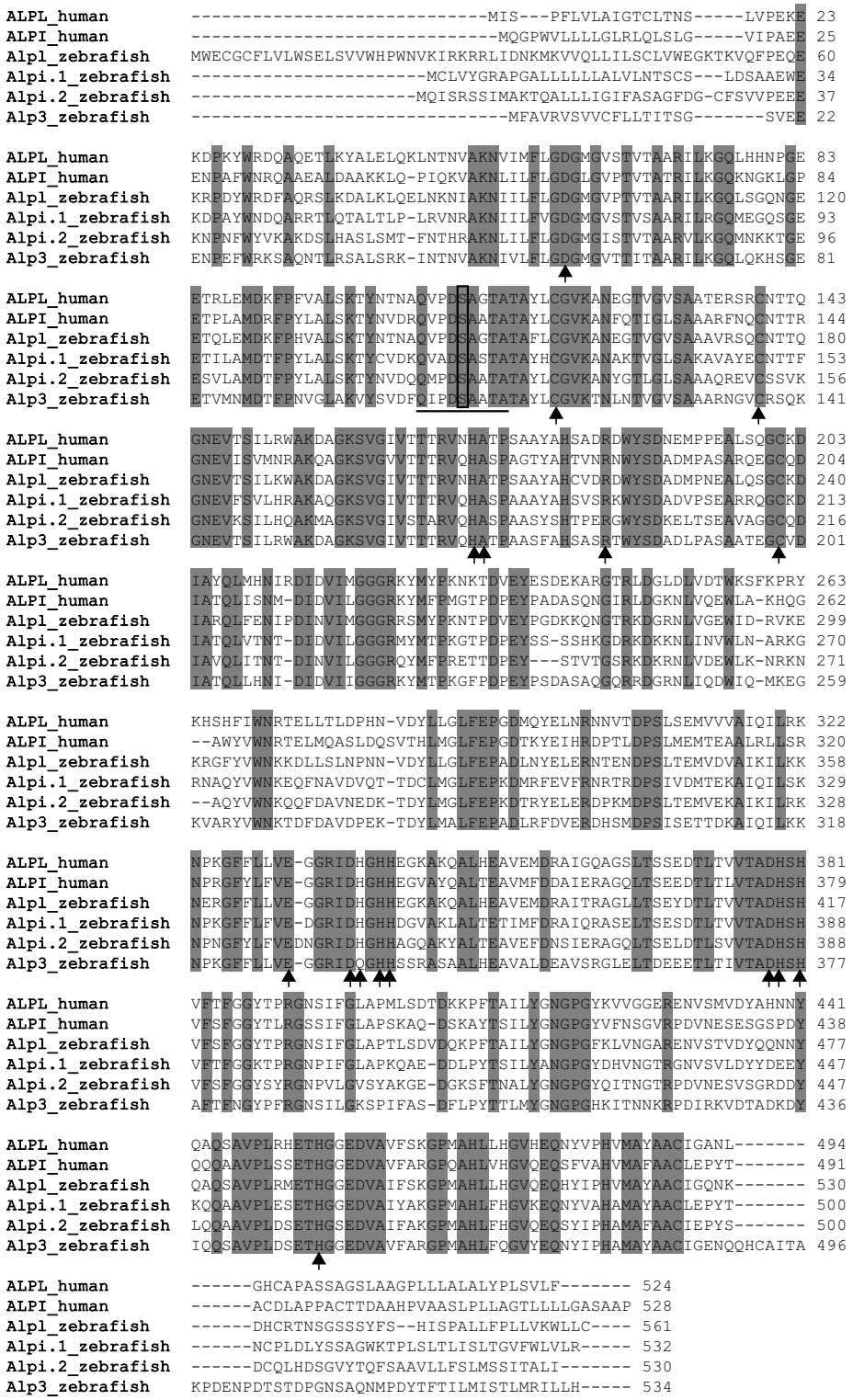


FIGURE 1 | ClustalW Sequence alignment of human and zebrafish Alps. Identical amino acids are shaded gray, putative metal-binding sites are indicated by arrows, active sites are underlined, and the conserved serine required for enzyme activity is boxed. The proteins used in the alignment are human ALPL (ENSP00000363973) and ALPI (ENSP00000295463); zebrafish Alpl (ENSDARP00000117214), Alpi.1 (ENSDARP00000016216), Alpi.2 (ENSDARP00000070354), and Alp3 (ENSDARP00000019098).

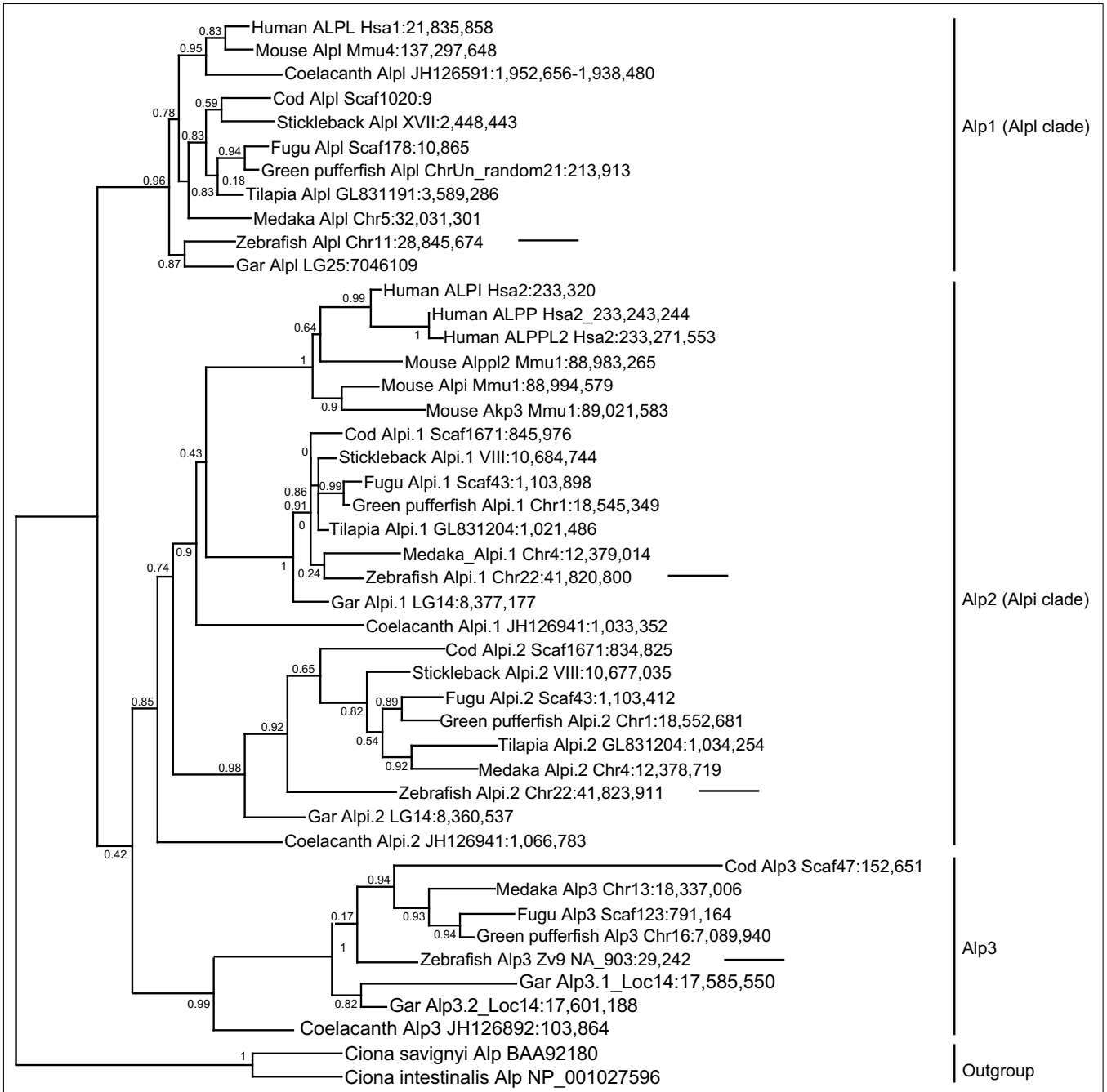


FIGURE 2 | Diversification of vertebrate Alp protein sequences. Sequences are indicated on this maximum likelihood tree with the species name, gene name, and genomic location. For example, the taxon "Human ALPL Hsa1:21,835,858" indicates the start of the human *ALPL* gene encoding this protein on human chromosome 1 at nucleotide position 21,835,858 according to Ensembl Release 68 (July 2012) and Pre-Ensembl Release 66 (April 2012). The tree has three major clades [Alp1 (Alpl clade), Alp2 (Alpi clade), and Alp3]. Arrows point to the four zebrafish Alp genes.

We also observed high levels of intestinal-specific expression of *alpi.2* (**Figures 4C,D**), indicating that the tissue-specific expression of this *alpi* coortholog has been maintained. Finally, we observed intestinal-specific expression of the teleost-specific *alp3* gene (**Figures 4C,E**), suggesting that intestinal-specific expression of *alp* genes is an ancestral trait possessed by the single *Alp* gene found in non-vertebrate chordates before the VGD1 and VGD2 events about 650 million years ago (Hedges et al., 2006; Braasch and Postlethwait, 2012).

MAMMALIAN *Alpi* GENES HAVE UNDERGONE RAPID AND DYNAMIC EVOLUTION

Our phylogenetic analysis of the Alp gene family (**Figure 1**) suggested a dynamic evolutionary history of clades containing the

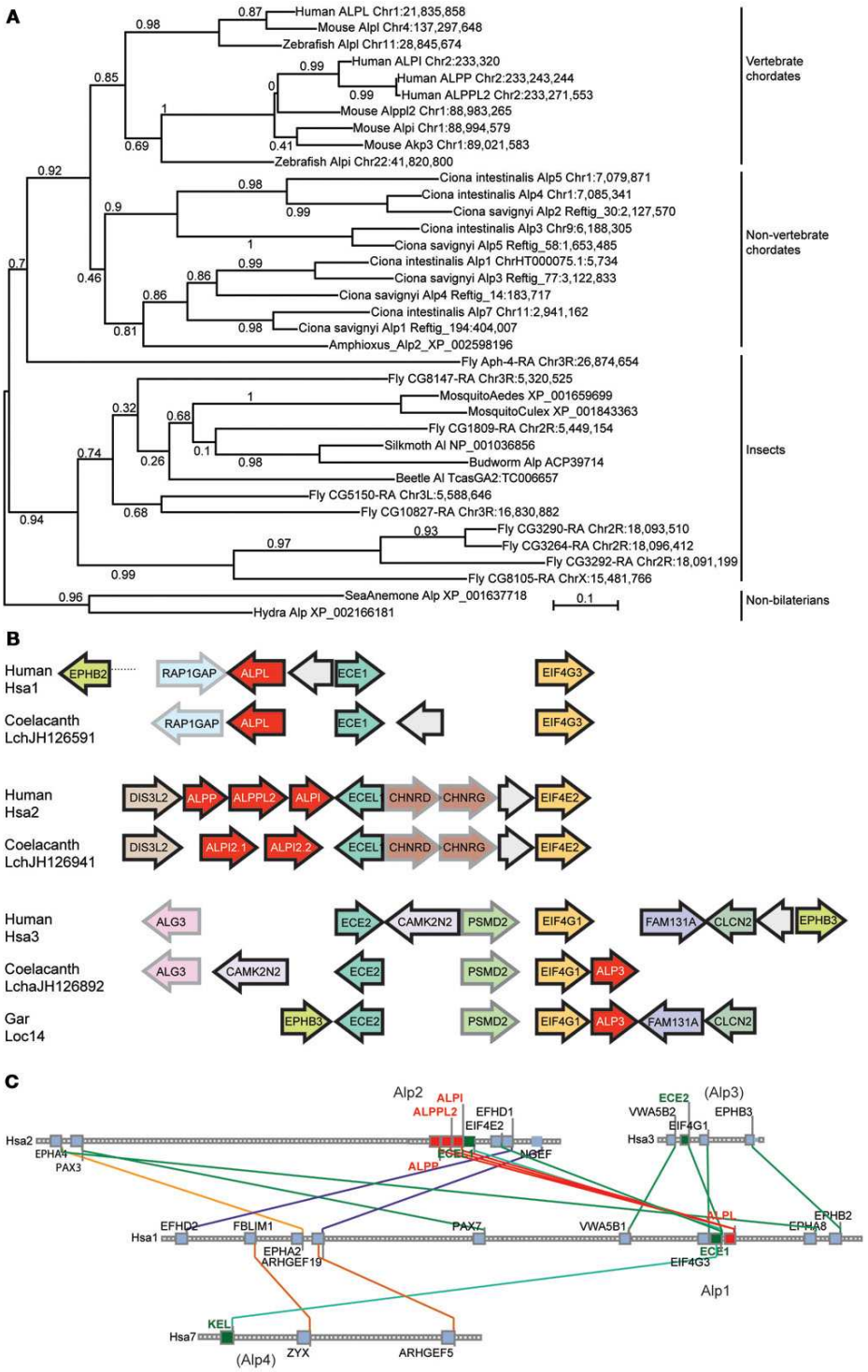


FIGURE 3 | Invertebrate Alp protein sequences and conserved syntenies. (A) *Alp* genes amplified independently in vertebrate chordate and non-vertebrate chordate lineages, and in deuterostomes (including vertebrates) and protostomes (including insects). **(B)** Conserved syntenies among paralogons including human, coelacanth, and gar *Alp* genes. *Alp* genes in red, and paralogs of other gene families, such as *Ece1*, *Ecel1*, and *Ece2* genes and *Eif4* genes shown in unifying colors. Arrows represent direction of transcription. **(C)** Conserved syntenies in four paralogons in the human genome, only two of which have *ALP* genes. The figure shows four human chromosomes, Hsa1, Hsa2, Hsa3, and Hsa7, with lines between chromosomes connecting paralogs. Each square represents a gene; small gray squares from which no lines extend are genes without paralogs in the regions shown.

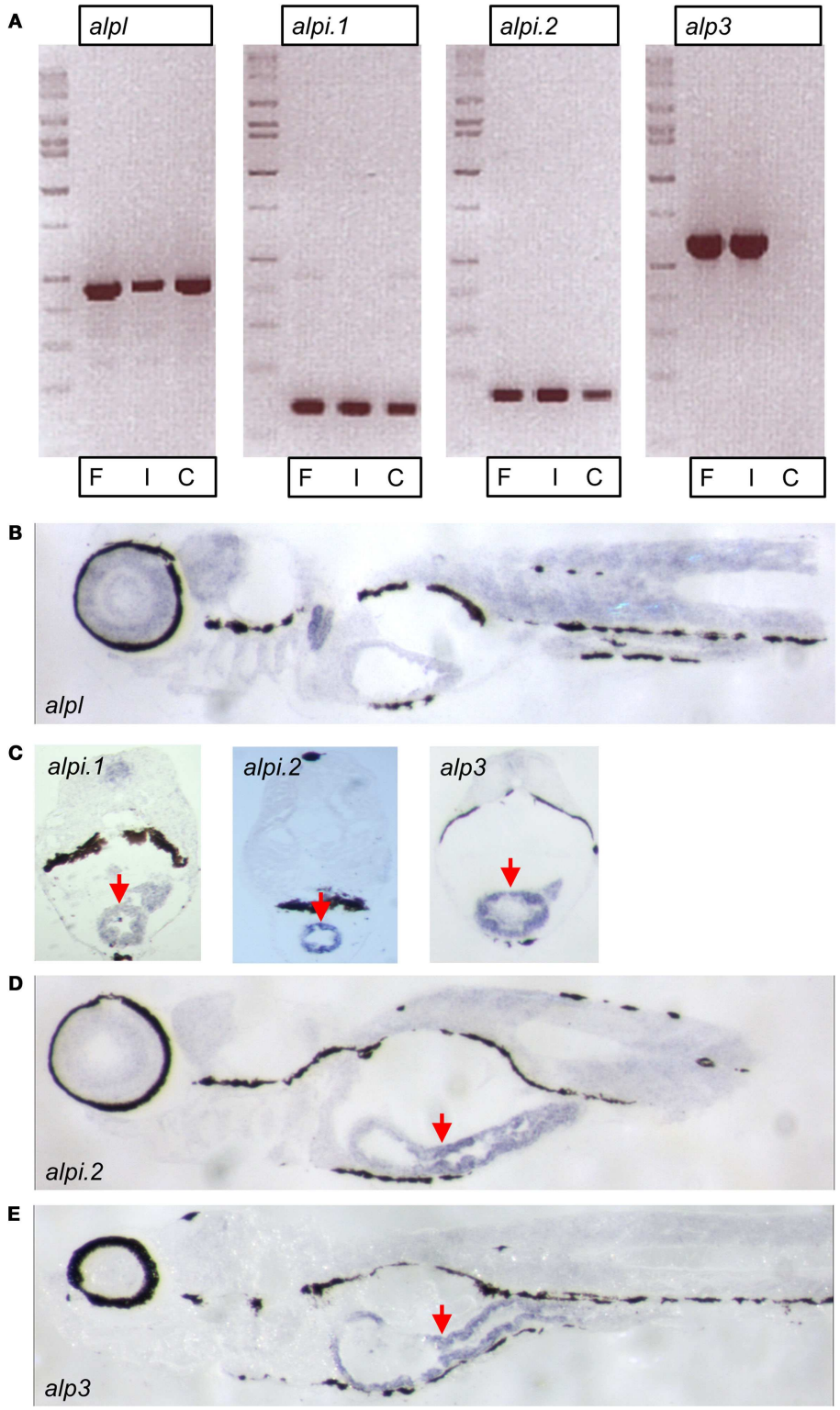


FIGURE 4 | Expression of zebrafish *alp* genes. (A) Semi-quantitative reverse transcription PCR analysis of *alp* gene transcript levels at 7 days post fertilization (dpf) in whole fish (F), dissected intestines (I), and carcasses with intestine removed (C). *In situ* hybridization of 7 dpf larval sagittal **(B,D,E)** and transverse **(C)** sections with probes to *alpl*, *alpi.1*, *alpi.2*, and *alp3* as indicated. The hybridization is visible in blue. Arrows point to the intestinal epithelium.

intestinal-specific zebrafish *alp* genes. We examined this further through a phylogenetic analysis of mammalian *Alp* genes. All mammals examined in this study had a single *Alp* gene at the *Alpl-like* locus, which was surrounded by genes that were orthologs in all species with locally well-assembled genomes, showing that this region of the genome has been well-conserved among mammals. The genomic situation at the *Alpi-like* locus, however, differed greatly among taxa, with several species having multiple *Alpi*-related genes (**Figure 5A**).

Phylogenetic analysis rooted on bird and non-vertebrate chordate Alp sequences showed that many of these Alpi-like genes have arisen very recently (**Figure 5B**). For example, none of the three human genes (*ALPI*, *ALPP*, and *ALPP2*) is a unique ortholog of any non-primate *Alpi-like* gene despite names in common use.

The rodents mouse and rat also show shared and independent *Alpi-like* gene duplications (**Figure 5B**). The mouse *Alpi-like* locus contains three genes called *Alpi*, *Alppl2*, and *Akp3*. Although the human genome has genes called *ALPI* and *ALPPL2,* the mouse and human genes are not one-to-one orthologs according to phylogenetic analysis (see also **Figure 3A**). If this were the case, then the human and mouse *ALPI/Alpi* genes would group together in the tree and the *ALPPLP/Appl2* genes would group together; instead, the tree clearly shows that the human and mouse lineages diverged long before *ALPI* diverged from *ALPP* and *ALPPL2* and before *Alpi* diverged from *Alppl2* and *Akp3*.

The phylogenetic tree suggests that mouse and rat genomes contain at least two pairs of orthologous *Alpi-like* genes called *Akp3/Akp3* and *Alppl2/LOC100359675*, respectively (**Figure 5B**). In addition, the tree shows that rat has two genes derived from a recent duplication called *Alpi* and *Alppl2*. The tree shows with strong support that rat Alpi and Alppl2 arose in a tandem gene duplication event and that the rat Alpi sequence is more closely related to the rat Alppl2 sequence than it is to the mouse Alpi sequence, despite the names. In addition, although genes flanking these mammal's *Alpi-like* locus are orthologs (**Figure 5A**), gene orientations are consistent with non-orthology or a gene-specific inversion event for the rodent genes called *Alpi*.

Similar findings come from a detailed phylogenetic analysis of the primate *Alp* genes (**Figure 6**). Although each primate had a single *Alpl* gene, the Alpi-like clade displayed several cases of independent tandem duplication events. The three *Alpi*-related genes of mouse formed a strong outgroup for the primate *Alpi* genes, suggesting that these three murine genes duplicated independently from the primate genes, confirming that none is a unique ortholog of any of the human paralogs. The single *Alpi* genes in the two Strepsirrhini species – lemur and bushbaby – grouped together at the base of the primate Alpi clade, which would be expected if ancestral primates had a single *Alpi* gene (**Figure 6**). A tandem duplication after the divergence of Strepsirrhini and Haplorrhini produced an Alpi clade and an Alpp clade; subsequently, after the divergence of the human and chimpanzee lineage about 6 million years ago, the human lineage experienced a tandem duplication in the ALPP clade and the chimpanzee lineage had independent duplications in the ALPP lineage, evidence of a dynamic diversification of ALP-related sequences in our recent history.

VARIABLE COPY NUMBERS OF *Alpi* GENES AMONG MAMMALS

We speculate that the rapid gains and losses of *Alpi* genes manifest in vertebrate lineages are driven by dynamic changes in host-microbe associations. In particular, if Alps play a conserved function in detoxifying LPS, then requirements for this gene could change rapidly with changing proportions of Gram-negative bacterial associates, possibly driven by adaptations to different diets that require different microbial physiologies for their metabolism. To explore this hypothesis, we determined the representation of Gram-negative phyla present in a published dataset of fecal samples from 60 mammalian species (Ley et al., 2008a; **Figure 7**). These samples contained an enormous range of proportional representation of Gram-negative phyla from 90% (rock hyrax) to 0% (cow). Even within one host species (humans), proportions ranged from 50 to 9%, emphasizing the variable nature of gut-associated microbial communities. We found no significant correlation between the proportional representation of Gram-negative bacteria and gut physiologies. We note, however, that fecal sampling is unlikely to capture the microbial composition of specialized gut compartments. For example, whereas the cow fecal sample from this study contained no Gram-negative phyla, surveys of cow rumen content typically find a high proportion of Gram-negative species (Jami and Mizrahi, 2012).

The limited number and sometimes low quality of whole genome sequences from mammalian species precluded us from performing a systematic correlation between *Alpi* gene copy number and gut morphology, but we noted some interesting trends. Although many of the mammals have multiple *Alpi-like* genes, three members of the order Carnivora that have simple gut morphologies – the giant panda, domestic dog, and domestic cat – all have only one *Alpi* gene (despite the fact that the panda is an herbivore). In contrast, Artiodactyls, which have more complex digestive systems, showed more complex *Alpi-like* genomics. The cow has six *Alpi*-related genes (called here *Alpi-like 1–6*), only four of which (*Alpil2*;*E1BJE3*, *Alpil4;ENSBTAT000002229*, *Alpil5;F1N2M5*, and *Alpil6*;*F1N6T5*) are annotated in Ensemble. These four annotated genes are arranged in neighboring pairs oriented in the same direction (**Figure 5A**), with Alpil5 and Alpil6 falling as sisters in the tree and as neighbors in the genome and Alpil4 and Alpil5 also diverging as sisters and neighbors, suggesting tandem duplication events (**Figure 5B**). The pig has about three *Alpi-like* genes, although only two of them are well-assembled in Ensemble. One of the pig genes (*Alpil2*; *ENSSSCG00000028816*) appears to be coorthologous to at least two of the cow genes (*Alpi2* and *4*), suggesting that tandem gene duplications occurred in the Artiodactyl lineage before the divergence of swine and bovine lineages. The tree suggests the loss of ancestral Artiodactyl genes in the pig lineage and the tree and gene orientations indicate that independent duplications occurred at least in the cow lineage as well as local inversions during lineage divergence. An open and interesting question is whether different bovine *Alpi-like* genes are expressed in specific portions of the cow's complex digestive system and are adapted to the different microbial contents of each compartment.

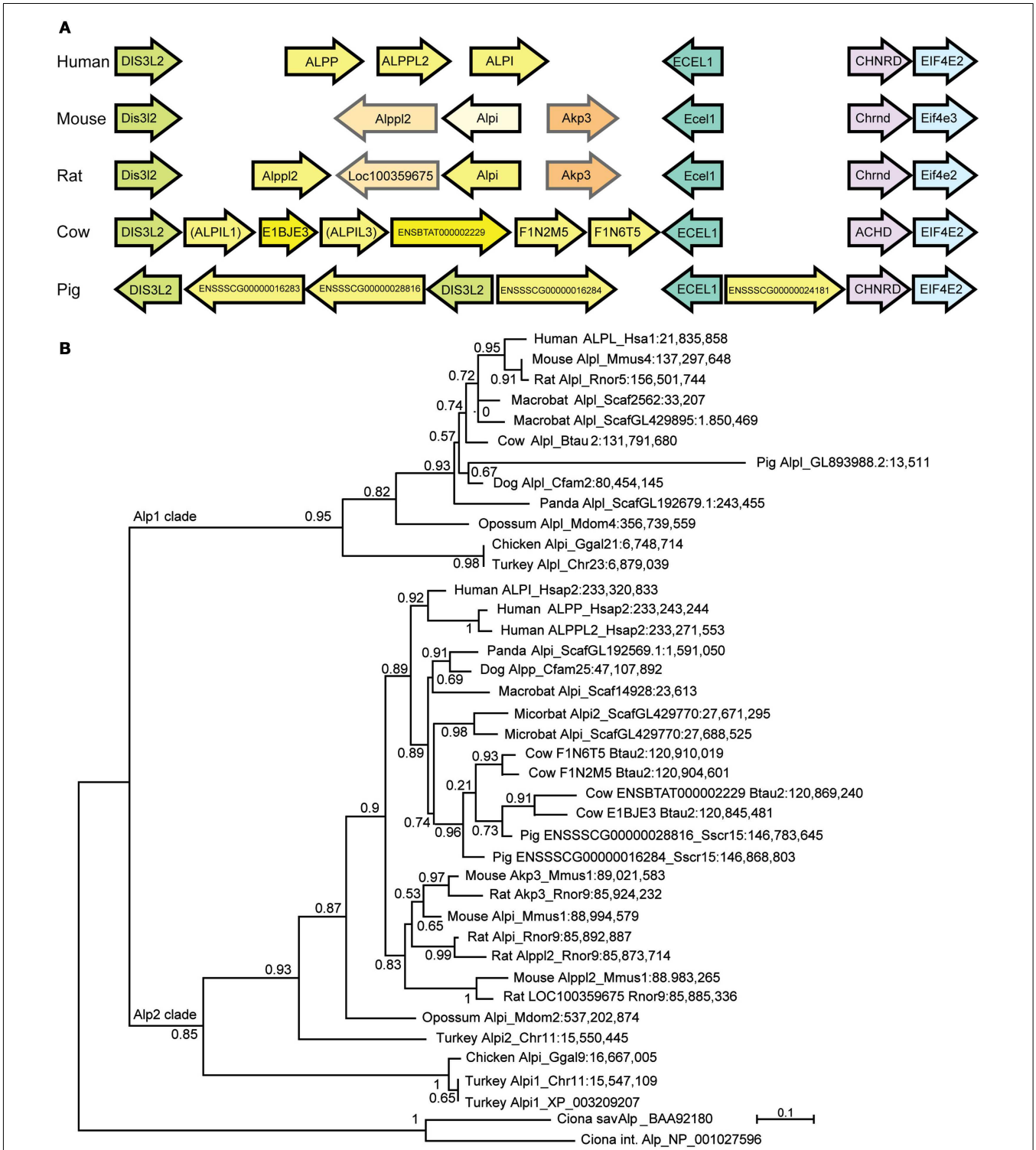


FIGURE 5 | The diversification of mammalian Alp protein sequences. (A) The *Alpi*-related locus in five mammals showing *Alp* genes in yellow. Orthologs between mouse and rat shown in similar shades. **(B)** Maximum likelihood tree showing mammalian and bird *Alp* genes. In general, although the *Alpi* (*Alp2*) locus is orthologous and the flanking genes are orthologs among the various mammals, the tree shows that the *Alpi*-related genes themselves are usually not one-to-one orthologs.

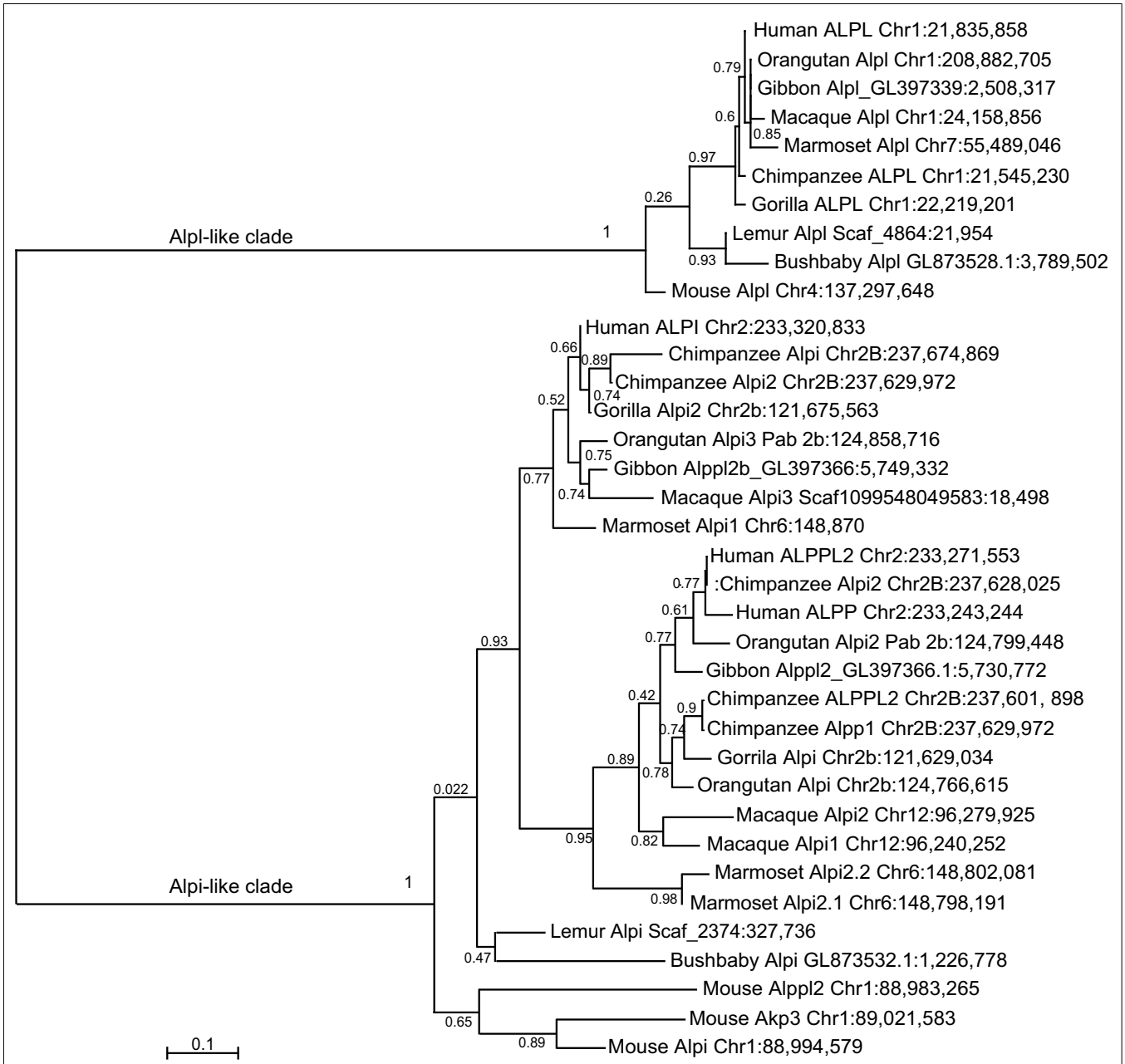


FIGURE 6 | Maximum likelihood tree of primate *Alp* sequences. After the divergence of the Strepsirrhini (lemur and bushbaby) lineage from the Haplorrhini lineages, a tandem gene duplication produced an *Alpi*-like gene and an *Alpp*-like gene with further duplications within some of the Haplorrhini sub-lineages.

DISCUSSION

Our investigation into the phylogenetic relationship between *Alp* genes from zebrafish and other vertebrates reveals a remarkable evolutionary history for this gene family. As summarized in **Figure 8**, our analysis supports the model that the vertebrate *Alp* genes arose from a single *Alp* gene in ancestral chordates that duplicated during two rounds of genome duplication events (VGD1 and VGD2) that preceded the diversification of extant vertebrates after the divergence of vertebrates from non-vertebrate chordates and initially yielded four genes. One of these genes became the modern *Alpl*. This gene was faithfully maintained as a single copy throughout the vertebrates, with the coorthologous gene that would have been generated during a third round of genome duplication, the teleost whole genome duplication (TGD), having been lost. Indeed none of the three predicted coorthologous Alp genes are found within the teleost genomes we examined, but this could be due to chance; because about 75% of zebrafish coorthologues have been lost, the likelihood of all three Alp coorthologues being lost by chance is 42%, clearly a frequent occurrence for any three genes taken at random. The predicted fourth *Alp* generated during the

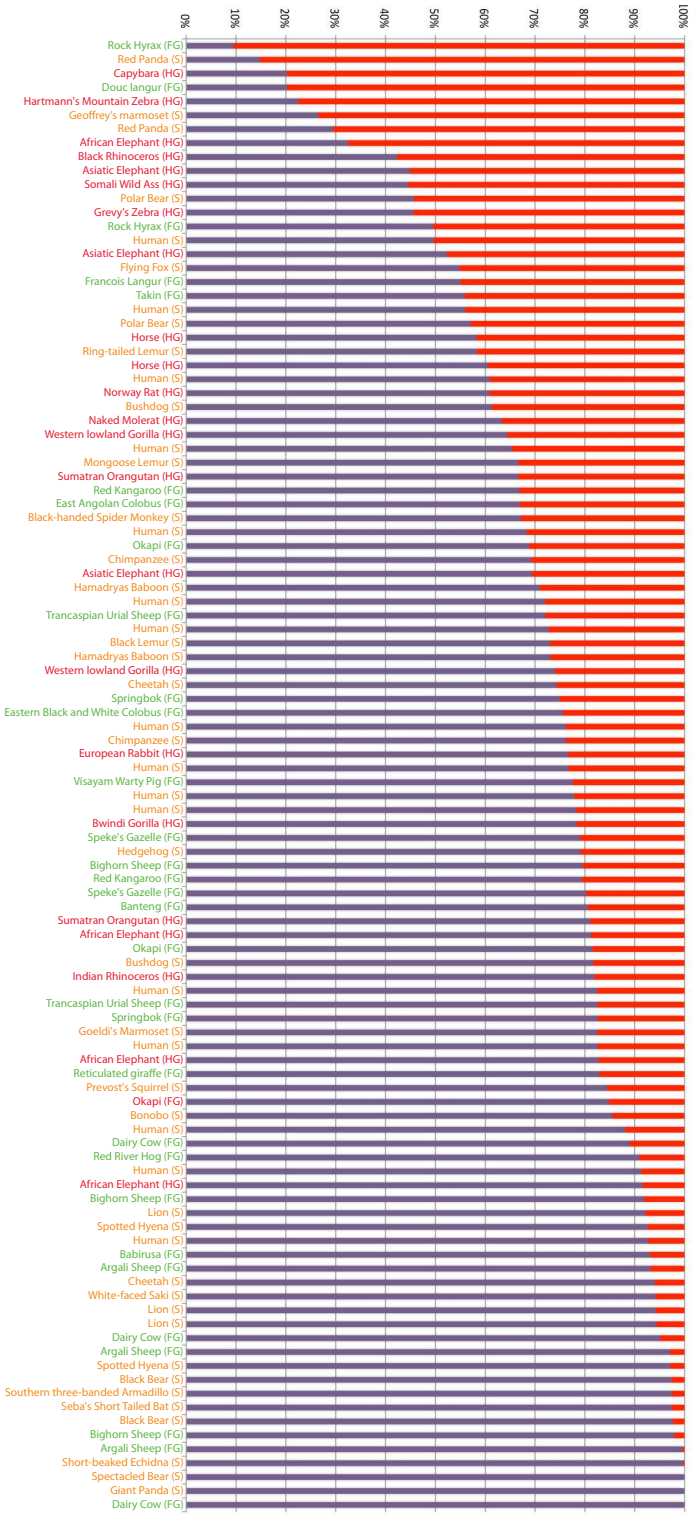


FIGURE 7 | Proportional representation of Gram-negative phyla in fecal samples from different mammals. Bacterial composition of mammalian fecal samples, as previously reported (Ley et al., 2008a), are represented as the proportion of phyla composed of Gram-negative (red) and Gram-positive (purple) bacteria. Host animals' gut morphologies are indicated and color coded (FG, foregut fermenter in green; S, simple in yellow; HG, hindgut fermenter in red).

two vertebrate genome duplications was lost before the divergence of lobefin and rayfin bony vertebrates. *Alp3* persisted in the rayfin lineage where it underwent tandem duplication in the gar and can be found in the basally diverging lobefin, the coelacanth, but was lost from crown group lobefins, the tetrapods. The evolutionary history of *Alp2* (*Alpi-like*) is the most dynamic. Examination of phylogenies of fishes, mammals, and primates reveals that this gene has undergone multiple independent losses and tandem duplications in its history. The role of gene conversion in homogenizing sequences after tandem duplications is as yet unstudied.

We speculate that the basis for the dramatically different evolutionary histories of the tissue non-specific *Alpl* and intestinal *Alpi* genes, which encode enzymes with well-conserved catalytic activities, lies in their different patterns of tissue expression. The ubiquitously expressed *Alpl* plays an important function in bone mineralization in mammals (Golub and Boesze-Battaglia, 2007). In humans, over 261 different mutations in *ALPL* have been linked to hypophosphatasia and skeletal abnormalities, of which 75% are missense mutations[1]. Several of these mutations result in dominant inheritance of the disease (Mornet et al., 2011), suggesting that modest changes in Alp function at sites of bone mineralization can lead to deleterious phenotypes, possibly restricting the evolution of copy number variation of this gene once it became dedicated to this function.

The frequent gains and losses of the intestinally expressed *Alp* genes across vertebrate lineages indicate much greater plasticity in the requirements for the enzyme encoded at this locus. The rapid evolution of gene copy number in the *Alpi* gene clade is especially striking among the mammals, which have undergone dramatic changes in their gut morphologies that accommodate different microbial fermentation strategies during adaptations to different diets, with innovations such as foregut fermentation arising independently multiple times in the mammalian tree (Stevens and Hume, 2004). Possibly the loss of *Alp3* prior to the divergence of the tetrapods put extra pressure on *Alpi* to accommodate changing requirements for Alp activity in the intestine. We hypothesize that requirements for Alpi function in detoxification of luminal LPS from Gram-negative gut bacteria changed dramatically with alterations in associated microbial communities, driven by dietary changes. Additionally, dietary changes may have altered the selective pressures on Alpi's lipid absorption function. *Alpi* gene expansion may have been a way to tune the levels and spatial distribution of this enzyme within the gut and to allow the evolution of different gut compartments adapted to different bacterial communities or other gut functions. For example, one of the three mouse *Alpi-like* genes, *Akp3*, is restricted in its expression to the proximal duodenum (Narisawa et al., 2007). It would be interesting to determine whether the six *Alpi* genes in the cow are restricted to different regions of the elaborate gastrointestinal tract anatomy of this foregut fermenter. The gut is also a major site of infections, and the capacity of intestinal Alp to detoxify LPS is likely to be important for modulating immune responses to pathogens as well as resident beneficial bacteria. Epidemics of infections with Gram-negative enteric pathogens may have been additional driving forces that shaped the rapid evolution of the intestinal *Alp* genes.

In summary, the stark contrast between the evolutionary history of the vertebrate *Alpl* and *Alpi* clades suggests that these genes diversified in the functions they perform in organisms. We propose that the highly dynamic pattern of gene evolution of the *Alpi* clade is indicative of a gene family that serves important functions in mediating host-microbe interactions, which can change dramatically over short periods of time and impose strong selective pressures on animals.

MATERIALS AND METHODS

USE OF VERTEBRATE ANIMALS

All zebrafish experiments were performed using protocols approved by the University of Oregon Institutional Care and Use Committee, and following standard protocols (Westerfield, 2007).

SANGER SEQUENCING OF ZEBRAFISH Alp GENE CODING SEQUENCES

Fresh RNA was extracted from 7 days post fertilization (dpf) zebrafish larvae (illustra RNAspin mini kit, GE Healthcare). Total cDNA was synthesized using RNA as template (SuperScript III reverse transcription kit, Life technologies). The coding sequences of Alp genes were amplified from cDNA in PCR (Phusion DNA polymerase, Thermo Scientific). Primers used in PCR included alplF: 5′-ATGTGGGAATGTGGATGCTTTCTTG-3′, alplR: 5′-TCAGCAAAGCAGCCATTTGACC-3′; alpi.1F: 5′-ATGTGTTTGGTTTACGGTCGGGC-3′, alpi.1R: 5′-TCATCTCAAAACAAGCCAAAACACG-3′; alpi.2F: 5′-ATGGCCAAAACACAAGCCCTG-3′, alpi.2R: 5′-CTAAATAAGAGCAGTAATGGAGGACATCAG-3′; alp3F: 5′-ATGTTTGCTGTCCGTGTGTCC-3′, alp3R: 5′-TCAGTGCAGTAAAATCCTCATCAGTG-3′. PCR products were evaluated by DNA gel electrophoresis for purity and then extracted from gel. The purified PCR products were cloned into the pCR® – Blunt II TOPO® vector (ZeroBlunt® TOPO® PCR Cloning kit, Life Technologies). Clones carrying Alp gene coding sequences were sequenced (Sequetech). The complete coding sequences were assembled using the plasmid editor ApE[2] and submitted to GenBank.

PCR DETECTION OF ZEBRAFISH Alp GENE TRANSCRIPTION

Total cDNA was synthesized from fresh RNA at 7 dpf from whole fish, dissected intestines, or carcasses with guts removed. Gene-specific primers were used in PCR to test the presence of gene transcripts (alplF: 5′-TATTTCTTGGAGATGGGATGGGTG-3′, alplR: 5′-TTCAAAGAGTTGTCTGGCGATGTC-3′; alpi.1F: 5′-GCACCGCGCCAAAGCACAAG-3′, alpi.1R: 5′-CGGGCTTCGGAGGGCACATC-3′; alpi.2F: 5′-TGCGCTTTACGGAAACGGTCCA-3′, alpi.2R: 5′ – TGCGCCATCGGGCCTTTAGC-3′; alp3F: 5′-ATGTTTGCTGTCCGTGTGTCC-3′, alp3R: 5′-ACGA GAAACCG CCTCATCCAG-3′). Equal amounts of cDNA (200 ng/μl, NanoDrop 1000) were added as template in PCR.

IN SITU HYBRIDIZATION

Zebrafish cDNA was obtained as describe above and used in PCR to amplify a unique fragment of *alpl*

[1] http://www.sesep.uvsq.fr/03_hypo_mutations.php#legend

[2] http://biologylabs.utah.edu/jorgensen/wayned/ape/

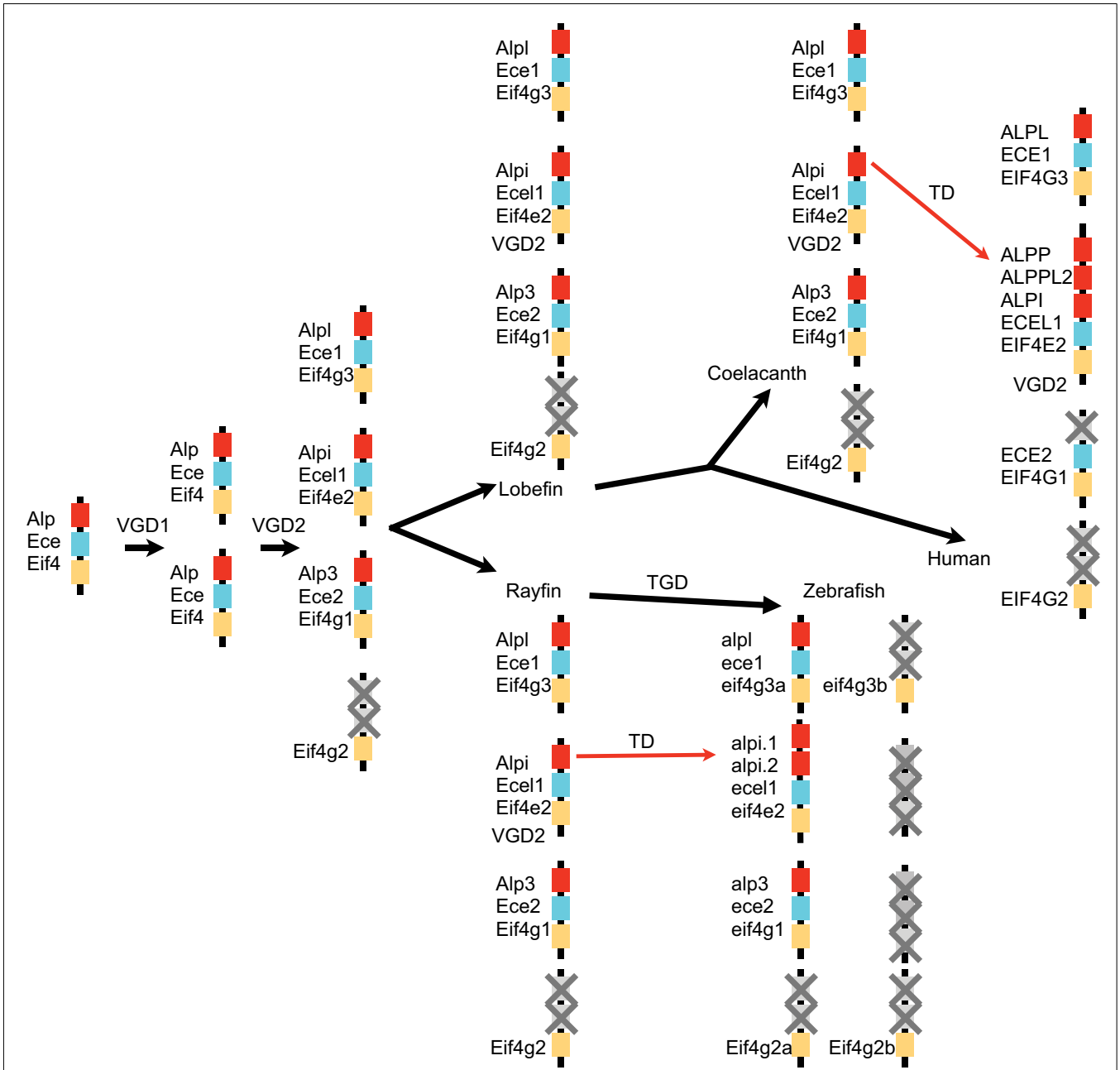


FIGURE 8 | A model for the history of *Alp* genes in vertebrates. A chromosome segment in a non-vertebrate chordate contained *Ece, Alp,* and *Eif4* genes and duplicated twice in the vertebrate genome duplication events (VGD1 and VGD2) to make four paralogons, one of which lost its *Alp* gene. After the divergence of rayfin and lobefin bony fish lineages, the teleost genome duplication event (TGD) did not result in further extant *Alp* genes, but tandem duplications (TD) in the *Alpi* locus provided a number of separate *Alp* genes.

(alplF: TTCCAGAGCAAGAGAAGCGG; alplR: GTCTTAGA GAGGGCGACGTG), *alpi.1* (alpi.1F: CGACCGGGCGATTCAGA GAG; alpi.1R: TGGTGTACGGCTCAAGGCAC), *alpi.2* (alpi.2F: TCACTAACGGGACTCGACCT; alpi.2R: AGGCCATAGCGT GAGGAATG), and alp3 (alp3F: CAGGGTCATCACTCCAGTCG; alp3R: TCTGGACGCTTGTTGTTGGT). The purified PCR product was cloned into the pCR® – Blunt II TOPO® vector (ZeroBlunt® TOPO® PCR Cloning kit, Life Technologies) and validated by sequencing (Sequetech). The verified plasmid was used in PCR to add T7 RNA polymerase binding sites to the gene-specific fragment (alplF: TTCCAGAGCAAGAGAAGCGGC, alplT7-R: GTAATACGACTCACTATAGGGGTCTTAGAGAGGG; ali.1p-F: CGACCGGGCGATTCAGAGAG, alpi.1T7-R: GTAAT-ACGACTCACTATAGGGTGGTGTACGGCTCAA; alpi.2F: TCAC-TAACGGGACTCGACCTGATGT, alpi.2T7-R: GTAATACGACT-CACTATAGGGAGGCCATAGCGTGAG; alp3F: CAGGGTCAT

CACTCCAGTCGGGC, alp3T7-R: GTAATACGACTCACTATAGG GTCTGGACGCTTGTTG). The purified PCR product was used as template to synthesize the DIG labeled RNA probe (DIG RNA Labeling Mix (T7), Roche Applied Science).

Seven dpf zebrafish larvae were fixed in 4% PFA overnight at 4˚C and then washed in 1×PBST. Rinsed larvae were embedded in 1.5% agar and cryo-cut into 16 μm thick sections. Sections were defrosted and air died overnight at room temperature. Diluted RNA probes were added on slides and incubated overnight at 70˚C. After hybridization, slides were rinsed with the wash solution (1×SSC, 50% formamide, 0.1% Tween-20). Slides were then treated with the block solution (MABT, 2% blocking reagent, 20% heat inactivated sheep serum) for 3 h. Following that, first antibody solution (AP conjugated anti-DIG) was applied to slides and incubated overnight at 4˚C. Slides were then washed in AP staining buffer (5 M NaCl, 1 M $MgCl_2$, 1 M Tris pH 9.5, 20% Tween-20) and incubated with NBT/BCIP color regents in the dark at 37˚C. The color developing process was stopped by washing slides in PBST. Slides were then rinsed in ddH_2O, dehydrated in ethanol and eventually mounted in Permount. Images were taken under a Nikon TE2000 inverted microscope with a CoolSNAP camera (Photometrics).

GENOMIC ANALYSES

ALP-related sequences were obtained from sequenced genomes present in Ensembl[3] and Pre-Ensembl[4] by BLASTP searches using the human ALPI (ENSG00000163295) and ALPL (ENSG00000162551) sequences as query. Protein sequences for non-annotated genes in Ensembl were assembled using GeneMark[5] (Lomsadze et al., 2005). Protein sequences were retrieved and subjected to phylogenetic analysis using MUSCLE for multiple alignment (Edgar, 2004), Gblocks to remove poorly aligned positions and divergent regions (Castresana, 2000), the PhyML maximum likelihood method for tree building (Guindon and Gascuel, 2003), and TreeDyn[6] for tree rendering (Chevenet et al., 2006; Dereeper et al., 2008). Conserved synteny analyses were performed using the Synteny Database (Catchen et al., 2009). The sequences used in phylogeny study were provided in supplements.

[3]http://www.ensembl.org
[4]http://pre.ensembl.org
[5]http://exon.gatech.edu/eukhmm.cgi

[6]http://www.phylogeny.fr/

ACKNOWLEDGMENTS

We thank Ruth Ley for sharing microbiota data, Rose Sockol and the staff of the University of Oregon Zebrafish Facility for excellent fish husbandry, and Poh Kheng Loi and the staff of the University of Oregon histology facility for histology services. This research was supported by NIH grants 1R01DK075667 (to Karen Guillemin), 1R01AG031922 (to John H. Postlethwait), and 8R01OD011116 (to John H. Postlethwait). NIH grant HD22486 provided support for the Oregon Zebrafish Facility.

AUTHOR NOTE

New sequences submitted to GenBank: zebrafish *alpl* complete CDS (accession number JX847415), zebrafish *alpi.1* complete CDS (accession number JX847416), zebrafish *alpi.2* complete CDS (accession number JX847417), and zebrafish *alp3* complete CDS (accession number JX847418).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Molecular_Innate_Immunity/10.3389/fimmu.2012.00314/abstract

Data sheet S1 | Alp sequences used to generate Figure 2.

Data sheet S2 | Alp sequences used to generate Figure 3.

Data sheet S3 | Alp sequences used to generate Figure 5.

Data sheet S4 | Alp sequences used to generate Figure 6.

REFERENCES

Akiba, Y., Mizumori, M., Guth, P. H., Engel, E., and Kaunitz, J. D. (2007). Duodenal brush border intestinal alkaline phosphatase activity affects bicarbonate secretion in rats. *Am. J. Physiol. Gastrointest. Liver Physiol.* 293, G1223–G1233.

Bates, J. M., Akerlund, J., Mittge, E., and Guillemin, K. (2007). Intestinal alkaline phosphatase detoxifies lipopolysaccharide and prevents inflammation in zebrafish in response to the gut microbiota. *Cell Host Microbe* 2, 371–382.

Bates, J. M., Mittge, E., Kuhlman, J., Baden, K. N., Cheesman, S. E., and Guillemin, K. (2006). Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Dev. Biol.* 297, 374–386.

Beumer, C., Wulferink, M., Raaben, W., Fiechter, D., Brands, R., and Seinen, W. (2003). Calf intestinal alkaline phosphatase, a novel therapeutic drug for lipopolysaccharide (LPS)-mediated diseases, attenuates LPS toxicity in mice and piglets. *J. Pharmacol. Exp. Ther.* 307, 737–744.

Beutler, B., and Rietschel, E. T. (2003). Innate immune sensing and its roots: the story of endotoxin. *Nat. Rev. Immunol.* 3, 169–176.

Braasch, I., and Postlethwait, J. H. (2012). "Polyploidy in fish and the teleost genome duplication," in *Polyploidy and Genome Evolution*, eds P. S. Soltis and D. E. Soltis (New York: Springer) (in press).

Brun, L. R., Brance, M. L., and Rigalli, A. (2012). Luminal calcium concentration controls intestinal calcium absorption by modification of intestinal alkaline phosphatase activity. *Br. J. Nutr.* 108, 229–233.

Campbell, E. L., MacManus, C. F., Kominsky, D. J., Keely, S., Glover, L. E., Bowers, B. E., et al. (2010). Resolvin E1-induced intestinal alkaline phosphatase promotes resolution of inflammation through LPS detoxification. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14298–14303.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552.

Catchen, J. M., Conery, J. S., and Postlethwait, J. H. (2009). Automated identification of conserved synteny after whole-genome duplication. *Genome Res.* 19, 1497–1505.

Chevenet, F., Brun, C., Banuls, A. L., Jacq, B., and Christen, R. (2006). TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7, 439. doi:10.1186/1471-2105-7-439

Dehal, P., and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3, e314. doi:10.1371/journal.pbio.0030314

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* 36, W465–W469.

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113. doi:10.1186/1471-2105-5-113

Goldberg, R. F., Austen, W. G. Jr., Zhang, X., Munene, G., Mostafa, G., Biswas, S., et al. (2008). Intestinal alkaline phosphatase is a gut mucosal defense factor maintained by enteral nutrition. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3551–3556.

Golub, E. E., and Boesze-Battaglia, K. (2007). The role of alkaline phosphatase in mineralization. *Curr. Opin. Orthop.* 18, 444–448.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704.

Hedges, S. B., Dudley, J., and Kumar, S. (2006). TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* 22, 2971–2972.

Heemskerk, S., Masereeuw, R., Moesker, O., Bouw, M. P., Van Der Hoeven, J. G., Peters, W. H., et al. (2009). Alkaline phosphatase treatment improves renal function in severe sepsis or septic shock patients. *Crit. Care Med.* 37, 417–423, e411.

Jami, E., and Mizrahi, I. (2012). Composition and similarity of bovine rumen microbiota across individual animals. *PLoS ONE* 7, e33306. doi:10.1371/journal.pone.0033306

Kats, S., Brands, R., Seinen, W., De Jager, W., Bekker, M. W., Hamad, M. A., et al. (2009). Anti-inflammatory effects of alkaline phosphatase in coronary artery bypass surgery with cardiopulmonary bypass. *Recent Pat. Inflamm. Allergy Drug Discov.* 3, 214–220.

Kim, H. M., Park, B. S., Kim, J. I., Kim, S. E., Lee, J., Oh, S. C., et al. (2007). Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* 130, 906–917.

Lalles, J. P. (2010). Intestinal alkaline phosphatase: multiple biological roles in maintenance of intestinal homeostasis and modulation by diet. *Nutr. Rev.* 68, 323–332.

Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., et al. (2008a). Evolution of mammals and their gut microbes. *Science* 320, 1647–1651.

Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., and Gordon, J. I. (2008b). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat. Rev. Microbiol.* 6, 776–788.

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y. O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33, 6494–6506.

Lukas, M., Drastich, P., Konecny, M., Gionchetti, P., Urban, O., Cantoni, F., et al. (2010). Exogenous alkaline phosphatase for the treatment of patients with moderate to severe ulcerative colitis. *Inflamm. Bowel Dis.* 16, 1180–1186.

Lynes, M., Narisawa, S., Millan, J. L., and Widmaier, E. P. (2011). Interactions between CD36 and global intestinal alkaline phosphatase in mouse small intestine and effects of high-fat diet. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 301, R1738–R1747.

Martinez-Moya, P., Ortega-Gonzalez, M., Gonzalez, R., Anzola, A., Ocon, B., Hernandez-Chirlaque, C., et al. (2012). Exogenous alkaline phosphatase treatment complements endogenous enzyme protection in colonic inflammation and reduces bacterial translocation in rats. *Pharmacol. Res.* 66, 144–153.

McConnell, R. E., Higginbotham, J. N., Shifrin, D. A. Jr., Tabb, D. L., Coffey, R. J., and Tyska, M. J. (2009). The enterocyte microvillus is a vesicle-generating organelle. *J. Cell Biol.* 185, 1285–1298.

Millán, J. L. (2006). *Mammalian Alkaline Phosphatases: From Biology to Applications in Medicine and Biotechnology*. Weinheim: Wiley-VCH.

Mizumori, M., Ham, M., Guth, P. H., Engel, E., Kaunitz, J. D., and Akiba, Y. (2009). Intestinal alkaline phosphatase regulates protective surface microclimate pH in rat duodenum. *J. Physiol. (Lond.)* 587, 3651–3663.

Mornet, E., Yvard, A., Taillandier, A., Fauvert, D., and Simon-Bouy, B. (2011). A molecular-based estimation of the prevalence of hypophosphatasia in the European population. *Ann. Hum. Genet.* 75, 439–445.

Narisawa, S., Hoylaerts, M. F., Doctor, K. S., Fukuda, M. N., Alpers, D. H., and Millan, J. L. (2007). A novel phosphatase upregulated in Akp3 knockout mice. *Am. J. Physiol. Gastrointest. Liver Physiol.* 293, G1068–G1077.

Narisawa, S., Huang, L., Iwasaki, A., Hasegawa, H., Alpers, D. H., and Millan, J. L. (2003). Accelerated fat absorption in intestinal alkaline phosphatase knockout mice. *Mol. Cell. Biol.* 23, 7525–7530.

Pickkers, P., Heemskerk, S., Schouten, J., Laterre, P. F., Vincent, J. L., Beishuizen, A., et al. (2012). Alkaline phosphatase for treatment of sepsis-induced acute kidney injury: a prospective randomized double-blind placebo-controlled trial. *Crit. Care* 16, R14.

Poelstra, K., Bakker, W. W., Klok, P. A., Hardonk, M. J., and Meijer, D. K. (1997a). A physiologic function for alkaline phosphatase: endotoxin detoxification. *Lab. Invest.* 76, 319–327.

Poelstra, K., Bakker, W. W., Klok, P. A., Kamps, J. A., Hardonk, M. J., and Meijer, D. K. (1997b). Dephosphorylation of endotoxin by alkaline phosphatase in vivo. *Am. J. Pathol.* 151, 1163–1169.

Rader, B. A., Kremer, N., Apicella, M. A., Goldman, W. E., and McFall-Ngai, M. J. (2012). Modulation of symbiont lipid A signaling by host alkaline phosphatases in the squid-vibrio symbiosis. *MBio* 3. doi: 10.1128/mBio.00093-12

Ramasamy, S., Nguyen, D. D., Eston, M. A., Alam, S. N., Moss, A. K., Ebrahimi, F., et al. (2011). Intestinal alkaline phosphatase has beneficial effects in mouse models of chronic colitis. *Inflamm. Bowel Dis.* 17, 532–542.

Rentea, R. M., Liedel, J. L., Welak, S. R., Cassidy, L. D., Mayer, A. N., Pritchard, K. A. Jr., et al. (2012). Intestinal alkaline phosphatase administration in newborns is protective of gut barrier function in a neonatal necrotizing enterocolitis rat model. *J. Pediatr. Surg.* 47, 1135–1142.

Shifrin, D. A. Jr., McConnell, R. E., Nambiar, R., Higginbotham, J. N., Coffey, R. J., and Tyska, M. J. (2012). Enterocyte microvillus-derived vesicles detoxify bacterial products and regulate epithelial-microbial interactions. *Curr. Biol.* 22, 627–631.

Stevens, C. E., and Hume, I. D. (2004). *Comparative Physiology of the Vertebrate Digestive System*. Cambridge, New York: Cambridge University Press.

Sullivan, C., Charette, J., Catchen, J., Lage, C. R., Giasson, G., Postlethwait, J. H., et al. (2009). The gene history of zebrafish tlr4a and tlr4b is predictive of their divergent functions. *J. Immunol.* 183, 5896–5908.

Tuin, A., Poelstra, K., De Jager-Krikken, A., Bok, L., Raaben, W., Velders, M. P., et al. (2009). Role of alkaline phosphatase in colitis in man and rats. *Gut* 58, 379–387.

Van Veen, S. Q., Dinant, S., Van Vliet, A. K., and Van Gulik, T. M. (2006). Alkaline phosphatase reduces hepatic and pulmonary injury in liver ischaemia – reperfusion combined with partial resection. *Br. J. Surg.* 93, 448–456.

Van Veen, S. Q., Van Vliet, A. K., Wulferink, M., Brands, R., Boermeester, M. A., and Van Gulik, T. M. (2005). Bovine intestinal alkaline phosphatase attenuates the inflammatory response in secondary peritonitis in mice. *Infect. Immun.* 73, 4309–4314.

Verweij, W. R., Bentala, H., Huizinga-Van Der Vlag, A., Van Loenen-Weemaes, A.-M., Kooi, K., Meijer, D. K., et al. (2004). Protection against an Escherichia coli-induced sepsis by alkaline phosphatase in mice. *Shock* 22, 174–179.

Westerfield, M. (2007). *The Zebrafish Book*, 4th Edn. Eugene: University of Oregon Press.

Whyte, M. P. (2010). Physiological role of alkaline phosphatase explored in hypophosphatasia. *Ann. N. Y. Acad. Sci.* 1192, 190–200.

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.

Received: 31 August 2012; accepted: 21 September 2012; published online: 12 October 2012.

Citation: Yang Y, Wandler AM, Postlethwait JH and Guillemin K (2012) Dynamic evolution of the LPS-detoxifying enzyme intestinal alkaline phosphatase in zebrafish and other vertebrates. Front. Immun. ***3****:314. doi: 10.3389/fimmu.2012.00314*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.

frontiers in
IMMUNOLOGY

REVIEW ARTICLE
published: 12 March 2012
doi: 10.3389/fimmu.2012.00025

Macroevolutionary immunology: a role for immunity in the diversification of animal life

Eric S. Loker*

Center for Evolutionary and Theoretical Immunology, Department of Biology, Division of Parasitology, Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM, USA

Edited by:
Larry J. Dishaw, University of South Florida, USA

Reviewed by:
Lynn Martin, University of South Florida, USA
Gérard Eberl, Institut Pasteur, France

****Correspondence:***
Eric S. Loker, Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA.
e-mail: esloker@unm.edu

An emerging picture of the nature of immune systems across animal phyla reveals both conservatism of some features and the appearance among and within phyla of novel, lineage-specific defense solutions. The latter collectively represent a major and underappreciated form of animal diversity. Factors influencing this macroevolutionary (above the species level) pattern of novelty are considered and include adoption of different life styles, life histories, and body plans; a general advantage of being distinctive with respect to immune defenses; and the responses required to cope with parasites, many of which afflict hosts in a lineage-specific manner. This large-scale pattern of novelty implies that immunological phenomena can affect microevolutionary processes (at the population level within species) that can eventually lead to macroevolutionary events such as speciation, radiations, or extinctions. Immunologically based phenomena play a role in favoring intraspecific diversification, specialization and host specificity of parasites, and mechanisms are discussed whereby this could lead to parasite speciation. Host switching – the acquisition of new host species by parasites – is a major mechanism that drives parasite diversity and is frequently involved in disease emergence. It is also one that can be favored by reductions in immune competence of new hosts. Mechanisms involving immune phenomena favoring intraspecific diversification and speciation of host species are also discussed. A macroevolutionary perspective on immunology is invaluable in today's world, including the need to study a broader range of species with distinctive immune systems. Many of these species are faced with extinction, another macroevolutionary process influenced by immune phenomena.

Keywords: macroevolution, immunology, host–parasite interactions, evolutionary immunology, host shifting, biodiversity

INTRODUCTION

Recent years have witnessed a dramatic increase in our understanding of the diversity of immune systems across animal phyla (Flajnik and Kasahara, 2010; Messier-Solek et al., 2010; Rast and Litman, 2010; this volume). Availability of genome sequences from a broad variety of animals coupled with an increased appreciation for the diversity of their defenses has given the study of immunity a much stronger evolutionary foundation, one that has been further enriched by studies of plant immunity and responses of bacteria and archaea to threats to their genomes (Horvath and Barrangou, 2010). The increasing depth and breadth of immunological studies is also bringing to light a greater awareness of the impact that immunity has had on all forms of life, especially parasites. Here "parasite" is used inclusively, referring to infectious agents ranging from viruses to bacteria to protists to multicellular helminths. The features uniting parasites are that they infect hosts, provoke some degree of fitness-diminishing harm, prompt the deployment of immune responses, and undertake immune evasive actions. "Immune systems" are referred to as those molecules, cells, tissues, and organs that protect hosts from parasites (see caveats below). This discussion excludes a broad range of behavioral defenses like preening (Bush and Clayton, 2006) or avoidance (e.g., Mooring et al., 2003; Garnick et al., 2010).

Here I attempt to draw together ideas that begin to put immunological phenomena into a broader macroevolutionary context. Macroevolution is the study of patterns, and the evolutionary processes that have generated them, at or above the species level (Stanley, 1998; Levinton, 2001). It is the study of how and why life has diversified, and attempts to document how and why lineages of organisms have come into being and either given rise to additional lineages or gone extinct. The process of speciation is germane to macroevolutionary studies because it is the process responsible for increasing the diversity of life forms. Extinction and its causes are also an essential part of such studies.

The attributes of immune systems across the spectrum of animal diversity provide a new way to view and reinterpret the diversity of animals. Immune systems exhibit unforeseen novelty and thus offer new insights into major selective forces influencing animal life. Also, phenomena that are fundamentally immunological provide fertile ground for investigating the impact of immunity as a driver of biodiversity. The role of immune systems in macroevolutionary processes is one that deserves recognition and more study.

In considering what is to follow, several caveats should be borne in mind:

(1) We are just beginning to view molecular components of immune systems from a broad sampling of animal phyla. Detailed analyses are still few for how these presumptive immune components actually function in defense, and how critical their roles might be in protecting the organisms in question.
(2) Also poorly known are the specific parasite threats faced by the more obscure groups of animals serving as hosts.
(3) Many of the examples of immunological novelty presented below emphasize differences at the phylum level. Some phyla such as the Arthropoda, Mollusca, and Chordata are immense in species numbers and undoubtedly collectively employ as yet many undiscovered immune capabilities. Also, some of the smaller animal phyla are essentially unexplored with respect to their immune systems. Once understood, these additional examples will only add to the overall diversity of immune responses.
(4) It is not always easy to circumscribe "the immune system" or an "immune response." This is particularly so in cases where potent defenses for parasites result from selection for variant alleles for genes like hemoglobin B or apolipoprotein L-1 that otherwise might not be considered a core part of the immune system (Anstee, 2010; Barreiro and Quintana-Murci, 2010; Genovese et al., 2010; Wheeler, 2010).

IMMUNOLOGICAL NOVELTY AMONG ANIMAL PHYLA: AN UNDERAPPRECIATED FORM OF DIVERSITY

Discoveries relating to the innate immune systems of plants, flies, and mammals have tended to accentuate the similarities among them, implying a grand conservatism even across kingdoms with respect to basic immune system design and function. Indeed, *there are* intriguing similarities between the membrane-associated and intracytoplasmic receptors of plants and animals suggestive that some basic solutions to recognition and response to parasites have been conserved since at least the time animal and plant lineages diverged. However, particularly given that some of these similarities are a likely result of convergent evolution rather than indicative of a common origin (Ausubel, 2005), conserved immune features are not the emphasis here. Rather, this overview accentuates the emergence of immunological novelty among and within animal phyla (**Figure 1**; **Table 1**).

The most basal animal group is the phylum Porifera, the sponges (Srivastava et al., 2010). Sponges lack the complex tissue and organ structure found in other animal phyla, and lack cells specialized for protection from parasites. Although sponge immunobiology is in its infancy, one of the best-known sponges, *Suberites domuncula*, possesses membrane-spanning molecules that contain an intracellular Toll-interleukin 1 receptor (TIR) domain, though it lacks an external leucine-rich repeat pattern recognition receptor more typical of TLRs. On the basis of having a MyD88 homolog, *S. domuncula* has at least the rudiments of an NF-κB signaling pathway. Sponges also have molecules for attacking bacterial membranes, presumptive antiviral responses (Schroder et al., 2008), and diversified scavenger receptor cysteine-rich molecules of unknown function (Wiens et al., 2007).

Among basal animals, it is members of the phylum Cnidaria (jellyfish, *Hydra*, anemones, and corals) that have proven most surprising with respect to the large size and content of their genomes, including their immune systems. Cnidarians have distinct tissues but lack organs and are generally considered to be diploblastic, meaning they have recognizable ecto- and endoderm, but lack well-developed mesoderm tissue. Like sponges, they lack recognizable specialized immune cells. However, the starlet sea anemone *Nematostella vectensis* has at least one TLR, an NF-κB signaling pathway, a homolog of a complement 3-like molecule, the likely presence of functioning intracellular NOD-like receptors (NLRs), perforin-like molecules, diverse C-type lectins (Wood-Charlson and Weis, 2009), and even a recognizable homolog of the recombination activating gene, RAG1 (Miller et al., 2007; Augustin et al., 2010). Cnidarians often live in colonies and have to contend with encroaching competitors, including conspecifics. For this they have well-developed mechanisms to recognize self and non-self. One of the responsible molecules has been identified, and is surface expressed, polymorphic and possesses three external immunoglobulin superfamily (IgSF) domains (Nicotra et al., 2009).

The remaining animals, the Bilateria, are bilaterally symmetrical and triploblastic, with well-developed tissues and organs. They often have specialized immune cells. Most fall into two major lineages, the protostomes and deuterostomes. Among the protostomes, representatives of the molting clade (Ecdysozoa) have been most extensively studied in an immune context, as this clade includes nematodes and arthropods, both containing well-studied model organisms. *Caenorhabditis elegans* and other nematodes have reduced genomes and from an immunological perspective are surprising for what they *do not* have. Although *C. elegans* has one bonafide TLR that plays a role in defense against some bacteria (Tenor and Aballay, 2008), it lacks Myd88, NF-κB and several other components of the canonical Toll pathway. NLRs are also lacking. Nonetheless, *C. elegans* can mount inducible, parasite specific responses. It has several novel signaling pathways for defense (Irazoqui et al., 2010) and produces many distinctive antimicrobial peptides (AMPs) for protection from bacteria (Roeder et al., 2010). C-type lectins may serve as recognition molecules in *C. elegans*. The preoccupation with production of AMPs by gut cells reflects their diet of bacteria, which could include potential parasites (Roeder et al., 2010).

Another prominent model of ecdysozoan immunity is *Drosophila* (Lemaitre and Hoffmann, 2007), but increasingly other insect are studied as well, such as mosquitoes because of their role in transmitting human parasites (Bartholomay et al., 2010). Insects have dedicated immune cells such as plasmatocytes and lamellocytes that circulate through their open circulatory system and phagocytose or encapsulate foreign objects. The *Drosophila* immune system also shows evidence of gene loss: it lacks the C3-like complement component and NLR homologs found in cnidarians. Their TLRs are different from those of vertebrates in that they do not engage microbial ligands directly. Of their nine TLR genes, only one or two function in immunity, activating NF-κB signaling pathways in the fat body to produce AMPs (Lemaitre

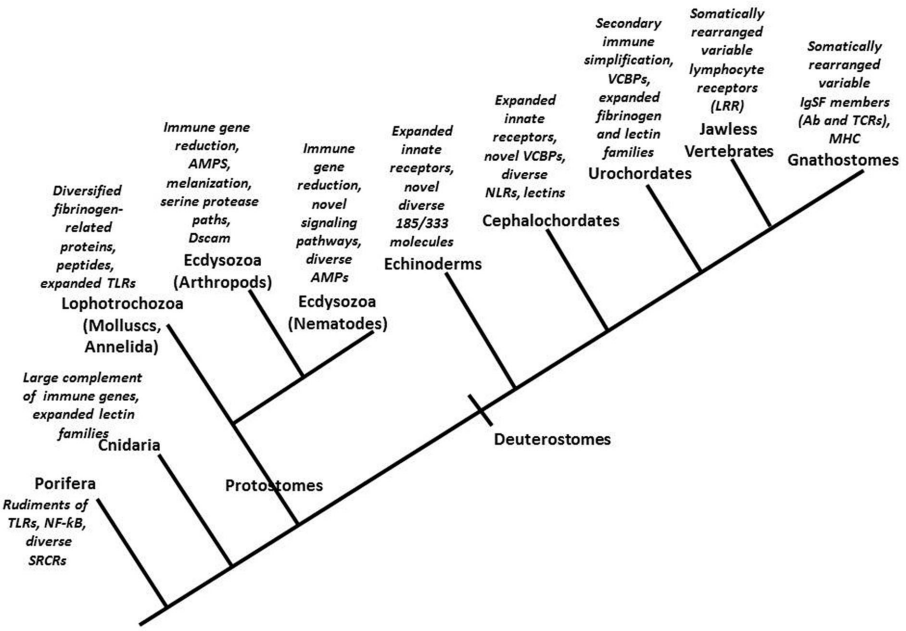


FIGURE 1 | An overview of some of the novel features associated with immune responses of representatives of major animal lineages (see text for details). TLR, Toll-like receptor; AMP, antimicrobial peptide; Dscam, Down syndrome cell adhesion protein; VCBPs, variable region-containing chitin-binding proteins; NLRs, intracellular NOD-like receptors; LRR, leucine-rich repeat; IgSF, immunoglobulin superfamily; Ab, antibodies; TCR, T cell receptor; MHC, major histocompatibility complex.

and Hoffmann, 2007). Insects are by no means immunologically bereft though. They have a number of other effective defense components not seen in many other organisms. They have elaborate cascades of CLIP-serine proteases that mediate and coordinate phagocytosis, nodule formation, encapsulation, and AMP formation, and they can deposit layers of melanin around foreign objects (Kanost et al., 2004). They engage multimeric fibrinogen-related proteins (FREPs) in parasite recognition (Dong and Dimopoulos, 2009) and employ Down syndrome cell adhesion molecule (Dscam), a member of the IgSF, in antigen recognition. Tens of thousands of Dscam isoforms can potentially be generated by alternative splicing (Schmucker and Chen, 2009) and parasite challenge-specific Dscam splice form repertoires can be produced (Dong et al., 2006).

Insect studies provide additional examples of immunological novelty, at the ordinal, family, or even genus level. One example is provided by *Drosophila* and *Anopheles*, both in the same order (Diptera), but representing very different life styles and having been separate lineages for 250 million years. Gene families involved in immunity have evolved rapidly and divergently in the two dipterans. For example, with respect to thioester containing proteins (TEPs), *Anopheles* has 10 genes and *Drosophila* only four, with only one orthologous pair between the two. *Anopheles* has 58 fibrinogen-like immunogenes whereas *Drosophila* has only 14, with only two shared orthologous pairs (Dong and Dimopoulos, 2009).

At the family level, a comparison of three different mosquito genera (*Aedes*, *Anopheles*, and *Culex*, all in the Culicidae) has revealed prominent genus specific expansion of some immune gene families (Bartholomay et al., 2010). Comparative studies of *Drosophila* species are particularly revealing, showing that novel immune genes and immune gene families have originated relatively recently, suggestive of a role of parasites in driving adaptive evolution in flies (Sackton et al., 2007). Furthermore, for particular immune proteins, the amino acids under positive selection vary between *Drosophila* species groups, suggesting different fly species experience different parasite pressures (Morales-Hojas et al., 2009). Insects with very different life styles, such as the social honey bees (Evans et al., 2006) and ants (Smith et al., 2011), and symbiont-dependent aphids (Pennisi, 2009) likewise have immune systems that are surprisingly divergent from the *Drosophila* immune system.

The other major lineage of protostomes, the Lophotrochozoa, includes prominent groups such as the flatworms, annelids, and mollusks. In the polychete annelid *Capitella capitata*, TLRs have undergone an expansion to over 100 genes, most of which are similar, suggestive of recent duplication. Another annelid, the leech *Helobdella robusta*, has only 16 TLR homologs which are not only highly divergent from one another but also are not orthologous to any of the polychete sequences (Davidson et al., 2008). In the freshwater snail *Biomphalaria glabrata*, FREPs are encoded by an expanded gene family, and are implicated in defense against gastropod parasites such as digenetic trematodes (Hanington et al., 2010a,b). In *B. glabrata*, FREPs are particularly noteworthy for being comprised of juxtaposed IgSF and fibrinogen domains, and for the fact they are somatically diversified during the production of hemocytes by the snails (Zhang et al., 2004). Expanded families of C-type lectins are present in other mollusks and the bivalve

Table 1 | An overview of animal immune systems emphasizing distinctive immune features and diversified defense molecules.

Phylum	Body plan	Specialized defense cells	Genome reduction	Distinctive immune features	Diversified defense molecules	Reference
Porifera (sponges)	Diploblastic	No	No	TIR domains, lack external LRR, MyD88 homolog, NF-κB rudiments, lack death domains, LPS-interacting proteins, perforin-like molecules, antiviral 2′-5′ oligoadenylate system	Scavenger receptor cysteine-rich molecules	Wiens et al. (2007), Schroder et al. (2008)
Cnidaria (anemones, corals, jellyfish, *Hydra*	Diploblastic with mesoglea	No	No	Bonafide TLR and NF-κB pathway, complement-3 component, multiple NACHT domains and NLRs likely, recognizable RAG1 homolog. *Hydra* lacks canonical TLR with both LRR and TIR, allorecognition molecules present including with IgSF domains	Diverse C-type lectins	Wood-Charlson and Weis (2009), Miller et al. (2007), Augustin et al. (2010), Hemmrich et al. (2007), Nicotra et al. (2009)
BILATERIA PROTOSTOMES ECDYSOZOA						
Nematoda (*Caenorhabditis* and many others)	Triploblastic	No	Yes	One TLR that plays a role in defense against some bacteria, lacks canonical Toll pathway and NLRs, but can mount inducible defense responses and have several novel defense-related signaling pathways	Produces many caenopores and other antimicrobial peptides including 42 NLPs caenacins, diverse C-type lectins	Irazoqui et al. (2010), Roeder et al. (2010)
Arthropoda (insects, crustaceans, and many others	Triploblastic	Yes	Yes	NLR and complement-3 components lacking, have one or two TLRs functioning with NF-κB pathways, TLRs do not engage ligands directly, produce antimicrobial peptides, CLIP-protease cascades, melanization reactions	Multimeric fibrinogen-related molecules, the IgSF member Dscam with multiple isoforms	Lemaitre and Hoffmann (2007), Kanost et al. (2004), Dong and Dimopoulos (2009), Schmucker and Chen (2009), Dong et al. (2006), Brites et al. (2008), Smith et al. (2011)
LOPHOTROCHOZOA						
Annelida (earthworms, leeches, polychaetes)	Triploblastic	Yes	No	Over 100 TLR genes, extensive involvement of coelomocytes from coelom in defense, cytotoxicity against allogeneic cells, hemolytic and clotting factors in body fluid, antimicrobial peptides, and protective body mucus	Expanded set of TLRs in polychaetes	Davidson et al. (2008), Salzet et al. (2006)

Mollusca (cephalopods, snails, bivalves chitons, others)	Triploblastic	Yes	No	Involvement of body mucus in protection, TLRs and Toll pathway present, little melanization, hemocytes working with lectins like galectins or fibrinogen-containing proteins (FREPs), mitogen-activated protein kinase pathways, complement-like factors antimicrobial peptides	Somatic diversification of FREPS by point mutation and gene conversion, large C-type lectin families, diversified myticin C in bivalves	Hanington et al. (2010a,b), Bayne (2009), Zhang et al. (2004), Yoshino et al. (2008), Costa et al. (2009), Loker (2010)
DEUTEROSTOMES						
Echinodermata (sea urchins, starfish, brittle stars, crinoids, sea cucumbers)	Triploblastic	Yes	No	"Expanded" innate immune system with >220 TLRs, >200 NLRs, >200 SRCR genes, Toll pathway, lectin and alternative pathways, RAG1 and RAG2 homologs present	In addition to expanded sets of TLRs, NLRs, and SRCRs, they also have novel Sp 185/333 gene family producing diverse immune proteins	Messier-Solek et al. (2010), Hibino et al. (2006), Ghosh et al. (2010)
CHORDATA						
Cephalochordata (amphioxus)	Triploblastic	Yes	No	"Expanded innate immune system with ∼72 TLRs, >92 NLRs, −270 SRCRs, >1200 C-type lectins, possesses distinctive variable region-containing chitin-binding proteins (VCBPs), have functioning complement, RAG1 and possibly RAG2 present	In addition to expanded sets of TLRs, NLRs, SRCRs, and lectins, they have polymorphic VCBPs arising from a variety of mechanisms	Huang et al. (2008), Messier-Solek et al. (2010), Dishaw et al. (2008, 2010), Cannon and Litman (2009)
Urochordata (tunicates)	Triploblastic	Yes	Yes	V-like and CI-like domains present, VCBPs present, have three TLRs, lack complement, or expansion of any gene family relevant to vertebrate immunity	Have expanded families of C-type lectins and fibrinogen-related proteins	Azumi et al. (2003), Cannon et al. (2004)
Agnathans (lampreys, hagfish)	Triploblastic	Yes	No	Lack RAG1 and RAG2 and do not produce TCRs or immunoglobulins, but do have two basic types of lymphocytes and produce variable lymphocyte receptors (VLRs) with LRRs	Produce diverse VLRs through somatic rearrangement of modules with leucine-rich repeats	Pancer et al. (2004), Rogozin et al. (2007)
Gnathostomes (fish, amphibians reptiles, birds, mammals)	Triploblastic	Yes	No	Modest numbers of TLRs (10–25) and NLRs (20–35), three complement pathways, somatic diversification of both Ig and TCR, involvement of MHC, memory, heightened secondary response, affinity maturation in some	Both Ig and TCRs diversified somatically	Messier-Solek et al. (2010), Litman et al. (2010)

Mytilus edulis is capable of generating diversified forms of the AMP myticin C both within and among individuals (Costa et al., 2009).

Among invertebrate deuterostomes, the sea urchin has proven surprising in featuring dramatic echinoderm-specific expansion of several recognition molecules (Hibino et al., 2006). Sea urchins possess >220 TLR genes (vertebrates usually have 21–25), >200 NLRs (mammals have 20–35), >200 SRCR genes (humans have 16; Messier-Solek et al., 2010), and a novel Sp 185/333 gene family. The latter gene family produces a repertoire of defense proteins more diverse than the sequence diversity encoded in the genes, indicative of the presence of another mechanism to generate diversity (Ghosh et al., 2010). Sea urchins also possess an NF-κB pathway, lectin and alternative complement pathways and homologs of RAG1 and RAG2, but do not produce immunoglobulins (Ig), T cell receptors (TCRs), or have a major histocompatibility complex (MHC; Hibino et al., 2006).

With respect to our own phylum, the Chordata, the cephalochordate *Branchiostoma* (better known as Amphioxus or the lancelet), is novel in having expanded families of TLRs, NLRs, and SRCRs (Huang et al., 2008), over 1,200 C-type lectins, and an extraordinary diversity in adaptors/facilitators and signaling/effector domains functioning downstream from their NLRs (Huang et al., 2008; Messier-Solek et al., 2010). Amphioxus also possesses distinctive variable region-containing chitin-binding proteins (VCBPs; Dishaw et al., 2008; Cannon and Litman, 2009) which are further distinguished by high levels of polymorphism, resulting in yet another distinct "hyper-diversified," multigene immune receptors family (Dishaw et al., 2010). Cephalochordates have a functioning complement system operating via the alternative and lectin pathways, including with a distinctive expanded number of C1q-like genes (Huang et al., 2008; Messier-Solek et al., 2010). A RAG1 gene is present, and possibly a RAG2 gene as well (Dong et al., 2005).

The urochordates, or tunicates, the sister group to the vertebrates, in the same immunological vein as nematodes and flies, are surprising for what they do not have. None of the genes playing a pivotal role in adaptive immunity in the jawed vertebrates are present. MHC, TCRs, Ig, RAG, and activation-induced cytidine deaminase (AID) genes are all lacking. V-like and C1-like domains are present and VCBPs have been identified (Cannon et al., 2004), and they do have complement components, three TLRs, an expanded family of C-type lectins and FREPs. However, urochordates lack obvious expansions of any gene family highly relevant to vertebrate immunity (Azumi et al., 2003). Based on what we know thus far, genome reduction is the hallmark of urochordate immunobiology.

Even closer to home are the agnathans or jawless vertebrates, lampreys and hagfish, the sister group to the jawed vertebrates or gnathostomes. We now know they lack RAG1 and RAG2 and do not produce TCRs or Ig, however, they have a remarkable ability to make highly diverse variable lymphocyte receptors (VLRs) that consist of somatically re-arranged modules containing leucine-rich repeats (Pancer et al., 2004). It is striking that agnathans and gnathostomes have adopted divergent solutions to the same problem of generation of recognition capability, both of which involve rearrangements of germ-line encoded genes, but in entirely different ways with different starting sets of molecules.

The basic gnathostome immune system, the one most familiar to immunologists, features a close collaboration between innate and adaptive arms. As noted above, relative to some of the invertebrate deuterostomes such as echinoderms or cephalochordates, gnathostome innate immune components are modest in numbers, typically possessing 10–25 TLRs and 20–35 NLRs (Messier-Solek et al., 2010). The gnathostome adaptive immune system features somatic diversification of both TCRs and Ig, requiring for this process RAG1 and RAG2, the former likely derived and modified from a *transib*-like transposon (Fugmann, 2010). The gnathostome immune system works in conjunction with a unique antigen processing and presentation system, the MHC, to limit self-reactivity. It is notable for its specificity, its emphasis on expansion of relevant clones of lymphocytes, and for its memory and capacity to produce heightened secondary responses long after primary stimulation (Litman et al., 2010). The ongoing discovery of new types of immune cells (Neill et al., 2010; Saenz et al., 2010) and novel receptors (Parra et al., 2007) strongly suggests there are more fundamental insights to come with respect to gnathostome immunology. Furthermore, and a point relevant for the general discussion here, there is considerable variability among gnathostomes in how their immune systems function (Flajnik and Kasahara, 2010).

To conclude this overview, it is indeed remarkable that organisms as diverse as cnidarians and humans have some immune architecture such as TLRs (and associated pathways) and NLRs in common. However, it is argued from the examples provided above that at least as compelling are the differences among and within phyla, even among species in a genus. Surprises abound, such as in the unexpectedly complete set of immune genes found in basal cnidarians and the immune genome reductions exhibited by nematodes, arthropods, and urochordates. Even the more familiar examples of conservatism such as TLRs and NLRs in arthropods and vertebrates may have been derived independently (Hughes, 1998; Hughes and Piontkivska, 2008; Zhang et al., 2010). Large lineage-specific gene expansions such as noted for echinoderms, and domain reshuffling such as for invertebrate NLRs (Zhang et al., 2010) have occurred, creating remarkable heterogeneities among and within phyla. Layered on top of this are other forms of innovation such as elaboration of novel signaling pathways and production of associated AMPs in nematodes, distinct antigen recognition and melanin-deposition systems in arthropods, and the emergence of several distinct mechanisms for generating diverse antigen receptors in mollusks, arthropods, echinoderms, cephalochordates, agnathans, and gnathostomes. From this it is concluded that immune systems across and within phyla have a remarkable propensity to generate novelty and distinctiveness. As we learn about the immune systems of more animals, this diversity is bound to increase.

DRIVERS OF IMMUNOLOGICAL NOVELTY

It is hardly surprising that immune systems are so variable. Animals have been extant and diversifying for up to 800 million years (Erwin et al., 2011). They have adopted a diversity of life styles: sessile, colonial predators; inhabitants of extreme environments

dependent on chemosynthetic symbionts; animals that are at times net producers of energy due to their photosynthetic symbionts; pelagic species that live in enormous schools; inhabitants of arid terrestrial environments; social species living in large colonies; and endoparasites that are so modified morphologically as to belie their origins, to name just a few. These different life styles will impose very different exposures to potential parasites.

Similarly, life histories vary radically from wind-dispersed organisms like tardigrades or rotifers that live in ephemeral habitats or that have life spans measured in hours, to sessile filter feeders like marine bivalves that routinely live for over 100 years in the same location. The role of life history traits such as survival rates and reproductive output are predicted to strongly influence the extent and kinds of particular immune responses that might be expected both among and within particular host species (Lee, 2006).

Another factor likely to have influenced immune capability is the nature and extent of commitment to mutualistic symbionts. Some animals have established mutualistic associations with what are essentially monocultures of specialized bacteria (Nyholm and Nishiguchi, 2008; Pais et al., 2008). Others, like ourselves, are dependent on a diversity of both archaeal and bacterial mutualists to which our immune system has made extensive accommodation. Third party symbionts have long played a role in educating, augmenting, and modulating animal immune systems (Turnbaugh et al., 2007). The outcome of host–parasite interactions is often influenced by third party symbionts which probably play a far great role in host defense in many animal groups than customarily realized (Loker, 1994; Welchman et al., 2009; Gross et al., 2009; Eberl, 2010).

The adoption of body plans differing in complexity and mass has also influenced immune system structure and function. It has been argued that evolution of the vertebrate jaw and an accompanying predatory life style introduced parasites into the gut and required a more elaborate adaptive immune system that now typifies gnathostomes (Matsunaga and Rahman, 1998). It has also been argued that the complexities of adaptive immunity could not have evolved in animals with limited numbers of cells or with small size or simplified body architecture (Hauton and Smith, 2007). With respect to body mass, for vertebrates, it has been suggested that the number of B and likely T cells in a clone scale with body mass as does the B cell repertoire (Wiegel and Perelson, 2004). The general point is that the adoption of different habitats, life histories, symbionts, and bodies of differing body mass and complexity are all factors that will influence immune system design and mode of action.

A GENERAL ADVANTAGE OF IMMUNOLOGICAL NOVELTY – BEING DIFFERENT FROM YOUR NEIGHBORS

In addition to the above considerations, all organisms have to contend with another category of symbionts, namely parasites. Because viruses, bacteria, and protists were present before animals arose, all animals from their inception would have had to contend with these parasites. Several modes of transmission of such parasites among early animals of disparate lineages were available, including: intimate proximity of many different kinds of animals (such as on a coral reef), predation, presence of vectors imbibing blood or plant juices containing parasites, and even one parasite serving as a vector for another, as for example a trematode vectoring a bacterium into new animal hosts.

In such a situation, where frequent transfer of parasites was possible among hosts from even disparate phyla, if all emerging animal lineages had the same defenses, it would be possible for an effective parasite that had overcome the defenses of one host group to simply spread into another host phylum. Consequently, having an immune system with distinctive means of antigen recognition and/or novel effector mechanisms would have been a distinct advantage when inevitably confronted with parasites that had evolved in other host groups (**Figure 2**). Being immunologically different increases the odds that parasites from other inhabitants of the same coral reef are not as easily acquired. The notion that a parasite can track and exploit a common host genotype creating an advantage of rareness has been predicted and observed in specific host–parasite systems (Trachtenberg et al., 2003; Wolinska and Spaak, 2009), further suggestive of a similar dynamic favoring distinctiveness or "avoidance of commonness" among members of different host lineages.

LINEAGE-SPECIFIC ANIMAL PARASITES AS ADDITIONAL DRIVERS OF NOVELTY

Once animals began to diversify, a major trend was for some animals to parasitize others. Some animal parasites became wholly or largely committed to particular lineages of animal hosts, in which they subsequently diversified. Such lineage-specific parasites (some examples in **Table 2**) are another general factor expected to drive immunological novelty. These parasites often establish prolonged, intimate, and extensive infections in their chosen hosts that have profound fitness consequences such as castration or death (Lafferty and Kuris, 2009a). Furthermore, given the phylogenetic diversity represented among these parasites, it is not surprising they would evolve novel methods of infectivity. For example, ichneumonid and braconid hymenopteran parasitoids have acquired mutualistic polydnaviruses that function to suppress the immune responsiveness of their hosts and facilitate parasitoid infection (Webb et al., 2009). In contrast, without the aid of viral symbionts, larvae of digenetic trematodes secrete both anti-oxidants and immunosuppressive factors that down-regulate snail host immune components for a period sufficient to enable them to complete their lengthy period of larval development (Hanington et al., 2010a). Therefore, the immune response devised by a particular host lineage afflicted with its own phylogenetically distinct, host specific, and harmful parasites would likely be divergent from the responses mounted by a different host group experiencing its own lineage-specific parasites (**Figure 2**). This is not to imply that only animal parasites have developed lineage-specific associations with hosts, but merely serves to show that specialized parasites can help us understand the origins of immunological novelties.

A ROLE FOR IMMUNITY IN GENERATING PARASITE BIODIVERSITY

The macroevolutionary patterns noted above with respect to novelty in defense strategies among animal lineages could not occur if there were not microevolutionary processes ultimately involved in

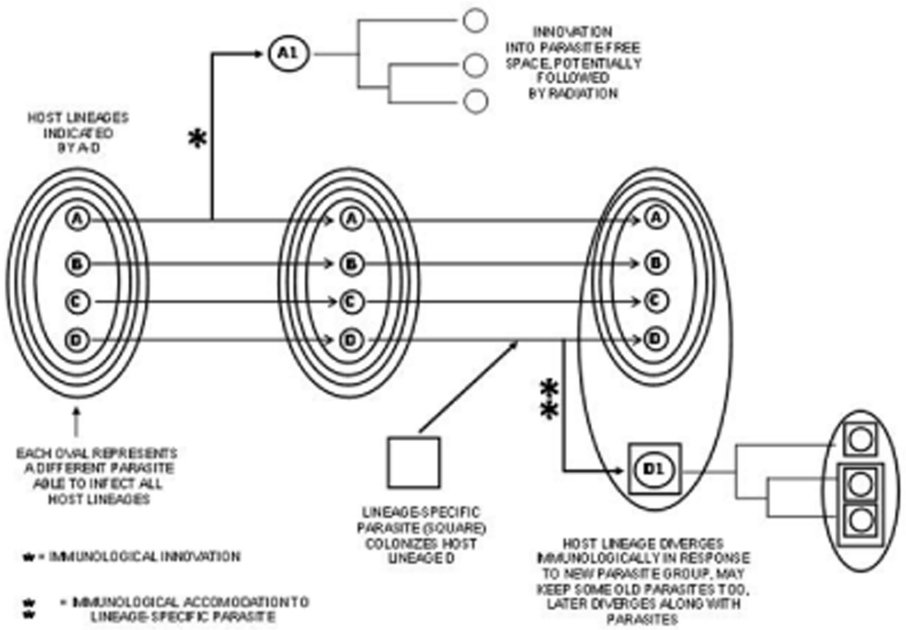


FIGURE 2 | One scenario for early in the divergence of animals is that different lineages (A–D) have fundamentally similar immune systems such that they all are colonized by the same parasites. In the case shown at the top, an immunological innovation occurred in lineage A, that allowed it to resist these parasites. This may have permitted a subsequent radiation in "parasite-free space" in this host lineage. At the bottom, lineage D has acquired a lineage-specific parasite different from those previously experienced. This requires an immunological accommodation that causes the immune system of lineage D1 to diverge. Both the host lineage and the lineage-specific parasites along with them may subsequently diverge.

Table 2 | Examples of parasites that are "lineage-specific" in particular host groups during at least part of their life cycles.

Parasites	Species (*n*)	Hosts
Digenetic trematodes (as larvae)	18,000	Mollusks (usually gastropods)
Unionid bivalves	1,000	Fish
Rhizocephalans	260	Decapod crustaceans
Poecilostome copepods	400	Cnidarians
Sisyridid sponge flies	50	FW sponges and some bryozoans
Hydracarina water mites	5,000	Aquatic insects
Tantulocaridans	30	Crustaceans
Acroceridae	>500	Spiders
Pipunculids	1,388	Leafhoppers and planthoppers
Tetracneminae chalcidoid wasps	815	Pseudococcid insects
Banchinid ichneumonid wasps	1,500	Lepidopteran insects
Ichneumoninae ichneumonid wasps	350 genera	Lepidopteran insects
Aphidiinae braconid wasps	400	Aphids
Conopidae	800	Mostly wasps and bees

This list is not exhaustive and merely serves to illustrate the concept that particular host lineages acquire unique parasites that are likely to have distinctive methods of infectivity that could influence how their host's immune systems are shaped by selection. Although cases where members of the parasite groups identified colonize hosts outside the indicated host lineage certainly occur, they do not negate the idea that the host groups indicated above have been far more affected in aggregate than a host lineage containing an isolated member harboring a peculiar outlying parasite. Also, for some huge groups, such as the ichneumonid wasps, although when viewed more inclusively they infect much broader groups of hosts (such as insects or terrestrial arthropods), the point remains they have had relatively little impact on other major host lineages beyond the insects.

generating them. These microevolutionary processes occur below the species level, within and among populations of either host or parasite species, and might culminate in speciation of either participant. Speciation may be accompanied by colonization of new habitats, and further divergence to create major new lineages. Starting with the process of parasite diversification, the sections

below discuss how the involvement of immunology in microevolutionary processes could lead to events that can help explain the macroevolutionary patterns discussed above.

The following quotes outline some of the key ideas for how immunity can play a role in generation of parasite diversity.

> "For a pathogen, the selective pressures arising from the host immune system are a major influence on the evolution of mechanisms of infectivity and of immune-recognition avoidance" (Acevedo-Whitehouse and Cunningham, 2006).
>
> "In parasitism an essential factor in survival is immune escape, which allows a parasite to resist host attack. Immune escape is a mechanism for reducing gene flow at the level of the compatibility filter because its result is assortative survival (Combes and Théron, 2000) as opposed to assortative mating." (Combes, 2001, p. 154).
>
> "The stepwise coevolutionary process results in extreme specialization and complex defense mechanisms. . .specialization is likely to increase the rate of speciation that may occur in both host and parasite" (Price, 1980)
>
> "Host specificity thus is an ideal prerequisite for rapid speciation" (Mayr, 1963).

IMMUNITY IS OFTEN RESPONSIBLE FOR THE SPECIALIZATION TO PARTICULAR HOST SPECIES SHOWN BY PARASITES

Parasites are often cited as examples of specialists because they have limited ranges of host species, often with restricted ranges of habitats within their chosen hosts. For example, the lineage-specific parasites mentioned above, often show considerable specificity to particular species or genera within their adopted host lineages. Most animal parasites are host specific (Poulin and Keeney, 2008; Agosta et al., 2010), but this by no means is to suggest that generalists do not occur: parasites like *Schistosoma japonicum*, *Toxoplasma gondii*, *Borrelia burgdorferi*, or the rabies virus routinely infect a remarkably broad range of host species. Using molecular techniques to identify parasites, species formerly considered to be generalists have in some cases been shown to be complexes of cryptic, host specific species (Poulin and Keeney, 2008).

As eloquently documented by Combes (2001), both encounter and compatibility filters operate to restrict the spectrum of host usage. Encounter filters pertain in situations where host and parasite live in different geographic localities, have different ecological circumstances, or where host or parasite behavioral tendencies preclude contact. The compatibility filter refers to barriers imposed by the host that prevent infection once contact has occurred. The compatibility filter includes both physiological and biochemical suitability of the host to support the parasite, and the active defense provided by the immune system.

Encounter filters are undeniably important in restricting parasite host range. Many examples of emerging infections (Goss et al., 2009; Gray et al., 2009; Pfeffer and Dobler, 2010) owe their emergence to a change in the encounter filter such that a new combination of parasite (often including its vector) and host are juxtaposed (Daszak et al., 2000; Parrish et al., 2008; Weissenbock et al., 2010). Emerging diseases are an indication that parasite infectivity is not always dependent on a long accommodation to a particular host species: lack of contact may have prevented prior infections. Similarly, experimental infections of new hosts with parasites essentially bypass the encounter filter, and are sometimes successful (Poulin and Keeney, 2008), affirming the reality and importance of the encounter filter.

With respect to the compatibility filter, a role for unsuitability should not be discounted and could be manifested in several ways, such as a lack of receptors needed for efficient viral entry into cells (Parrish et al., 2008), lack of appropriate structures for parasite attachment (Tompkins and Clayton, 1999), or by a general failure to provide the biochemical environment needed for the parasite to survive (Sullivan and Richards, 1981).

Although unsuitability is probably underappreciated with respect to preventing infections, relying on the possibility of being unsuitable is not a cogent defense strategy. The importance of active immunity to the compatibility filter is illustrated by several lines of evidence.

(1) As illustrated by HIV, when the immune system is compromised, the door is opened to opportunists that themselves can become life-threatening (Holmes et al., 2003).
(2) Genetic defects in the immune system, such as with TLRs, are associated with increased susceptibility to several different pathogens (Qureshi and Medzhitov, 2003).
(3) Experimental exposures of hosts to parasites they have not previously encountered often fail (Bowen, 1976; Bozeman et al., 1981; Vidal-Martinez et al., 1994; Philips and Clarkson, 1998; Sapp and Loker, 2000; Duke, 2004; de Vienne et al., 2009; Giraud et al., 2010) or the parasite replicates poorly or is inefficiently transmitted in a new host (Komar et al., 2003; Parrish et al., 2008). **Table 3** provides examples of parasites in novel hosts that are engaged and killed by immune responses.
(4) Host defense genes are under strong selection and are conspicuous for evolving quickly (Sackton et al., 2007; Viljakainen et al., 2009; Barreiro and Quintana-Murci, 2010; Schulte et al., 2010).
(5) The extraordinary diversity of strategies undertaken by parasites to evade, manipulate, or suppress the immune system is testament to the impact of immunity on their success (Schmid-Hempel, 2008). These evasive strategies have been shown in some cases to be specific with respect to the particular hosts involved (**Table 4**), providing a mechanistic basis for the connection between immunity and parasite specialization.

Given that the consequences to a parasite for engaging the compatibility filter of an atypical host could be disastrous and result in its death, strong selection to avoid colonization of such hosts would be expected in some cases. Similar considerations may also apply to the host as well, as mounting an immune response can be costly and detrimental (Graham et al., 2011). This means that some avoidance behaviors attributed to the encounter filter may actually be a consequence of the operation of a strong host immune response (Kuris et al., 2007; Keesing et al., 2009).

To conclude this section, specialization and attendant host specificity is a central, emergent property of parasitism and has multiple underlying determinants, involving both encounter and compatibility filters. The ubiquity of host defenses and the evidence that they often eliminate novel parasites argue that host immune systems play a critical role in limiting parasite host

Table 3 | Examples of colonizing parasites, or parasites placed in novel hosts, that are killed or limited by immune responses.

Infection of the crab *Pachygrapsus marmoratus* with the rhizocephalan barnacle *Sacculina carcini* results in melanization of larvae in thoracic ganglia (Kuris et al., 2007)
Antibody/factor that activates complement in serum of the non-host *Raja radiata* kills the tapeworm *Acanthobothrium quadripartitum* whereas larvae survive in serum of the normal host, *Raja naevus* (McVicar and Fletcher, 1970)
Destruction of cercariae of avian schistosomes in the skin of mammals associated with a mixed Th1/Th2 lymphocyte cytokine response followed by more polarized Th2 response upon repeated exposures (Horak and Kolarova, 2005)
Encapsulation of hymenopteran parasitoids by hemocytes of non-permissive insect hosts (Schmidt et al., 2001)
Lysis of the trypanosome *Trypanosoma brucei brucei* by apolipoprotein L-1 in serum of humans who are refractory to this subspecies (Wheeler, 2010).
Disruption of the Erk-STAT1 signaling pathway allows cross species transmission of the normally rabbit-specific myxoma virus to mice (Wang et al., 2004)
Animal handlers who were exposed to a new coronavirus developed antibodies to the new virus and did not develop clinical infections (Guan et al., 2003)
Species specific forms of APOBEC3G and other innate, intracellular defense components, can prevent cross species transfer of lentiviruses (Mangeat et al., 2004; VandeWoude et al., 2010)

Table 4 | Examples of parasite immune evasive factors that are host specific in their action.

A staphylococcal complement inhibitor that specifically blocks human C4b2a and C3bBb, interfering with additional C3b deposition through classical, lectin or alternative pathways (Rooijakkers et al., 2005). Sung et al. (2008) found several genes conserved in all *Staphylococcus aureus* isolates from humans were variable or missing in one or more animal isolates, including *fnbA*, *fnbB*, and *coa*.
Human and murine chlamydial infections depend on different virulence factor genes that coevolved to counter host species specific IFN-γ-mediated effector responses mounted by the particular host species (Nelson et al., 2005).
Orf virus encodes a secreted protein GIF that binds to and inhibits GM-CSF and IL-2 of ovines but not humans or murines, consistent with the idea that Orf virus is evolutionarily adapted to sheep as its primary host (Seet et al., 2003).
Different strains of influenza A virus likely have NS1 genes adapted to antagonize the IFNα/β antiviral system of their specific host species (Garcia-Sastre, 2006).
In a review of the interactions between monogenean parasites and their fish hosts, Buchmann and Lindenstrøm (2002) concluded that "immune evasion mechanisms are probably a main factor in host specificity."
Rosengard et al. (2002) noted that the smallpox inhibitor of complement enzymes (SPICE) is nearly 100-fold more potent than the vaccinia homolog in inactivating human C3b and sixfold more potent at inactivating C4b, providing evidence for how variola proteins are particularly adept at overcoming human immunity relative to vaccinia.
The host specificity of three species of *Bacillus* (*B. cereus*, *B. thuringiensis*, and *B. anthracis*) is determined by the presence of virulence plasmids that determine the type of particular virulence factors produced (Gohar et al., 2005).

ranges and thereby at least in part dictate the specialization so characteristic of parasitic organisms. Also in support of this claim is that some patterns of parasite host specificity can be attributed to the operation of specific immune evasion strategies, and that such strategies are pervasive among parasites.

DOES SPECIALIZATION DICTATED BY IMMUNITY ACTUALLY LEAD TO PARASITE SPECIATION?

As exemplified by the parasites indicated in **Table 2**, in addition to being specialized to exploit particular host groups, they are remarkably diverse in species, as are the lineages of many parasites. One of the potential consequences of specialization, including in an immunological context, is diversification in species, of both parasite and host lineages. The mechanisms involved in promoting speciation remain a matter of active investigation and for the discussion below, the purpose is to indicate that immunological phenomena may play a role in this process that deserves further attention. One prominent mechanism of parasite speciation is switching to a new host species, and the role of accommodation to the immune system of new hosts to permit such switches is discussed in a separate section below.

A second mechanism is co-speciation. For some parasite groups closely tied to their hosts and with limited options for colonization of new hosts, such as sucking lice on burrowing mammals (Light and Hafner, 2008), speciation may occur if the hosts upon which they are found themselves speciate, often following a physical separation of populations of the host species. In such cases, persistence of new daughter parasite species should be favored by the fact that the parental species had already achieved successful accommodation to the parental host species. Although the actual role of specific immune phenomena in influencing the persistence of incipient parasite species in co-speciating systems is not known, an important underlying role for a preexisting immunological accommodation between parental host and parasite species that also favors persistence of the new parasite species should not be discounted.

Another important way in which specialization dictated by immunological phenomena can increase the probability of formation of new parasite species is by promoting intraspecific diversification. The interactions between a particular host and parasite species can be expected to be variable across space (Wood et al., 2007). Parasite abundance will vary across local scales, possibly because of the variable presence of other hosts needed to complete

its life cycle (Byers et al., 2008). Other parasite species impacting the same host may be present or absent, such that the host experiences different overall parasite pressure in different locations within its range. Furthermore, the host itself will be variable across its range owing to its responses to other local circumstances. All of these factors conspire to create heterogeneities with respect to how the host potentially mounts immune responses to the parasite (Kraaijeveld and Godfray, 1999; Thomas et al., 2000; Lindstrom et al., 2004; Kalbe and Kurtz, 2006; Blais et al., 2007; Bryan-Walker et al., 2007; Scharsack et al., 2007; Matthews et al., 2010). Variability within parasite species with respect to infectivity to their hosts is a pervasive phenomenon (Carius et al., 2001; Schulenburg and Ewbank, 2004; Seppala et al., 2007; Vorburger et al., 2009) and this is likely driven in part by variations in immune evasive measures taken by parasites (Hammerschmidt and Kurtz, 2005; Cornet et al., 2009; Vorburger et al., 2009). These dynamics are compatible with general theoretical predictions that parasite variation is driven by immunity, and hosts themselves are variable with respect to immunity due to pressure posed by parasitism (Frank, 2002). Immune responses are drivers for parasite diversification (Summers et al., 2003; Lazzaro and Little, 2009; McKeever, 2009). An overall increase in intraspecific genetic variability, with that variation partitioned into regionally differentiated parasite populations accommodated to local host populations provide rich opportunities for further divergence.

One possibility for further divergence is that local adaptation to host immunity could potentially lead to "assortative survival" (Combes and Théron, 2000; Combes, 2001, p. 154), meaning that the only options for mating (parasites frequently seek mates and undergo sexual reproduction within their hosts) occur between individuals able to survive in hosts with similar immune capability that are vulnerable to the same parasite immune evasive capacity (Giraud et al., 2010). This would further accentuate local differentiation of parasites, potentially leading to ecological speciation of the parasite, particularly if subsequent gene flow is prevented by failure of immune adapted parasites to thrive in hosts (from other localities) with different immune capacities.

It must be noted that fluctuations in local patterns of abundance of hosts and parasites may diminish the strength of local adaptation and promote gene flow such that speciation is precluded (Lazzaro and Little, 2009), and that in general, evidence that parasite speciation is effected by underlying immune mechanisms is sparse. However, given the need for parasites to accommodate to a host's internal environment and that a host species is likely to be confronted with varying parasite pressure, it seems host immune responses will favor diversification in parasite lineages. To add an additional dimension to the concept that spatially variable relationships favor parasite diversification, it has recently been argued a general underlying mechanism favoring biological diversification is the existence of localized parasite-coevolutionary races that select hosts to prefer immunologically similar conspecifics and to avoid out-group individuals, thereby minimizing the risk of exotic disease acquisition (Fincher and Thornhill, 2008). By promoting strong intraspecific diversification within host species based on avoidance of contagion, this mechanism has also been predicted to lead to parasite diversification (Fincher and Thornhill, 2008).

To conclude this section, all of these observations fit into the more generalized geographic mosaic theory of coevolved relationships (Thompson, 2005): in this particular case, local adaptations based on immunological accommodation of host and parasite can lead to diversification of parasites and potentially speciation.

A ROLE FOR IMMUNITY IN HOST SWITCHING AND PARASITE DIVERSIFICATION

An important way diversity in parasite lineages is generated, one that has increasingly come to light from molecular phylogenetic studies and the study of emerging diseases, is via switching to new hosts (**Table 5**). Although successful host switching cannot be a ubiquitous process, otherwise we might expect to find only a few species of generalist parasites instead of a predominance of host specific parasites, clearly it has been an important factor historically and examples continue to be regularly documented. *A priori*, it seems logical that most successful switches would be to hosts not phylogenetically distant from the original host species. Such close range switches are likely favored by a degree of phenotypic plasticity and preadaptation (exaptation) of the parasite and its use of phylogenetically conserved resources in the new host species such that new attributes are not needed to overcome a new host's immune system (de Vienne et al., 2009; Agosta et al., 2010). For example, in a study of host switches in bats involving the fast-evolving RNA virus causing rabies, the success of cross species transfers diminished as the phylogenetic distances among the hosts involved increased (Streicker et al., 2010).

It is also possible for parasite switches to occur when the original and new host species are not closely related (Brant and Loker, 2005). This has been observed for emerging human parasites for which ungulates and carnivores were more likely originating host species than primates, and it was concluded that an already broad host range as opposed to the phylogenetic relatedness of the new and old host species was the more important factor dictating success in interspecific parasite jumps (Woolhouse et al., 2005).

In any case, a host switch can lead to a speciation event if the parasite in the new host becomes isolated from the founding stock, or can have even more profound effects if the switch is into a new host lineage and leads to the founding of a diverse new parasite lineage (Agosta, 2006; Janz et al., 2006; Hoberg and Brooks, 2008; Martinsen et al., 2008; Refrégier et al., 2008; Winkler et al., 2009; Giraud et al., 2010; Nyman, 2010). The isolation of the switching parasite from the founding stock is reinforced because even a single individual may be able to establish a new population and because differing ecological circumstances of the new host may preclude mixing of parasite progeny with the source population: the new parasites may never get back into the original host and thus mate only with other parasites in the new host. Assortative survival and mating would again be factors favoring isolation of the founding parasites. Host switching is also relevant to the idea stated above that if a host lineage acquires a new parasite, it may then have its immune system substantially altered. Particularly if the parasite is successful and radiates, then the immune system of the new host lineage may be forced to diverge to adjust to the new challenge.

Table 5 | Examples of parasite groups exhibiting hosts switches likely to have played a major role in diversification of that group.

With respect to *Plasmodium* and related genera of blood parasites, major clades are associated with shifts into different families of dipteran vectors, and the *Plasmodium* species of birds and squamate reptiles show evidence of repeated switching back and forth (Martinsen et al., 2008).
Major lineages within the blood fluke genus *Schistosoma* are defined by acquisition of different genera of even families of snail intermediate hosts, by host switching (Morgan et al., 2003). Long-range host shifts involving acquisition of both new snail and vertebrate hosts appear to have occurred during the history of schistosomes (Brant and Loker, 2005).
Zietara and Lumme (2002) note that as many as 20,000 species of the monogenean genus *Gyrodactylus* may exist, and note that in a study of one subgenus (*Limnonephrotus*) that several host switch events were statistically confirmed, including into new host families, supporting the idea that host switching is a means to drive innovation and adaptive radiation in these ectoparasites.
It appears that host switching has been common in trypanorhynch tapeworms, one of the most diverse and abundant groups of metazoan parasites of elasmobranchs (Olson et al., 2010)
Coronaviruses have likely undergone several host switches, between mouse and rat, chicken and turkey, birds and mammals, and between humans and other mammals (Rest and Mindell, 2003).
Braconid wasps of the subfamily Euphorinae have undergone extensive host switching among phylogenetically distantly related insect host groups, often followed by adaptive radiations of the parasitoids within a particular host lineage (Shaw, 1988).
"Infection of a novel host is the most frequent cause of fungal emerging disease" (Stukenbrock and McDonald, 2008; Giraud et al., 2010)

Several examples from the literature of emerging infectious disease indicate that switches are often favored by changes in the encounter and not the compatibility filter (Woolhouse et al., 2005). Ecological circumstances have exposed humans to a parasite they previously did not encounter. Such examples of host switches, particularly if the new host is distantly related to the original host, would seem to argue against the points made in the preceding sections regarding the importance of parasite accommodation to the idiosyncrasies of their host's immune system. If immunity is important in restricting parasite host ranges, how can such switches occur?

First, these conspicuous successes need to be weighed against all the encounters between novel parasite and host combinations that fail and therefore go unnoticed, which is likely a far more frequent outcome (de Vienne et al., 2009; Tunaz and Stanley, 2009; Giraud et al., 2010; see also **Table 3**). In cross species transfers of rabies into bats, the vast majority are dead ends: they did not establish sustained infections (Streicker et al., 2010). Although some of the failures could be explicable because of less frequent contact among more distantly related bats (the encounter filter), increasingly divergent defense systems leading to higher levels of innate resistance were also invoked as an explanation (Daszak, 2010; Streicker et al., 2010). The role of immune systems in preventing such infections would be easy to underestimate because the result is a failed experiment that in all probability we never even knew had happened. In a similar vein, a survey of field-trapped insects in turkey revealed that 98% exhibited some kind of melanotic hemocyte nodule (Tunaz and Stanley, 2009). Such host reactions provide a convenient historical record of previous parasite encounters (Kuris et al., 2007). It was concluded that insects are regularly challenged by infections from which they recover. The action of innate immunity in routinely preventing acquisition of new parasites is probably considerable and easy to underestimate.

Secondly, host switches would be favored if the new parasite, as exemplified by HIV, directly attacks the host's immune system and compromises it, or if the new host is immunocompromised by some other means. Diminished levels of immune competence can occur for several reasons, including ones likely to have been in operation throughout animal evolution. One possible means is that the host's indigenous parasites might use immunosuppression to favor initiation and persistence of their own long-term infections (**Table 6**) and thereby facilitate colonization of that host by other parasites (Krasnov et al., 2005). An intriguing possibility is that the successful colonization of a host species by one or more immunosuppressive parasites might then favor colonization by opportunist parasites, resulting in an unusually diverse parasite fauna supported by that host. The large number of species of larval digenetic trematodes known to be supported by some snail species (Loker et al., 1981; Lafferty and Kuris, 2009b) might exemplify this possibility.

High host density, stressful thermal (Bruno et al., 2007) or oxygen regimes (Aeby and Santavy, 2006), and even mating (Rolff and Siva-Jothy, 2002) are some natural situations that can also lower immune competence. To this list can be added a number of human-imposed immune stressors including crowded aquaculture conditions (Flemming et al., 2007), use of harmful chemicals on fields or roads (Rohr et al., 2008; Karraker and Ruthig, 2009), altered diets (Sahu et al., 2008), elevated or altered environmental nutrient conditions (McKenzie and Townsend, 2007; Wedekind et al., 2010), and deliberate implementation of immunosuppressive therapies. Not only can these situations lower host immune competence, they may also increase parasite virulence and thereby alter probabilities of successful colonization in a new host species (Wedekind et al., 2010).

A switch into an immunocompromised individual of a new host species is likely to be temporary and not lead to speciation unless the new parasite can better adapt to its new host, at the same time with minimal gene flow occurring with conspecifics from its ancestral host. Availability of populations of similarly immunocompromised new hosts that allow continued transmission and adaptation of the new parasite host could favor divergence from the founding parasite and speciation.

Table 6 | Examples of immunosuppression by one parasite that could favor acquisition of new parasites, and potentially an eventual speciation event.

Varroa mites (*Varroa destructor*) in honey bees (*Apis mellifera*) suppress the activity of several immune-related genes (encoding both antimicrobial peptides and enzymes) and favor higher infection titers with the deformed wing virus (Yang and Cox-Foster, 2005).
Drosophila simulans infected with *Wolbachia* have reduced ability to encapsulate eggs of the parasitoid *Leptopilina heterotoma* (Fytrou et al., 2006).
The malaria parasite *Plasmodium gallinaceum* suppresses the encapsulation response of the mosquito *Aedes aegypti* (Boëte et al., 2004).
Two acanthocephalan parasites *Pomphorhynchus laevis* and *Polymorphus minutus* both have the effect of decreasing the standing level of immune defense (as measured by reduced phenoloxidase enzyme activity) in their local gammarid hosts, *Gammarus pulex*, but not in their more recently introduced host *Gammarus roeseli* (Rigaud and Moret, 2003).
Hymenopteran parasitoids induce immunosuppression in their host insects in part by the injection of polydnaviruses which target and inhibit both cellular and humoral components of the host response (e.g., Labropoulou et al., 2008) and the parasitized hosts become increasingly susceptible to opportunistic infections by viruses (Rivkin et al., 2006), bacteria (Shelby et al., 1998), and other parasitoids (Guzo and Stoltz, 1985).
As noted by Lie (1982), interference by trematode larvae with gastropod defense responses appears to be a common consequence of infection (Hanington et al., 2010a), and the presence of one trematode infection can facilitate the colonization of an infected snail by trematodes that would not ordinarily be able to infect that species of host (Walker, 1979; Southgate et al., 1989).
HIV in people was associated with parasites that rarely if ever had been implicated in causing human disease including microsporidia, cryptosporidia, JC virus, and *Mycobacterium avium* (Kovacs and Masur, 2008).
Studies of parasite communities suggest that taxonomic distinctiveness of ectoparasites and endoparasite richness are positively correlated across species of rodent hosts, indicative of immune responses to some parasites depleting energy reserves and facilitating colonization by others (Krasnov et al., 2005).

A ROLE FOR IMMUNITY IN DRIVING HOST DIVERSITY?

How might immunological phenomena influence the degree of diversity shown among the hosts contending with parasites? As highlighted below, parasite pressure clearly favors immunological diversification at microevolutionary scales. Whether this diversification contributes in a meaningful way to host speciation remains controversial, but has attracted considerable attention and is gaining some support, as discussed below.

As early as Haldane (1949), and as more recently underscored (Frank, 2002; Lazzaro and Little, 2009), it has been recognized that parasites drive polymorphisms in host immune competence, particularly in variable environments. This can occur by balancing (Wegner, 2008) or disruptive selection (Duffy et al., 2008; Matthews et al., 2010). High levels of polymorphism are found in several genes of both the innate and adaptive immune systems (Hill, 2001; Trowsdale and Parham, 2004; Acevedo-Whitehouse and Cunningham, 2006). MHC genes show a predominance of non-synonymous over synonymous mutations in their peptide-binding regions, and have extensive allelic diversity, indicative of strong role of selection, generally considered to be mediated by parasites. For example, parasite pressure is believed to have promoted maintenance of high MHC diversity in sticklebacks (Wegner et al., 2003) and Atlantic salmon (Dionne et al., 2007). In humans, regional differences in HLA class I diversity has been associated with intracellular parasite richness (Prugnolle et al., 2005). Across species comparisons of rodents have associated helminth species richness with increased MHC class II polymorphisms (de Bellocq et al., 2008). MHC genes are also known for their role in mediating mate choice through olfactory systems in humans (McClintock et al., 2002), rodents (Sommer, 2005), and fishes (Landry et al., 2001; Milinski et al., 2005).

The evidence supporting the idea that variability in immune response driven by parasites can be a factor favoring speciation of host lineages is mostly correlational, but is supported by a growing body of literature. One general factor favoring this is the development of strong local immunological accommodations of particular host populations to the distinctive parasites they encounter, such that across a broader host range different host populations differ significantly with respect to the nature of their responses (Wheatley, 1980; Blais et al., 2007; Scharsack et al., 2007; Matthews et al., 2010). For example, malaria might be encountered in some but not all parts of the host range where appropriate mosquitoes were present, or different suites of parasites might be encountered in different foraging habitats such as rivers or lakes (Eizaguirre et al., 2011). Differences among populations of the same host species with respect to their immune defenses have been noted in *Drosophila* against parasitoids (Kraaijeveld and Godfray, 1999), Darwin's finches (Lindstrom et al., 2004), sticklebacks coping with eye flukes (Kalbe and Kurtz, 2006), and marine amphipods infected with trematodes (Thomas et al., 2000; Bryan-Walker et al., 2007). Differentiation resulting from spatially variable antagonistic interactions with parasites would in this case provide the substrate for further diversification of host lineages (Thompson, 2005, 2009).

The impact of this local adaptation could be augmented by assortative mating mediated by sexual selection to favor further divergence (van Doorn et al., 2009). According to this line of thought, those locally adapted hosts that best resist parasites are able to elaborate ornaments that signal superior resistance to parasites, such that local mates preferentially breed with them. If these hosts were transplanted to other host populations with their own distinct parasite challenges, they would not be as resistant, their sexual ornamentation would suffer, and they would not be selected for mating. Thus a combination of natural and sexual selection could favor divergence of new host species.

Parasites can exert strong selection on traits known to affect mate choice (Hamilton and Zuk, 1982; Poulin and Thomas, 1999; MacColl, 2009) and in some cases the genes involved also have immune functions, such as genes of the MHC (Milinski et al., 2005; Milinski, 2006; Eizaguirre et al., 2009, 2011). MHC genes

have been considered as possible "magic traits" in sticklebacks, influencing both defense and mate choice (Matthews et al., 2010). MHC divergence in a closely related and sympatric pair of cichlid species from Lake Malawi has been proposed to be a result of local host–parasite-coevolutionary associations, and to have influenced odor-mediated mate choice, and ultimately to have favored speciation (Blais et al., 2007).

Diverging cichlids of the genus *Pundamilia* in Lake Victoria provide another example of how local adaptation and assortative mating (both potentially influenced by immune traits) may work together to promote host speciation. In this system, females have a preference for conspicuously colored males. These bright colors seem to be reliable indicators of male fitness, including resistance to parasites. Conditions of light penetration favor blue males in shallow depths and red males in deeper waters. The parasites encountered at different depths also vary in density and composition such that habitat-specific defenses could occur. Females from the depths prefer brighter red males whereas those from shallow water prefer brighter blue males, potentially leading to divergence, with visual cues playing a key role. This example points out the immunology may work in concert with a number of other forces such as predator avoidance or dietary preferences which all conspire to reinforce divergence of the two incipient species by visual means (Maan and Seehausen, 2010). Mating among hosts between different locations would break down these differences, but might be disfavored if the progeny had reduced resistance to any local set of parasites.

"Infectious speciation" as exemplified by interactions between *Drosophila* species and inherited, endosymbiotic *Wolbachia* bacteria provides another possible mechanism for the involvement of immune processes in host speciation. For a group of six related species in the *D. paulistorum* complex, each species has its own distinct host specific obligatory and mutualistic *Wolbachia* with which it has achieved accommodation. This accommodation likely involves suppression of immune pathways involving apoptosis of infected cells. In hybrids, the *Wolbachia* over-replicate and cause embryonic inviability and male sterility, suggesting the unique host accommodation has been lost. In addition to such post-mating isolation, it has also been shown that females can detect and will reject males harboring the wrong symbiont, thus further reinforcing isolation (Miller et al., 2010).

Among animal hosts, hybrids are often more susceptible to parasites than parental species (see overviews provided by Fritz, 1999; Wolinska et al., 2008), potentially favoring isolation of the parental host species, but other outcomes have also been noted and the responsiveness of hybrids to infection recorded 1 year might differ from those reported the next. This implies that the interactions between hybrids and the parasites they experience exhibit complex temporal dynamics and that the parasites themselves have undergone complicated changes as a result of their hosts' hybridization history that have not been sufficiently investigated (Detwiler and Criscione, 2010). Certainly some studies suggest that isolation of incipient parental species might be reinforced by a breakdown of co-adapted immune gene complexes among their hybrids. Also, in some cases, the act of hybridization contributes to the formation of new species by allopolyploidy, as has been postulated for anurans of the genus *Xenopus*. Hybrids in this case often have increased resistance for parasites, potentially providing a selective advantage to favor the persistence of new species of recent hybrid origin (Jackson and Tinsley, 2003).

Another hypothesized mechanism favoring diversification of host lineages is that localized interactions with particular parasites allows immunological accommodation to them, resulting in strong preference for interactions with individuals with similar immune accommodation and philopatry (limited host dispersal). This is coupled with avoidance of out-group individuals that might lead to introduction of exotic parasites (Fincher and Thornhill, 2008). Diverse parasite populations are thus hypothesized to drive diverse host populations, and ultimately speciation, leading to a general correlation between host and parasite biodiversity (Fincher and Thornhill, 2008).

THE ROLE OF IMMUNOLOGY IN EXTINCTION EVENTS

Extinction too is a macroevolutionary process, including the major pulse of extinction events currently underway. Habitat destruction, human overpopulation, industrialization, threats from introduced species, emerging diseases, and global climate change, have lead to predictions that up to 50% of all species will be lost in the next 50 years (Pimm and Raven, 2000; Koh et al., 2004; Thomas et al., 2004). The study of immunology is relevant to extinction in at least three broad contexts outlined below.

First, as noted above, the interactions between parasite and host often lead the parasite to specialization and host specificity which may in part be dictated by interactions with the host immune system. It has long been argued that specialization leads to a greater probability of extinction because if the host on which the specialized parasite is dependent undergoes severe population fluctuations or itself goes extinct, the parasite will soon follow: a co-extinction event has occurred. Co-extinctions involving pairs of mutualists or host–parasite units may be the most common forms of loss of biodiversity (Koh et al., 2004; Dunn et al., 2009). By comparison, a generalist parasite able to exploit alternative hosts would have a greater likelihood of surviving under similar circumstances. The evolutionary trajectories taken by parasites have been much debated, and specialized parasites have been shown to give rise to large lineages of specialized, or even to more generalized descendents (Johnson et al., 2009). However, host specificity remains a salient feature affecting the odds of extinction (Koh et al., 2004; Poulin and Keeney, 2008).

Second, an inescapable feature of the modern world is the frequency with which invasions of exotic species occur (Torchin et al., 2002). Invasive species might be either hosts or parasites (possibly including their vectors), and all can present dire and often unpredictable consequences, including extinction, for indigenous organisms (Wyatt et al., 2008). A role for immunology in influencing invasions can occur in at least three ways:

(a) Introduced host species often leave their native parasites behind and although they are likely to be colonized by parasites in their new environment, in some cases this colonization is slow to occur, such that they experience relatively low parasite burdens for a long time. Insofar as immune responses are costly to mount and the harmful effects of parasites are avoided, the invading species may enjoy a distinct advantage

over its competitors in its relative freedom from parasites, particularly if they can adopt relatively low cost defense measures against their maladapted parasites (Lee and Klasing, 2004). The relative inability of parasites to colonize the new intruder is a testament to the specialization often required to achieve infectivity, as noted above.

(b) If an introduced host is accompanied by some of its indigenous parasite fauna, then related hosts in the newly colonized area may then have to contend with a parasite to which they are not accustomed, a process that may take decades to achieve (Hedrick et al., 2003; Taraschewski, 2006). This is particularly likely to cause problems if the colonized area is an isolated habitat like an island where hosts have simply not encountered comparable parasites previously. The devastating impact of the introduction of mosquitoes followed by avian malaria and avian pox on the endemic honeycreepers of the Hawaiian islands, likely causing both extinctions and slowing recovery of still extant species, is an iconic example (Atkinson and Samuel, 2010). Islands often favor speciation yet the subsequent colonization of mainland habitats by island species is likely to be limited and unsuccessful due to active transmission of parasites there that a migrant is unable to handle immunologically.

(c) In some cases, invasion of a parasite can occur even without the benefit of a simultaneous introduction of its indigenous host. An example is the eel swimbladder nematode *Anguillicola crassus* which has been introduced from the Orient into Europe where it provokes intense tissue reactions from European eel species (Taraschewski, 2006). By contrast, Oriental eels mount immune responses that prevent a high and robust population of worms from building up.

All of these examples centered on introductions have macroevolutionary implications as they might lead to expansion and radiation of hosts and their parasites into new habitats, or may directly cause extirpations of indigenous hosts (and possibly their co-adapted parasites as well). Immunological phenomena provide proximal causal explanations. Once again, as with the discussion above regarding host switching, there may be significance to what we do not see as well: many parasite introductions fail because the colonist is unable to breach indigenous host defenses and host introductions fail because they are ill-prepared for indigenous parasites. For example, the introduction of the American rainbow trout *Oncorhynchus mykiss* into Europe failed because they encountered the parasite *Myxobolus cerebralis* which is normally transmitted by the indigenous brown trout, *Salmo trutta*: rainbows succumbed to whirling disease in Europe (Hedrick et al., 2003). When this parasite was introduced into North America, brown trout (themselves also introduced) were already well-adapted to it but indigenous rainbows and other salmonids were not and have suffered outbreaks of whirling disease as a consequence.

A third broad context in which immunology becomes relevant to extinction, and one that is much in evidence today, is the role played by diminution of diversity in key immune loci such as the MHC. This can result in vulnerability of endangered host species to general parasite attack and thus extinction (Radwan et al., 2010). Loci other than the MHC may also play major roles in dictating susceptibility to particular groups of parasites (Behnke et al., 2003), and polymorphism in non-MHC genes are relevant in resistance to both tuberculosis (Ottenhoff et al., 2005) and malaria (Hill, 2001). MHC genes are estimated to account for only about half of the genetic variability for resistance traits (Acevedo-Whitehouse and Cunningham, 2006).

Regardless of the immune locus involved, the role of random drift in diminishing allelic diversity in bottlenecked or fragmented host populations would seem to increase the risk of successful parasite attack and greatly increase extinction probabilities. There is a need to determine if other factors like genome-wide inbreeding depression are more important in causing extinction, to understand why some species depauperate in MHC diversity seemingly continue to thrive (Hedrick, 2003; Radwan et al., 2010), for increased transfer of information from the study of the main immunological models to endangered species, and to study other polymorphic genes involved in effective defense (Acevedo-Whitehouse and Cunningham, 2006).

CONCLUSION AND FUTURE STUDIES

A picture of the structure and function of immune systems across animal phyla is slowly emerging, but so many organisms remain unstudied we lack perspective on how representative our current picture is. Do all tunicates, for example, reveal evidence of pronounced genome reduction with respect to immune function, or is this a characteristic of just the few species studied to any extent? What kind of pressure from parasites or otherwise has driven the expansion of all the innate immunity genes evident in organisms like sea urchins, and how do these animals regulate and orchestrate effective responses given the multiplicity of defense molecules they possess? In their 100+ years life span, how frequently must marine bivalves mount immune responses and how do they avoid the problem of fast-evolving parasites from "locking on" and overrunning them? In the meantime, it is clear that the approaches taken to achieve immune defense are diverse, frequently innovative and often capable of generating diverse repertoires of defense molecules that blur the distinctions commonly made to distinguish adaptive from innate immunity. The diversity in immune systems among and within phyla is in and of itself a major macroevolutionary pattern that should become a more central part of how we characterize animal diversity, including in textbooks. We need to know what the full pattern actually is, and the pattern begs an explanation for the processes involved in generating it.

Study of model parasite–host systems shows that the strength of selection imposed on particular immune genes is strong, and can result in some of the fastest evolutionary rates known for metazoans. In other words, parasite–host immune interactions strongly influence anagenesis, the evolutionary changes occurring within parasite or host lineages. If we adopt a broader perspective with a longer time frame, it seems likely such intense interactions will be seen to have an impact on cladogenesis, the origin of lineages, as well. It seems the overall impact of infection and immunity should attract as much attention as predation, competition, or other biotic interactions in shaping the overall diversity of animal life. Needed are more empirical studies over longer time

frames to provide more specific mechanistic insights as to how host immune responses drive diversification of parasites and how this can lead to speciation events, potentially in both parasite and host lineages. Further revelation of specific genes, often not considered part of the immune system *per se*, and how they facilitate defense against particular parasites and might favor evolutionary divergence among parasites are needed. Finally, placing immunology in a macroevolutionary perspective can hopefully provide insights for understanding today's world in which a host of rapid changes greatly increase the odds for extinction of many animals.

ACKNOWLEDGMENTS

This work was supported by NIH grants AI024340, and by 1P20RR18754, the latter from the Institute Development Award (IDeA) Program of the National Center for Research Resources, NIH.

REFERENCES

Acevedo-Whitehouse, K., and Cunningham, A. A. (2006). Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecol. Evol.* 21, 433–438.

Aeby, G. S., and Santavy, D. L. (2006). Factors affecting susceptibility of the coral *Montastraea faveolata* to black-band disease. *Mar. Ecol. Prog. Ser.* 318, 103–110.

Agosta, S. J. (2006). On ecological fitting, plant–insect associations, herbivore host shifts, and host plant selection. *Oikos* 114, 556–565.

Agosta, S. J., Janz, N., and Brooks, D. R. (2010). How specialists can be generalists: resolving the "parasite paradox" and implications for emerging infectious disease. *Zoologia* 27, 151–162.

Anstee, D. J. (2010). The relationship between blood groups and disease. *Blood* 115, 4635–4643.

Atkinson, C. T., and Samuel, M. D. (2010). Avian malaria *Plasmodium relictum* in native Hawaiian forest birds: epizootiology and demographic impacts on 'apapane *Himatione sanguinea*. *J. Avian Biol.* 41, 357–366.

Augustin, R., Sebastian, S., and Bosch, T. C. G. (2010). How Hydra senses and destrys microbes. *Semin. Immunol.* 22, 54–58.

Ausubel, F. M. (2005). Are innate immune signaling pathways in plants and animals conserved? *Nat. Immunol.* 6, 973–979.

Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., Arai, M., Inoue, Y., Shimizu, T., Satoh, N., Rokhsar, D. S., Du Pasquier, L., Kasahara, M., Satake, M., and Nonaka, M. (2003). Genomic analysis of immunity in a urochordate and the emergence of the vertebrate immune system: "waiting for Godot". *Immunogenetics* 55, 570–581.

Barreiro, L. B., and Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat. Rev. Genet.* 11, 17–30.

Bartholomay, L. C., Waterhouse, R. M., Mayhew, G. F., Campbell, C. L., Michel, K., Zou, Z., Ramirez, J. L., Das, S., Alvarez, K., Arensburger, P., Bryant, B., Chapman, S. B., Dong, Y. M., Erickson, S. M., Karunaratne, S. H. P. P., Kokoza, V., Kodira, C. D., Pignatelli, P., Shin, S. W., Vanlandingham, D. L., Atkinson, P. W., Birren, B., Christophides, G. K., Clem, R. J., Hemingway, J., Higgs, S., Megy, K., Ranson, H., Zdobnov, E. M., Raikhel, A. S., Christensen, B. M., Dimopoulos, G., and Muskavitch, M. A. T. (2010). Pathogenomics of *Culex quinquefasciatus* and meta-analysis of infection responses to diverse pathogens. *Science* 330, 88–90.

Bayne, C. J. (2009). Successful parasitism of vector snail Biomphalaria glabrata by the human blood fluke (trematode) Schistosoma mansoni: a 2009 assessment. *Mol. Biochem. Parasitol.* 165, 8–18.

Behnke, J. M., Iraqi, F., Menge, D., Baker, R. L., Gibson, J., and Wakelin, D. (2003). Chasing the genes that control resistance to gastrointestinal nematodes. *J. Helminthol.* 77, 99–109.

Blais, J., Rico, C., van Oosterhout, C., Cable, J., Turner, G. F., and Bernatchez, L. (2007). MHC adaptive divergence between closely related and sympatric African cichlids. *PLoS ONE* 2, e734. doi:10.1371/journal.pone.0000734

Boëte, C., Paul, R. E., and Koella, J. C. (2004). Direct and indirect immunosuppression by a malaria parasite in its mosquito vector. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 1611–1615.

Bowen, G. S. (1976). Experimental infection of North American mammals with epidemic Venezuelan Encephalitis Virus. *Am. J. Trop. Med. Hyg.* 25, 891–899.

Bozeman, F. M., Sonenshine, D. E., Williams, M. S., Chadwick, D. P., Lauer, D. M., and Elisberg, B. L. (1981). Experimental infection of ectoparasitic arthropods with *Rickettsia prowazekii* (GVF-16 Strain) and transmission to flying squirrels. *Am. J. Trop. Med. Hyg.* 30, 253–263.

Brant, S. V., and Loker, E. S. (2005). Can specialized pathogens colonize distantly related hosts? Schistosome evolution as a case study. *PLoS Pathog.* 1, 167–169. doi:10.1371/journal.ppat.0010038

Brites, D., McTaggert, S., Morris, K., Anderson, J., Thomas, K., Colson, I., Fabbro, T., Little, T. J., Ebert, D. and duPasquier, L. (2008). The Dscam homologue of the crustacean Daphnia is diversified by alternative splicing like insects. *Mol. Biol. Evol.* 25, 1429–1439.

Bruno, J. F., Selig, E. R., Casey, K. S., Page, C. A., Willis, B. L., Harvell, C. D., Sweatman, H., and Melendy, A. M. (2007). Thermal stress and coral cover as drivers of coral disease outbreaks. *PLoS Biol.* 5, 1220–1227. doi:10.1371/journal.pbio.0050124

Bryan-Walker, K., Leung, T. L. F., and Poulin, R. (2007). Local adaptation of immunity against a trematode parasite in marine amphipod populations. *Mar. Biol.* 152, 687–695.

Buchmann, K., and Lindenstrøm, T. (2002). Interactions between monogenean parasites and their fish hosts. *Int. J. Parasitol.* 32, 309–319.

Bush, S. E., and Clayton, D. H. (2006). The role of body size in host specificity: reciprocal transfer experiments with feather lice. *Evolution* 60, 2158–2167.

Byers, J. E., Blakeslee, A. M. H., Linder, E., Cooper, A. B., and Maguire, T. J. (2008). Controls of spatial variation in the prevalence of trematode parasites infecting a marine snail. *Ecology* 89, 439–451.

Carius, H. J., Little, T. J., and Ebert, D. (2001). Genetic variation in a host–parasite association: potential for coevolution and frequency-dependent selection. *Evolution* 55, 1136–1145.

Cannon, J. P., Haire, R. N., Schnitker, N., Mueller, M. G., and Litman, G. W. (2004). Individual protochordates have unique immune-type receptor repertoires. *Curr. Biol.* 14, R465–R466.

Cannon, J. P. and Litman, G. W. (2009). Plasticity of the immunoglobulin domain in the evolution of immunity. *Integr. Comp. Biol.* 49, 187–196.

Combes, C. (2001). *Parasitism: The Ecology and Evolution of Intimate Interactions*. Chicago, IL: The University of Chicago Press, 728 p.

Combes, C., and Théron, A. (2000). Metazoan parasites and resource heterogeneity: constraints and benefits. *Int. J. Parasitol.* 30, 299–304.

Cornet, S., Franceschi, N., Bollache, L., Rigaud, T., and Sorci, G. (2009). Variation and covariation in infectivity, virulence and immunodepression in the host–parasite association *Gammarus pulex–Pomphorhynchus laevis*. *Proc. R. Soc. B Biol. Sci.* 276, 4229–4236.

Costa, M. M., Dios, S., Alonso-Gutierrez, J., Romero, A., Novoa, B., and Figueras, A. (2009). Evidence of high individual diversity on myticin C in mussel (*Mytilus galloprovincialis*). *Dev. Comp. Immunol.* 33, 162–170.

Daszak, P. (2010). Bats, in black and white. *Science* 329, 634–635.

Daszak, P., Cunningham, A. A., and Hyatt, A. D. (2000). Wildlife ecology – emerging infectious diseases of wildlife – threats to biodiversity and human health. *Science* 287, 443–449.

Davidson, C. R., Best, N. M., Francis, J. W., Cooper, E. L., and Wood, T. C. (2008). Toll-like receptor genes (TLRs) from *Capitella capitata* and *Helobdella robusta* (Annelida). *Dev. Comp. Immunol.* 32, 608–612.

de Bellocq, J. G., Charbonnel, N., and Morand, S. (2008). Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents. *J. Evol. Biol.* 21, 1144–1150.

Detwiler, J. T., and Criscione, C. D. (2010). An infectious topic in reticulate evolution: introgression and hybridization in animal parasites. *Genes* 1, 102–123.

de Vienne, D. M., Hood, M. E., and Giraud, T. (2009). Phylogenetic determinants of potential host shifts in fungal pathogens. *J. Evol. Biol.* 22, 2532–2541.

Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., and Bernatchez, L. (2007). Clinal variation in MHC diversity with temperature: evidence for the role of host–pathogen interaction on local adaptation in Atlantic salmon. *Evolution* 61, 2154–2164.

Dishaw, L. J., Mueller, M. G., Gwatney, N., Cannon, J. P., Haire, R. N., Litman, R. T., Amemiya, C. T., Ota, T., Rowen, L., Glusman, G., and Litman, G. W. (2008). Genomic complexity of the variable region-containing chitin-binding proteins in amphioxus. *BMC Genet.* 9, 78. doi:10.1186/1471-2156-9-78

Dishaw, L. J., Ota, T., Mueller, M. G., Cannon, J. P., Haire, R. N., Gwatney, N. R., Litman, R. T., and Litman, G. W. (2010). The basis for haplotype complexity in VCBPs, an immune-type receptor in amphioxus. *Immunogenetics* 62, 623–631.

Dong, M., Fu, Y., Yu, C., Su, J., Huang, S., Wu, X., Wei, J., Yuan, S., Shen, Y., and Xu, A. (2005). Identification and characterisation of a homolog of an activation gene for the recombination activating gene 1 (RAG 1) in amphioxus. *Fish Shellfish Immunol.* 19, 165–174.

Dong, Y., and Dimopoulos, G. (2009). Anopheles fibrinogen-related proteins provide expanded pattern recognition capacity against bacteria and malaria parasites. *J. Biol. Chem.* 284, 9835–9844.

Dong, Y., Taylor, H. E., and Dimopoulos, G. (2006). AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. *PLoS Biol.* 4, 1137–1146. doi:10.1371/journal.pbio.0040229

Duffy, M. A., Brassil, C. E., Hall, S. R., Tessier, A. J., Caceres, C. E., and Conner, J. K. (2008). Parasite-mediated disruptive selection in a natural *Daphnia* population. *BMC Evol. Biol.* 8, 1–9. doi:10.1186/1471-2148-8-80

Duke, B. (2004). Failed attempts at experimental transplantation and transmission of nocturnally-periodic simian Loa from monkey to man. *Filaria J.* 3, 5.

Dunn, R. R., Harris, N. C., Colwell, R. K., Koh, L. P., and Sodhi, N. S. (2009). The sixth mass coextinction: are most endangered species parasites and mutualists? *Proc. R. Soc. B Biol. Sci.* 276, 3037–3045.

Eberl, G. (2010). A new vision of immunity: homeostasis of the superorganism. *Mucosal Immunol.* 3, 450–460.

Eizaguirre, C., Lenz, T. L., Sommerfeld, R. D., Harrod, C., Kalbe, M., and Milinski, M. (2011). Parasite diversity, patterns of MHC II variation and olfactory based mate choice in diverging three-spined stickleback ecotypes. *Evol. Ecol.* 25, 605–622.

Eizaguirre, C., Yeates, S. E., Lenz, T. L., Kalbe, M., and Milinski, M. (2009). MHC-based mate choice combines good genes and maintenance of MHC polymorphism. *Mol. Ecol.* 18, 3316–3329.

Erwin, D. H., Laflamme, M., Tweedt, S. M., Sperling, E. A., Pisani, D., and Peterson, K. J. (2011). The Cambrian conundrum: early divergence and later ecological success in the early history of animals. *Science* 334, 1091–1097.

Evans, J. D., Aronstein, K., Chen, Y. P., Hetru, C., Imler, J. L., Jiang, H., Kanost, M., Thompson, G. J., Zou, Z., and Hultmark, D. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol. Biol.* 15, 645–656.

Fincher, C. L., and Thornhill, R. (2008). A parasite-driven wedge: infectious diseases may explain language and other biodiversity. *Oikos* 117, 1289–1297.

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

Flemming, L., Rawlings, D., and Chenia, H. (2007). Phenotypic and molecular characterisation of fish-borne *Flavobacterium johnsoniae*-like isolates from aquaculture systems in South Africa. *Res. Microbiol.* 158, 18–30.

Frank, S. A. (2002). *Immunology and Evolution of Infectious Disease.* Princeton, NJ: Princeton University Press, 348 p.

Fritz, R. S. (1999). Resistance of hybrid plants to herbivores: genes, environment, or both? *Ecology* 80, 382–391.

Fugmann, S. D. (2010). The origins of the Rag genes – from transposition to V(D)J recombination. *Semin. Immunol.* 22, 10–16.

Fytrou, A., Schofield, P. G., Kraaijeveld, A. R., and Hubbard, S. F. (2006). Wolbachia infection suppresses both host defence and parasitoid counter-defence. *Proc. R. Soc. B Biol. Sci.* 273, 791–796.

Garcia-Sastre, A. (2006). Antiviral response in pandemic influenza viruses. *Emerging Infect. Dis.* 12, 44–47.

Garnick, S. W., Elgar, M. A., Beveridge, I., and Coulson, G. (2010). Foraging efficiency and parasite risk in eastern grey kangaroos (*Macropus giganteus*). *Behav. Ecol.* 21, 129–137.

Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., Bowden, D. W., Langefeld, C. D., Oleksyk, T. K., Knob, A. L. U., Bernhardy, A. J., Hicks, P. J., Nelson, G. W., Vanhollebeke, B., Winkler, C. A., Kopp, J. B., Pays, E., and Pollak, M. R. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science* 329, 841–845.

Ghosh, J., Buckley, K. M., Nair, S. V., Raftos, D. A., Miller, C., Majeske, A. J., Hibino, T., Rast, J. P., Roth, M., and Smith, L. C. (2010). Sp185/333: a novel family of genes and proteins involved in the purple sea urchin immune response. *Dev. Comp. Immunol.* 34, 235–245.

Giraud, T., Gladieux, P., and Gavrilets, S. (2010). Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol. Evol.* 25, 387–395.

Gohar, M., Gilois, N., Graveline, R., Garreau, C., Sanchis, V., and Lereclus, D. (2005). A comparative study of *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus anthracis* extracellular proteomes. *Proteomics* 5, 3696–3711.

Goss, E. M., Carbone, I., and Grunwald, N. J. (2009). Ancient isolation and independent evolution of the three clonal lineages of the exotic sudden oak death pathogen *Phytophthora ramorum*. *Mol. Ecol.* 18, 1161–1174.

Graham, A. L., Shuker, D. M., Pollitt, L. C., Auld, S. K. J. R., Wilson, A. J., and Little, T. J. (2011). Fitness consequences of immune responses: strengthening the empirical framework for ecoimmunology. *Funct. Ecol.* 25, 5–17.

Gray, M. J., Miller, D. L., and Hoverman, J. T. (2009). Ecology and pathology of amphibian ranaviruses. *Dis. Aquat. Org.* 87, 243–266.

Gross, R., Vavre, F., Heddi, A., Hurst, G. D., Zchori-Fein, E., and Bourtzis, K. (2009). Immunity and symbiosis. *Mol. Microbiol.* 73, 751–759.

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S., and Poon, L. L. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302, 276–278.

Guzo, D., and Stoltz, D. B. (1985). Obligatory multiparasitism in the tussock moth, *Orgyia leucostigma*. *Parasitology* 90, 1–10.

Haldane, J. B. S. (1949). Disease and evolution. *Ric. Sci.* 19(Suppl.), 68–76.

Hamilton, W. D., and Zuk, M. (1982). Heritable true fitness and bright birds – a role for parasites. *Science* 218, 384–387.

Hammerschmidt, K., and Kurtz, J. (2005). Surface carbohydrate composition of a tapeworm in its consecutive intermediate hosts: individual variation and fitness consequences. *Int. J. Parasitol.* 35, 1499–1507.

Hanington, P. C., Lun, C. M., Adema, C. M., and Loker, E. S. (2010a). Time series analysis of the transcriptional responses of *Biomphalaria glabrata* throughout the course of intramolluscan development of *Schistosoma mansoni* and *Echinostoma paraensei*. *Int. J. Parasitol.* 40, 819–831.

Hanington, P. C., Forys, M. A., Drago, J. W., Zhang, S. M., Adema, C. M., and Loker, E. S. (2010b). Role for a somatically diversified lectin in resistance of an invertebrate to parasite infection. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21087–21092.

Hauton, C., and Smith, V. J. (2007). Adaptive immunity in invertebrates: a straw house without a mechanistic foundation. *Bioessays* 29, 1138–1146.

Hedrick, P. W. (2003). "The major histocompatibility complex (MHC) in declining populations: an example of adaptive variation," in *Reproduction Science and Integrated Conservation*, eds W. V. Holt, A. R. Pickard, J. C. Rodger, and D. E. Wildt (Cambridge, UK: Cambridge University Press), 97–113.

Hedrick, R. P., McDowell, T. S., Marty, G. D., Fosgate, G. T., Mukkatira, K., Myklebust, K., and El-Matbouli, M. (2003). Susceptibility of two strains of rainbow trout (one with suspected resistance to whirling disease) to *Myxobolus cerebralis* infection. *Dis. Aquat. Org.* 55, 37–44.

Hemmrich, G., Miller, D. J. and Bosch, T. C. G. (2007). The evolution of immunity: a low life perspective. *Trends Immunol.* 28, 449–454.

Hibino, T., Loza-Coll, M., Messier, C., Majeske, A. J., Cohen, A. H., Terwilliger, D. P., Buckley, K. M., Brocton, V., Nair, S. V., Berney, K., Fugmann, S. D., Anderson, M. K., Pancer, Z., Cameron, R. A., Smith, L. C., and Rast, J. P. (2006). The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* 300, 349–365.

Hill, A. V. (2001). Immunogenetics and genomics. *Lancet* 357, 2037–2041.

Hoberg, E. P., and Brooks, D. R. (2008). A macroevolutionary mosaic: episodic host-switching, geographical colonization and diversification in complex host–parasite systems. *J. Biogeogr.* 35, 1533–1550.

Holmes, C. B., Losina, E., Walensky, R. P., Yazdanpanah, Y., and Freedberg, K. A. (2003). Review of human

immunodeficiency virus type 1-related opportunistic infections in sub-Saharan Africa. *Clin. Infect. Dis.* 36, 652–662.

Horak, P., and Kolarova, L. (2005). Molluscan and vertebrate immune responses to bird schistosomes. *Parasite Immunol.* 27, 247–255.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and Archaea. *Science* 327, 167–170.

Huang, S., Yuan, S., Guo, L., Yu, Y., Li, J., Wu, T., Liu, T., Yang, M., Wu, K., Liu, H., Ge, J., Huang, H., Dong, M., Yu, C., Chen, S., and Xu, A. (2008). Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. *Genome Res.* 18, 1112–1126.

Hughes, A. L. (1998). Protein phylogenies provide evidence of a radical discontinuity between arthropod and vertebrate immune systems. *Immunogenetics* 47, 283–296.

Hughes, A. L., and Piontkivska, H. (2008). Functional diversification of the toll-like receptor gene family. *Immunogenetics* 60, 249–256.

Irazoqui, J. E., Urbach, J. M., and Ausubel, F. M. (2010). Evolution of host innate defence: insights from *Caenorhabditis elegans* and primitive invertebrates. *Nat. Rev. Immunol.* 10, 47–58.

Jackson, J. A., and Tinsley, R. C. (2003). Parasite infectivity to hybridising host species: a link between hybrid resistance and allopolyploid speciation? *Int. J. Parasitol.* 33, 137–144.

Janz, N., Nylin, S., and Wahlberg, N. (2006). Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evol. Biol.* 6, 1–10. doi:10.1186/1471-2148-6-4

Johnson, K. P., Malenke, J. R., and Clayton, D. H. (2009). Competition promotes the evolution of host generalists in obligate parasites. *Proc. R. Soc. B Biol. Sci.* 276, 3921–3926.

Kalbe, M., and Kurtz, J. (2006). Local differences in immunocompetence reflect resistance of sticklebacks against the eye fluke *Diplostomum pseudospathaceum*. *Parasitology* 132, 105–116.

Kanost, M. R., Jiang, H. B., and Yu, X. Q. (2004). Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol. Rev.* 198, 97–105.

Karraker, N. E., and Ruthig, G. R. (2009). Effect of road deicing salt on the susceptibility of amphibian embryos to infection by water molds. *Environ. Res.* 109, 40–45.

Keesing, F., Brunner, J., Duerr, S., Killilea, M., LoGiudice, K., Schmidt, K., Vuong, H., and Ostfeld, R. S. (2009). Hosts as ecological traps for the vector of Lyme disease. *Proc. R. Soc. B Biol. Sci.* 276, 3911–3919.

Koh, L. P., Dunn, R. R., Sodhi, N. S., Colwell, R. K., Proctor, H. C., and Smith, V. S. (2004). Species coextinctions and the biodiversity crisis. *Science* 305, 1632–1634.

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R., and Bunning, M. (2003). Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerging Infect. Dis.* 9, 311–322.

Kovacs, J., and Masur, H. (2008). HIV related opportunistic infections: still relevant after 25 years of AIDS progress. *Enferm. Infecc. Microbiol. Clin.* 26, 323–324.

Kraaijeveld, A. R., and Godfray, H. C. J. (1999). Geographic patterns in the evolution of resistance and virulence in *Drosophila* and its parasitoids. *Am. Nat.* 153, S61–S74.

Krasnov, B. R., Mouillot, D., Khokhlova, I. S., Shenbrot, G. I., and Poulin, R. (2005). Covariance in species diversity and facilitation among noninteractive parasite taxa: all against the host. *Parasitology* 131, 557–568.

Kuris, A. M., Goddard, J. H. R., Torchin, M. E., Murphy, N., Gurney, R., and Lafferty, K. D. (2007). An experimental evaluation of host specificity: the role of encounter and compatibility filters for a rhizocephalan parasite of crabs. *Int. J. Parasitol.* 37, 539–545.

Labropoulou, V., Douris, V., Stefanou, D., Magrioti, C., Swevers, L., and Iatrou, K. (2008). Endoparasitoid wasp bracovirus-mediated inhibition of hemolin function and lepidopteran host immunosuppression. *Cell. Microbiol.* 10, 2118–2128.

Lafferty, K. D., and Kuris, A. M. (2009a). Parasitic castration: the evolution and ecology of body snatchers. *Trends Parasitol.* 25, 564–572.

Lafferty, K. D., and Kuris, A. M. (2009b). Parasites reduce food web robustness because they are sensitive to secondary extinction as illustrated by an invasive estuarine snail. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1659–1663.

Landry, C., Garant, D., Duchesne, P., and Bernatchez, L. (2001). 'Good genes as heterozygosity': the major histocompatibility complex and mate choice in Atlantic salmon (*Salmo salar*). *Proc. R. Soc. Lond. B Biol. Sci.* 268, 1279–1285.

Lazzaro, B. P., and Little, T. J. (2009). Immunity in a variable world. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 15–26.

Lee, K. A. (2006). Linking immune defenses and life history at the levels of the individual and the species. *Integr. Comp. Biol.* 46, 1000–1015.

Lee, K. A., and Klasing, K. C. (2004). A role for immunology in invasion biology. *Trends Ecol. Evol.* 19, 523–529.

Lemaitre, B., and Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annu. Rev. Immunol.* 25, 697–743.

Levinton, J. S. (2001). *Genetics, Paleontology and Macroevolution.* Cambridge, MA: Cambridge University Press, 634 p.

Lie, K. J. (1982). Survival of *Schistosoma mansoni* and other trematode larvae in the snail *Biomphalaria glabrata* – a discussion of the interference theory. *Trop. Geogr. Med.* 34, 111–122.

Light, J. E., and Hafner, M. S. (2008). Codivergence in heteromyid rodents (Rodentia: Heteromyidae) and their sucking lice of the genus *Fahrenholzia* (Phthiraptera: Anoplura). *Syst. Biol.* 57, 449–465.

Lindstrom, K. M., Foufopoulos, J., Parn, H., and Wikelski, M. (2004). Immunological investments reflect parasite abundance in island populations of Darwin's finches. *Proc. R. Soc. Lond. B Biol. Sci.* 271, 1513–1519.

Litman, G. W., Rast, J. P., and Fugmann, S. D. (2010). The origins of vertebrate adaptive immunity. *Nat. Rev. Immunol.* 10, 543–553.

Loker, E. S. (1994). On being a parasite in an invertebrate host – a short survival course. *J. Parasitol.* 80, 728–747.

Loker, E. S. (2010). "Gastropod immunobiology," in *Invertebrate Immunity*, ed. K. Söderhall, (Austin/New York: Landes Bioscience/Springer), 17–43.

Loker, E. S., Moyo, H. G., and Gardner, S. L. (1981). Trematode-gastropod associations in 9 non-lacustrine habitats in the Mwanza region of Tanzania. *Parasitology* 83, 381–399.

MacColl, A. D. C. (2009). Parasites may contribute to 'magic trait' evolution in the adaptive radiation of three-spined sticklebacks, *Gasterosteus aculeatus* (Gasterosteiformes: Gasterosteidae). *Biol. J. Linn. Soc.* 96, 425–433.

Mangeat, B., Turelli, P., Liao, S. Y., and Trono, D. (2004). A single amino acid determinant governs the species-specific sensitivity of APOBEC3G to vif action. *J. Biol. Chem.* 279, 14481–14483.

Maan, E., and Seehausen, O. (2010). Mechanisms of species divergence through visual adaptation and sexual selection: perspectives from a cichlid model system. *Curr. Zool.* 56, 285–299.

Martinsen, E. S., Perkins, S. L., and Schall, J. J. (2008). A three-genome phylogeny of malaria parasites (*Plasmodium* and closely related genera): evolution of life-history traits and host switches. *Mol. Phylogenet. Evol.* 47, 261–273.

Matsunaga, T., and Rahman, A. (1998). What brought the adaptive immune system to vertebrates? The jaw hypothesis and the seahorse. *Immunol. Rev.* 166, 177–186.

Matthews, B., Harmon, L. J., M'Gonigle, L., Marchinko, K. B., and Schaschl, H. (2010). Sympatric and allopatric divergence of MHC Genes in Threespine Stickleback. *PLoS ONE* 5, e10948. doi:10.1371/journal.pone.0010948

Mayr, E. (1963). *Animal Species and Evolution.* Cambridge, MA: Belknap Press of Harvard University Press.

McClintock, M. K., Jacob, S., Zelano, B., and Ober, C. (2002). Preferences for human odors are associated with paternally inherited HLA alleles. *Horm. Behav.* 41, 451.

McKeever, D. J. (2009). Bovine immunity – a driver for diversity in *Theileria* parasites? *Trends Parasitol.* 25, 269–276.

McKenzie, V. J., and Townsend, A. R. (2007). Parasitic and infectious disease responses to changing global nutrient cycles. *EcoHealth* 4, 384–396.

McVicar, A. H., and Fletcher, T. C. (1970). Serum factors in *Raja radiata* toxic to *Acanthobothrium quadripartitum* (Cesoda–Tetraphyllidea), a parasite specific to *R. naevus*. *Parasitology* 61, 55–63.

Messier-Solek, C., Buckley, K. M., and Rast, J. P. (2010). Highly diversified innate receptor systems and new forms of animal immunity. *Semin. Immunol.* 22, 39–47.

Milinski, M. (2006). The major histocompatibility complex, sexual selection, and mate choice. *Annu. Rev. Ecol. Evol. Syst.* 37, 159–186.

Milinski, M., Griffiths, S., Wegner, K. M., Reusch, T. B. H., Haas-Assenbaum, A., and Boehm, T. (2005). Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4414–4418.

Miller, W. J., Ehrman, L., and Schneider, D. (2010). Infectious speciation

revisited: impact of symbiont depletion on female fitness and mating behavior of *Drosophila paulistorum*. *PLoS Pathog.* 6, e1001214. doi:10.1371/journal.ppat.1001214.

Miller, D. J., Hemmrich, G., Ball, E. E., Hayward, D. C., Khalturin, K., Funayama, N., Agata, K., and Bosch, T. C. G. (2007). The innate immune repertoire in Cnidaria – ancestral complexity and stochastic gene loss. *Genome Biol.* 8, R59.1–R59.13.

Mooring, M. S., Fitzpatrick, T. A., Fraser, I. C., Benjamin, J. E., Reisig, D. D., and Nishihira, T. T. (2003). Insect-defense behavior by desert bighorn sheep. *Southwest. Nat.* 48, 635–643.

Morales-Hojas, R., Vieira, C. P., Reis, M., and Vieira, J. (2009). Comparative analysis of five immunity-related genes reveals different levels of adaptive evolution in the virilis and melanogaster groups of *Drosophila*. *Heredity* 102, 573–578.

Morgan, J. A., DeJong, R. J., Kazibwe, F., Mkoji, G. M., and Loker, E. S. (2003). A newly-identified lineage of Schistosoma. *Int. J. Parasitol.* 33, 977–985.

Neill, D. R., Wong, S. H., Bellosi, A., Flynn, R. J., Daly, M., Langford, T. K. A., Bucks, C., Kane, C. M., Fallon, P. G., Pannell, R., Jolin, H. E., and McKenzie, A. N. J. (2010). that mediates type-2 immunity. *Nature* 464, 1367–1370.

Nelson, D. E., Virok, D. P., Wood, H., Roshick, C., Johnson, R. M., Whitmire, W. M., Crane, D. D., Steele-Mortimer, O., Kari, L., McClarty, G., and Caldwell, H. D. (2005). Chlamydial IFN-gamma immune evasion is linked to host infection tropism. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10658–10663.

Nicotra, M. L., Powell, A. E., Rosengarten, R. D., Moreno, M., Grimwood, J., Lakkis, F. G., Dellaporta, S. L., and Buss, L. W. (2009). A hypervariable invertebrate allodeterminant. *Curr. Biol.* 19, 583–589.

Nyholm, S. V., and Nishiguchi, M. K. (2008). The evolutionary ecology of a sepiolid squid-vibrio association: from cell to environment. *Vie Milieu* 58, 175–184.

Nyman, T. (2010). To speciate, or not to speciate? Resource heterogeneity, the subjectivity of similarity, and the macroevolutionary consequences of niche-width shifts in plant-feeding insects. *Biol. Rev.* 85, 393–411.

Olson, P. D., Caira, J. N., Jensen, K., Overstreet, R. M., Palm, H. W., and Beveridge, I. (2010). Evolution of the trypanorhynch tapeworms: parasite phylogeny supports independent lineages of sharks and rays. *Int. J. Parasitol.* 40, 223–242.

Ottenhoff, T. H. M., Verreck, F. A. W., Hoeve, M. A., and van de Vosse, E. (2005). Control of human host immunity to mycobacteria. *Tuberculosis* 85, 53–64.

Pais, R., Lohs, C., Wu, Y., Wang, J., and Aksoy S. (2008). The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Appl. Environ. Microbiol.* 74, 5965–5974.

Pancer, Z., Amemiya, C. T., Ehrhardt, G. R. A., Ceitlin, J., Gartland, G. L., and Cooper, M. D. (2004). Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 430, 174–180.

Parra, Z. E., Baker, M. L., Schwarz, R. S., Deakin, J. E., Lindblad-Toh, K., and Miller, R. D. (2007). A unique T cell receptor discovered in marsupials. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9776–9781.

Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E. C., Burke, D. S., Calisher, C. H., Laughlin, C. A., Saif, L. J., and Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol. Mol. Biol. Rev.* 72, 457–470.

Pennisi, E. (2009). The Biology of Genomes, 5–9 May 2009, Cold Spring Harbor, New York. The bug and the bacterium: interdependent genomes. *Science* 324, 1253.

Philips, H. L., and Clarkson, M. J. (1998). Experimental infection of pregnant ewes with *Chlamydia pecorum*. *Infect. Immun.* 66, 2818–2821.

Pimm, S. L., and Raven, P. (2000). Biodiversity extinction by numbers. *Nature* 403, 843–845.

Pfeffer, M., and Dobler, G. (2010). Emergence of zoonotic arboviruses by animal trade and migration. *Parasit. Vectors* 3, 1–15.

Poulin, R., and Keeney, D. B. (2008). Host specificity under molecular and experimental scrutiny. *Trends Parasitol.* 24, 24–28.

Poulin, R., and Thomas, F. (1999). Phenotypic variability induced by parasites: extent and evolutionary implications. *Parasitol. Today* 15, 28–32.

Price, P. W. (1980). *Evolutionary Biology of Parasites*. Princeton, NJ: Princeton University Press, 237 p.

Prugnolle, F., Manica, A., Charpentier, M., Guegan, J. F., Guernier, V., and Balloux, F. (2005). Pathogen-driven selection and worldwide HLA class I diversity. *Curr. Biol.* 15, 1022–1027.

Qureshi, S. T., and Medzhitov, R. (2003). Toll-like receptors and their role in experimental models of microbial infection. *Genes Immun.* 4, 87–94.

Radwan, J., Biedrzycka, A., and Babik, W. (2010). Does reduced MHC diversity decrease viability of vertebrate populations? *Biol. Conserv.* 143, 537–544.

Rast, J., and Litman, G. (2010). Evolution of innate and adaptive immune recognition structures. *Semin. Immunol.* 22, 1–2.

Refrégier, G., Le Gac, M., Jabbour, F., Widmer, A., Shykoff, J. A., Yockteng, R., Hood, M. E., and Giraud, T. (2008). Cophylogeny of the anther smut fungi and their caryophyllaceous hosts: prevalence of host shifts and importance of delimiting parasite species for inferring cospeciation. *BMC Evol. Biol.* 8, 100. doi:10.1186/1471-2148-8-100

Rest, J. S., and Mindell, D. P. (2003). SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infect. Genet. Evol.* 3, 219–225.

Rigaud, T., and Moret, Y. (2003). Differential phenoloxidase activity between native and invasive gammarids infected by local acanthocephalans: differential immunosuppression? *Parasitology* 127, 571–577.

Rivkin, H., Kroemer, J. A., Bronshtein, A., Belausov, E., Webb, B. A., and Chejanovsky, N. (2006). Response of immunocompetent and immunosuppressed *Spodoptera littoralis* larvae to baculovirus infection. *J. Gen. Virol.* 87, 2217–2225.

Roeder, T., Stanisak, M., Gelhaus, C., Bruchhaus, I., Grotzinger, J., and Leippe, M. (2010). Caenopores are antimicrobial peptides in the nematode *Caenorhabditis elegans* instrumental in nutrition and immunity. *Dev. Comp. Immunol.* 34, 203–209.

Rogozin, I. B., Iyer, L. M., Liang, L. Z., Glazko, G. V., Liston, V. G., Pavlov, Y. I., Aravind, L., and Pancer, Z. (2007). Evolution and diversification of lamprey antigen receptors: evidence for involvement of AID-APOBEC family cytosine deaminase. *Nature Immunol.* 8, 647–656.

Rohr, J. R., Schotthoefer, A. M., Raffel, T. R., Carrick, H. J., Halstead, N., Hoverman, J. T., Johnson, C. M., Johnson, L. B., Lieske, C., Piwoni, M. D., Schoff, P. K., and Beasley, V. R. (2008). Agrochemicals increase trematode infections in a declining amphibian species. *Nature* 455, 1235–1239.

Rolff, J., and Siva-Jothy, M. T. (2002). Copulation corrupts immunity: a mechanism for a cost of mating in insects. *Proc. Natl. Acad. Sci. U.S.A.* 99, 9916–9918.

Rooijakkers, S. H., Ruyken, M., Roos, A., Daha, M. R., Presanis, J. S., Sim, R. B., van Wamel, W. J., van Kessel, K. P., and van Strijp, J. A. (2005). Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. *Nat. Immunol.* 6, 920–927.

Rosengard, A. M., Liu, Y., Nie, Z., and Jimenez, R. (2002). Variola virus immune evasion design: expression of a highly efficient inhibitor of human complement. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8808–8813.

Sackton, T. B., Lazzaro, B. P., Schlenke, T. A., Evans, J. D., Hultmark, D., and Clark, A. G. (2007). Dynamic evolution of the innate immune system in *Drosophila*. *Nat. Genet.* 39, 1461–1468.

Saenz, S. A., Siracusa, M. C., Perrigoue, J. G., Spencer, S. P., Urban, J. F., Tocker, J. E., Budelsky, A. L., Kleinschek, M. A., Kastelein, R. A., Kambayashi, T., Bhandoola, A., and Artis, D. (2010). IL25 elicits a multipotent progenitor cell population that promotes T(H)2 cytokine responses. *Nature* 464, 1362–1366.

Sahu, S., Das, B. K., Mishra, B. K., Pradhan, J., Samal, S. K., and Sarangi, N. (2008). Effect of dietary *Curcuma longa* on enzymatic and immunological profiles of rohu, *Labeo rohita* (Ham.), infected with *Aeromonas hydrophila*. *Aquac. Res.* 39, 1720–1730.

Salzet, M., Tasiemski, A., and Cooper, E. (2006). Innate immunity in lophotrochozoa: the annelids. *Curr. Pharm. Des.* 12, 3043–3050.

Sapp, K. K., and Loker, E. S. (2000). Mechanisms underlying digenean-snail specificity: role of miracidial attachment and host plasma factors. *J. Parasitol.* 86, 1012–1019.

Scharsack, J. P., Koch, K., and Hammerschmidt, K. (2007). Who is in control of the stickleback immune system: interactions between *Schistocephalus solidus* and its specific vertebrate host. *Proc. R. Soc. B Biol. Sci.* 274, 3151–3158.

Schmid-Hempel, P. (2008). Parasite immune evasion: a momentous molecular war. *Trends Ecol. Evol.* 23, 318–326.

Schmidt, O., Theopold, U., and Strand, M. (2001). Innate immunity and its evasion and suppression by hymenopteran endoparasitoids. *Bioessays* 23, 344–351.

Schmucker, D., and Chen, B. (2009). Dscam and DSCAM: complex genes in simple animals, complex animals

yet simple genes. *Genes Dev.* 23, 147–156.

Schroder, H. C., Natalio, F., Wiens, M., Tahir, M. N., Shukoor, M. I., Tremel, W., Belikov, S. I., Krasko, A., and Muller, W. E. (2008). The 2′-5′-oligoadenylate synthetase in the lowest metazoa: isolation, cloning, expression and functional activity in the sponge *Lubomirskia baicalensis*. *Mol. Immunol.* 45, 945–953.

Schulenburg, H., and Ewbank, J. J. (2004). Diversity and specificity in the interaction between *Caenorhabditis elegans* and the pathogen *Serratia marcescens*. *BMC Evol. Biol.* 4, 49. doi:10.1186/1471-2148-4-49

Schulte, R. D., Makus, C., Hasert, B., Michiels, N. K., and Schulenburg, H. (2010). Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7359–7364.

Seet, B. T., Johnston, J. B., Brunetti, C. R., Barrett, J. W., Everett, H., Cameron, C., Sypula, J., and Nazarian, S. H., Lucas, A., and McFadden, G. (2003). Poxviruses and immune evasion. *Annu. Rev. Immunol.* 21, 377–423.

Seppala, O., Karvonen, A., and Valtonen, E. T. (2007). Phenotypic variation in infectivity of *Diplostomum spathaceum* cercariae within a population. *J. Parasitol.* 93, 1244–1246.

Shaw, S. R. (1988). Euphorine phylogeny – the evolution of diversity in host utilization by parasitoid wasps (Hymenoptera, Braconidae). *Ecol. Entomol.* 13, 323–335.

Shelby, K. S., Cui, L., and Webb, B. A. (1998). Polydnavirus-mediated inhibition of lysozyme gene expression and the antibacterial response. *Insect Mol. Biol.* 7, 265–272.

Smith, C. R., Smith, C. D., Robertson, H. M., Helmkampf, M., Zimin, A., Yandell, M., Holt, C., Hu, H. E., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C. R., Elhak, E., Elsik, C. G., Fave, M. N., Fernandes, V., Gibson, J. D., Garaur, D., Gronenberg, W., Grubbs, K. J., Hagen, D. E., Viniegra, A. S. I., Johnson, B. R., Johnson, R. M., Khila, A., Kim, J. W., Mathis, K. A., Munoz-Torres, M. C., Murphy, M. C., Mustard, J. A., Nakamura, R., Niehuis, O., Nigam, S., Overson, R. P., Placek, J. E., Rajakumar, R., Reese, J. T., Suen, G., Tao, S., Torres, C. W., Tsutsui, N. D., Viljakainen, L., Wolschin, F., and Gadau, J. (2011). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5667–5672.

Sommer, S. (2005). Major histocompatibility complex and mate choice in a monogamous rodent. *Behav. Ecol. Sociobiol. (Print)* 58, 181–189.

Southgate, V. R., Brown, D. S., Warlow, A., Knowles, R. J., and Jones, A. (1989). The influence of *Calicophoron microbothrium* on the susceptibility of *Bulinus tropicus* to *Schistosoma bovis*. *Parasitol. Res.* 75, 381–391.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. A., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N. H., Stanke, M., Adamska, M., Darling, A., Degnan, S. M., Oakley, T. H., Plachetzki, D. C., Zhai, Y. F., Adamski, M., Calcino, A., Cummins, S. F., Goodstein, D. M., Harris, C., Jackson, D. J., Leys, S. P., Shu, S. Q., Woodcroft, B. J., Vervoort, M., Kosik, K. S., Manning, G., Degnan, B. M., and Rokhsar, D. S. (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466, 720–726.

Stanley, S. M. (1998). *Macroevolution: Pattern and Process*. Baltimore, MD: The Johns Hopkins University Press, 370 p.

Streicker, D. G., Turmelle, A. S., Vonhof, M. J., Kuzmin, I. V., McCracken, G. F., and Rupprecht, C. E. (2010). Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science* 329, 676–679.

Stukenbrock, E. H., and McDonald, B. A. (2008). The origins of plant pathogens in agro-ecosystems. *Annu. Rev. Phytopathol.* 46, 75–100.

Sullivan, J. T., and Richards, C. S. (1981). Schistosoma mansoni, NIH-SM-PR-2 strain, in susceptible and non-susceptible stocks of *Biomphalaria glabrata*: comparative histology. *J. Parasitol.* 67, 702–708.

Summers, K., McKeon, S., Sellars, J., Keusenkothen, M., Morris, J., Gloeckner, D., Pressley, C., Price, B., and Snow, H. (2003). Parasitic exploitation as an engine of diversity. *Biol. Rev. Camb. Philos. Soc.* 78, 639–675.

Sung, J. M., Lloyd, D. H., and Lindsay, J. A. (2008). *Staphylococcus aureus* host specificity: comparative genomics of human versus animal isolates by multi-strain microarray. *Microbiology* 154, 1949–1959.

Taraschewski, H. (2006). Hosts and parasites as aliens. *J. Helminthol.* 80, 99–128.

Tenor, J. L., and Aballay, A. (2008). A conserved Toll-like receptor for *Caenorhabditis elegans* innate immunity. *EMBO Rep.* 9, 103–109.

Thomas, F., Guldner, E., and Renaud, F. (2000). Differential parasite (Trematoda) encapsulation in *Gammarus aequicauda* (Amphipoda). *J. Parasitol.* 86, 650–654.

Thomas, J. A., Telfer, M. G., Roy, D. B., Preston, C. D., Greenwood, J. J. D., Asher, J., Fox, R., Clarke, R. T., and Lawton, J. H. (2004). Comparative losses of British butterflies, birds and plants and the global extinction crisis. *Science* 303, 1879–1881.

Thompson, J. N. (2005). *The Geographic Mosaic of Coevolution*. Chicago, IL: The University of Chicago Press.

Thompson, J. N. (2009). The coevolving web of life. *Am. Nat.* 173, 125–140.

Tompkins, D. M., and Clayton, D. H. (1999). Host resources govern the specificity of swiftlet lice: size matters. *J. Anim. Ecol.* 68, 489–500.

Torchin, M. E., Lafferty, K. D., and Kuris, A. M. (2002). Parasites and marine invasions. *Parasitology* 124, S137–S151.

Trachtenberg, E., Korber, B., Sollars, C., Kepler, T. B., Hraber, P. T., Hayes, E., Funkhouser, R., Fugate, M., Theiler, J., Hsu, Y. S., Kunstman, K., Wu, S., Phair, J., Erlich, H., and Wolinsky, S. (2003). Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* 9, 928–935.

Trowsdale, J., and Parham, P. (2004). Defense strategies and immunity-related genes. *Eur. J. Immunol.* 34, 7–17.

Tunaz, H., and Stanley, D. (2009). An immunological axis of biocontrol: infections in field-trapped insects. *Naturwissenschaften* 96, 1115–1119.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810.

VandeWoude, S., Troyer, J., and Poss, M. (2010). Restrictions to cross-species transmission of lentiviral infection gleaned from studies of FIV. *Vet. Immunol. Immunopathol.* 134, 25–32.

van Doorn, G. S., Edelaar, P., and Weissing, F. J. (2009). On the origin of species by natural and sexual selection. *Science* 326, 1704–1707.

Vidal-Martinez, V. M., Osoriosarabia, D., and Overstreet, R. M. (1994). Experimental infection of *Contracaecum multipapillatum* (Nematoda, Anisakinae) from Mexico in the domestic cat. *J. Parasitol.* 80, 576–579.

Viljakainen, L., Evans, J. D., Hasselmann, M., Rueppell, O., Tingek, S., and Pamilo, P. (2009). Rapid evolution of immune proteins in social insects. *Mol. Biol. Evol.* 26, 1791–1801.

Vorburger, C., Sandrock, C., Gouskov, A., Castaneda, L. E., and Ferrari, J. (2009). Genotypic variation and the role of defensive endosymbionts in an all parthenogenetic host–parasitoid interaction. *Evolution* 63, 1439–1450.

Walker, J. C. (1979). *Austrobilharzia terrigalensis* – schistosome dominant in interspecific interactions in the molluscan host. *Int. J. Parasitol.* 9, 137–140.

Wang, F., Ma, Y. Y., Barrett, J. W., Gao, X. J., Loh, J., Barton, E., Virgin, H. W., and McFadden, G. (2004). Disruption of Erk-dependent type I interferon induction breaks the myxoma virus species barrier. *Nat. Immunol.* 5, 1266–1274.

Webb, B., Fisher, T., and Nusawardani, T. (2009). The natural genetic engineering of polydnaviruses. *Ann. N. Y. Acad. Sci.* 1178, 146–156.

Wedekind, C., Gessner, M. O., Vazquez, F., Maerki, M., and Steiner, D. (2010). Elevated resource availability sufficient to turn opportunistic into virulent fish pathogens. *Ecology* 91, 1251–1256.

Wegner, K. M. (2008). Historical and contemporary selection of teleost MHC genes: did we leave the past behind? *J. Fish Biol.* 73, 2110–2132.

Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H., and Milinski, M. (2003). Parasite selection for immunogenetic optimality. *Science* 301, 1343.

Weissenbock, H., Hubalek, Z., Bakonyi, T., and Nowotny, N. (2010). Zoonotic mosquito-borne flaviviruses: worldwide presence of agents with proven pathogenicity and potential candidates of future emerging diseases. *Vet. Microbiol.* 140, 271–280.

Welchman, D. P, Aksoy, S., Jiggins, F., and Lemaitre, B. (2009). Insect immunity: from pattern recognition to symbiont-mediated host defense. *Cell Host Microbe* 6, 107–114.

Wheatley, B. P. (1980). Malaria as a possible selective factor in the speciation of macaques. *J. Mammal.* 61, 307–311.

Wheeler, R. J. (2010). The trypanolytic factor-mechanism, impacts and applications. *Trends Parasitol.* 26, 457–464.

Wiegel, F. W., and Perelson, A. S. (2004). Some scaling principles for the immune system. *Immunol. Cell Biol.* 82, 127–131.

Wiens, M., Korzhev, M., Perovic-Ottstadt, S., Luthringer, B., Brandt, D.,

Klein, S., and Muller, W. E. G. (2007). Toll-like receptors are part of the innate immune defense system of sponges (Demospongiae: Porifera). *Mol. Biol. Evol.* 24, 792–804.

Winkler, I. S., Mitter, C., and Scheffer, S. J. (2009). Repeated climate-linked host shifts have promoted diversification in a temperate clade of leafmining flies. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18103–18108.

Wolinska, J., Lively, C. M., and Spaak, P. (2008). Parasites in hybridizing communities: the Red Queen again? *Trends Parasitol.* 24, 121–126.

Wolinska, J., and Spaak, P. (2009). The cost of being common: evidence from natural *Daphnia* populations. *Evolution* 63, 1893–1901.

Wood, M. J., Cosgrove, C. L., Wilkin, T. A., Knowles, S. C., Day, K. P., and Sheldon, B. C. (2007). Within-population variation in prevalence and lineage distribution of avian malaria in blue tits, *Cyanistes caeruleus*. *Mol. Ecol.* 16, 3263–3273.

Wood-Charlson, E. M., and Weis, V. M. (2009). The diversity of C-type lectins in the genome of a basal metazoan, *Nematostella vectensis*. *Dev. Comp. Immunol.* 33, 881–889.

Woolhouse, M. E. J., Haydon, D. T., and Antia, R. (2005). Emerging pathogens: the epidemiology and evolution of species jumps. *Trends Ecol. Evol.* 20, 238–244.

Wyatt, K. B., Campos, P. F., Gilbert, M. T., Kolokotronis, S. O., Hynes, W. H., DeSalle, R., Ball, S. J., Daszak, P., MacPhee, R. D., and Greenwood, A. D. (2008). Historical mammal extinction on Christmas Island (Indian Ocean) correlates with introduced infectious disease. *PLoS ONE* 3, e3602. doi:10.1371/journal.pone.0003602

Yang, X., and Cox-Foster, D. L. (2005). Impact of an ectoparasite on the immunity and pathology of an invertebrate: evidence for host immunosuppression and viral amplification. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7470–7475.

Yoshino, T. P., Dinguirard, N., Kunert, J., and Hokk, C. H. (2008). Molecular and functional characterization of a tandem-repeat galectin from the freshwater snail Biomphalaria glabrata, intermediate host of the human blood fluke Schistosoma mansoni. *Gene* 411, 46–58.

Zhang, Q., Zmasek, C. M., and Godzik, A. (2010). Domain architecture evolution of pattern-recognition receptors. *Immunogenetics* 62, 263–272.

Zhang, S. M., Adema, C. M., Kepler, T. B., and Loker, E. S. (2004). Diversification of Ig superfamily genes in an invertebrate. *Science* 305, 251–254.

Zietara, M. S., and Lumme, J. (2002). Speciation by host switch and adaptive radiation in a fish parasite genus *Gyrodactylus* (Monogenea, Gyrodactylidae). *Evolution* 56, 2445–2458.

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.

Received: 30 November 2011; accepted: 08 February 2012; published online: 12 March 2012.

Citation: Loker ES (2012) Macroevolutionary immunology: a role for immunity in the diversification of animal life. Front. Immun. ***3****:25. doi: 10.3389/fimmu.2012.00025*

This article was submitted to Frontiers in Molecular Innate Immunity, a specialty of Frontiers in Immunology.