

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 1, 2009):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/314/5801/952>

Supporting Online Material can be found at:

<http://www.sciencemag.org/cgi/content/full/314/5801/952/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/314/5801/952#related-content>

This article **cites 40 articles**, 14 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/314/5801/952#otherarticles>

This article has been **cited by** 66 article(s) on the ISI Web of Science.

This article has been **cited by** 14 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/314/5801/952#otherarticles>

This article appears in the following **subject collections**:

Genetics

<http://www.sciencemag.org/cgi/collection/genetics>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

20892, USA. ⁵Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy. ⁶Department of Biology, Boston College, Chestnut Hill, MA 02467, USA. ⁷Department of Biology, Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada, V8W 3N5. ⁸Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672, USA. ⁹Human Genetics Section, Laboratory of Genomic Diversity, National Cancer Institute–Frederick, Frederick, MD 21702, USA. ¹⁰School of Biological and Chemical Sciences, Queen Mary, University of London, London E1 4NS, UK. ¹¹Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA. ¹²Department Molecular, Cellular and Developmental Biology and the Marine Science Institute, University of California, Santa Barbara, Santa Barbara, CA 93106–9610, USA. ¹³Hopkins Marine Station, Stanford University, Pacific Grove, CA 93950, USA. ¹⁴Howard Hughes Medical Institute, Center for Cancer Research, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. ¹⁵Departments of Biochemistry and Molecular Biology, University of Texas, M. D. Anderson Cancer Center, Houston, TX, 77030, USA. ¹⁶Molecular Biology and Biotechnology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. ¹⁷Department of Biology, Duke University, Durham, NC 27708, USA. ¹⁸Department of Biology, Wheaton College, Norton, MA 02766, USA. ¹⁹Stowers Institute for Medical Research, Kansas City, MO 64110, USA. ²⁰Department of Microbiology, Kansas University Medical Center, Kansas City, KS 66160, USA. ²¹Sunnybrook Research Institute and Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada M4N 3M5. ²²Department of Immunology, University of Toronto, Toronto, Ontario, Canada, M4N 3M5. ²³Department of Biological Sciences, George Washington University, Washington, DC 20052, USA. ²⁴Royal Swedish Academy of Sciences, Kristineberg Marine Research Station, Fiskebackskil, 450 34, Sweden. ²⁵Marine Biology, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093–0202, USA. ²⁶Department of Molecular and Cellular Biology and Biochemistry, Brown University Providence, RI 02912, USA. ²⁷Department of Biology and Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA. ²⁸Department of Animal Science, Texas A&M University, College Station, TX 77843, USA. ²⁹National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD 20894, USA. ³⁰Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA. ³¹National Center for Biotechnology Information, NIH, Bethesda, MD 20892, USA. ³²Penn Genomics Institute, University of Pennsylvania, Philadelphia, PA 19104, USA. ³³Evolution and Development Group, Max-Planck Institut für Molekulare Genetik, 14195 Berlin, Germany. ³⁴Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. ³⁵Center for Cancer Research, MIT, Cambridge, MA 02139, USA. ³⁶Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720–3200, USA. ³⁷Department of Biology, University of South Florida, Tampa, FL 33618, USA. ³⁸Université Pierre et Marie Curie (Paris 6), UMR 7150, Equipe Cycle Cellulaire et Développement, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. ³⁹CNRS, UMR 7150, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. ⁴⁰CNRS, UMR7628, Banyuls-sur-Mer, F-66650, France. ⁴¹Université Pierre et Marie Curie (Paris 6), UMR7628, Banyuls-sur-Mer, F-66650, France. ⁴²Center for Bioinformatics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. ⁴³Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. ⁴⁴Tethys Research, LLC, 2115 Union Street, Bangor, Maine 04401, USA. ⁴⁵Department of Molecular, Cellular, and Developmental Biology, University of California, Berkeley, Berkeley, CA 94720, USA. ⁴⁶Center for Computational Molecular Biology, and Computer Science Department, Brown University, Providence, RI 02912, USA. ⁴⁷Genome Research Facility, National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA 94035,

USA. ⁴⁸Systemix Institute, Cupertino, CA 95014, USA. ⁴⁹Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6. ⁵⁰Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6. ⁵¹Department of Biology, Center for Cancer Research, MIT, Cambridge, MA 02139, USA. ⁵²Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089–0740, USA. ⁵³Department of Biology, University of Central Florida, Orlando, FL 32816–2368, USA. ⁵⁴Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA. ⁵⁵Center for Computational Regulatory Genomics, Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA. ⁵⁶Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA. ⁵⁷Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, BC, Canada, V5Z 4E6. ⁵⁸Department of Biology and the Institute of Systems Research, University of Maryland, College Park, MD 20742, USA. ⁵⁹Laboratory of Cellular and Molecular Biology, National Institute on Aging, NIH, Baltimore, MD 21224, USA. ⁶⁰Department of Biological Sciences, Macquarie University, Sydney NSW 2109, Australia. ⁶¹Center of Marine Biotechnology, UMBI, Columbus Center, Baltimore, MD 21202, USA. ⁶²Department of Cell Biology and Anatomy, Louisiana State University Health Sciences Center, New Orleans, LA 70112, USA. ⁶³Department of Biology, University of Victoria, Victoria, BC, Canada, V8W 2Y2. ⁶⁴Department of Neuroscience, Uppsala University, Uppsala, Sweden. ⁶⁵Laboratory of Cellular and Molecular Biophysics, National Institute of Child Health and Development, NIH, Bethesda, MD 20895, USA. ⁶⁶Developmental Unit, EMBL, 69117 Heidelberg, Germany. ⁶⁷Computational Unit, EMBL, 69117 Heidelberg, Germany. ⁶⁸Biotechnology Insti-

tute, Universidad Nacional Autónoma de Mexico (UNAM), Cuernavaca, Morelos, Mexico 62250. ⁶⁹Department of Cellular and Developmental Biology "Alberto Monroy," University of Palermo, 90146 Palermo, Italy. ⁷⁰Laboratoire de Biologie du Développement (UMR 7009), CNRS and Université Pierre et Marie Curie (Paris 6), Observatoire Océanologique, 06230 Villefranche-sur-Mer, France. ⁷¹Department of Biology, University of Patras, Patras, Greece. ⁷²Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. ⁷³Departament de Genètica, Universitat de Barcelona, 08028–Barcelona, Spain. ⁷⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. ⁷⁵Institut Jacques Monod, CNR-UMR 7592, 75005 Paris, France. ⁷⁶Consiglio Nazionale delle Ricerche, Istituto di Biomedicina e Immunologia Molecolare "Alberto Monroy," 90146 Palermo, Italy. ⁷⁷Razavi-Newman Center for Bioinformatics, Salk Institute for Biological Studies, La Jolla, CA 92186, USA. ⁷⁸Department of Zoology, University of Hawaii at Manoa, Honolulu, HI 96822, USA.

*Present address: GlaxoSmithKline, 1250 South Collegeville Road, Collegeville, PA 19426, USA.

†Present address: Massachusetts General Hospital Cancer Center, Charlestown, MA 02129, USA.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5801/941/DC1

Materials and Methods

SOM Text

Figs. S1 to S6

Tables S1 to S8

Reference

8 August 2006; accepted 17 October 2006

10.1126/science.1133609

REVIEW

Genomic Insights into the Immune System of the Sea Urchin

Jonathan P. Rast,^{1*} L. Courtney Smith,² Mariano Loza-Coll,¹ Taku Hibino,¹ Gary W. Litman^{3,4}

Comparative analysis of the sea urchin genome has broad implications for the primitive state of deuterostome host defense and the genetic underpinnings of immunity in vertebrates. The sea urchin has an unprecedented complexity of innate immune recognition receptors relative to other animal species yet characterized. These receptor genes include a vast repertoire of 222 Toll-like receptors, a superfamily of more than 200 NACHT domain–leucine-rich repeat proteins (similar to nucleotide-binding and oligomerization domain (NOD) and NALP proteins of vertebrates), and a large family of scavenger receptor cysteine-rich proteins. More typical numbers of genes encode other immune recognition factors. Homologs of important immune and hematopoietic regulators, many of which have previously been identified only from chordates, as well as genes that are critical in adaptive immunity of jawed vertebrates, also are present. The findings serve to underscore the dynamic utilization of receptors and the complexity of immune recognition that may be basal for deuterostomes and predicts features of the ancestral bilaterian form.

Animal immune mechanisms are classified as acquired (adaptive), in which immune recognition specificity is the product of somatic diversification and selective clonal proliferation, or as innate, in which recognition specificity is germline encoded. Collectively, these systems act to protect the individual from invasive bacteria, viruses, and eukaryotic pathogens by detecting molecular signatures of infection and initiating effector responses. Innate immune mechanisms probably originated early

in animal phylogeny and are closely allied with wound healing and tissue maintenance functions. In many cases, their constituent elements are distributed throughout the cells of the organism. In bilaterally symmetrical animals (Bilateria), immune defense is carried out and tightly coordinated by a specialized set of mesoderm-derived cells that essentially are committed to this function (1–3). Overlaid onto this conserved core of developmental and immune programs are a variety of rapidly evolving recognition and

effector mechanisms, which likely are responsive to the dynamic nature of host-pathogen interactions (4) and are among the most rapidly evolving animal systems (5).

For a variety of reasons, the field of immunology has been overwhelmingly focused on the rearranging adaptive immune system, which is based on the activities of immunoglobulin and T cell-antigen receptors (TCR) and which, at this point, seems to be restricted to the jawed vertebrates (6). Interest in comparative approaches to immunity was broadened by the recognition of common features of innate immunity between *Drosophila melanogaster* (fruit fly) and mammals (7, 8). Recent findings suggest that somatic mechanisms of receptor diversification analogous to those of the acquired system of jawed vertebrates may be a more widespread feature of animal immunity than previously supposed. Examples of these include a gene conversion-like process that diversifies variable leucine-rich repeat (LRR)-containing receptor (VLR) proteins in jawless vertebrates (9, 10), somatic mutation of fibrinogen-related protein (FREP) receptors in a mollusc (11), and extensive alternative splicing of the Down syndrome cell adhesion molecule (DSCAM), a molecule that principally guides neuronal patterning, to generate immune reactive isoforms in insects (12, 13). On the basis of this narrow sampling, it is likely that a universe of novel and dynamic immune mechanisms exists among the invertebrates, further validating their role as significant immune models.

Of the ~30 bilaterian phyla that are recognized, only chordates, molluscs, nematodes, arthropods, and echinoderms have been the subject of extensive molecular immune research (Fig. 1). The overwhelming majority of functional and genetic data regarding immune systems comes from just two animal phyla: Chordata (mainly from mammals) and Arthropoda (*D. melanogaster*). Comprehensive genomic analyses of immunity also have been conducted in three other invertebrate species, the sea squirt (*Ciona intestinalis*) (14), the mosquito (*Anopheles gambiae*) (15), and the nematode worm (*Caenorhabditis elegans*) (16). More focused molecular studies include investigations of an immunelike transplantation reaction in *Botryllus schlosseri* (a urochordate) (17) and the immune response of a gastropod mollusc, *Biomphalaria glabrata*, to trematode parasites (11). Here we describe high-

lights from a community-wide genome analysis effort (18) on the purple sea urchin, *Strongylocentrotus purpuratus*, a member of the phylum Echinodermata, with both biological and phylogenetic attributes that are of compelling interest from an immune perspective.

Genes Related to Immune Function in the Sea Urchin

It is likely that between 4 and 5% of the genes identified in the sea urchin genome are involved directly in immune functions (18). Considering only those components that exhibit distinct homology to forms found in other phyla, the repertoire of immune-related genes (18) that has been shown to participate in the recognition of conserved pathogen-associated molecular patterns (PAMPs) includes 222 Toll-like receptor (TLR) genes, 203 NACHT domain-LRR (NLR) genes with similarity to vertebrate nucleotide-binding and oligomerization domain (NOD)/NALP cytoplasmic receptors (19), and a greatly expanded superfamily of 218 gene models encoding scavenger receptor cysteine-rich (SRCR) proteins (20, 21). In considering these estimates, it is critical to note that the sea urchin genome sequence was derived from sperm taken from a single animal (18). Although in certain cases inadvertent inclusion of both haplotypes in genome assembly may artificially inflate estimations of complexity of multigene families, this risk is likely to be small for the gene sets that we report here and, in any event, would not change the major conclusion of the findings [see supporting online material (SOM) for a more detailed explanation]. Furthermore, gene expansion is not a uniform characteristic of immune genes in sea urchin. Other classes of immune mediators, such as key components of the complement system, peptidoglycan-recognition proteins (PGRPs), and Gram-negative binding proteins (GNBPs) are equivalent in numbers to their homologs in protostomes and other deuterostomes.

Of the three major expansions of multigene families encoding immune genes, the TLRs are particularly informative. Two broad categories of these genes can be recognized: a greatly expanded multigene family consisting of 211 genes and a more limited group of 11 divergent genes (22), which includes 3 genes with ectodomain structures characteristic of most protostome TLR proteins, such as *Drosophila Toll* (23) (Fig. 2A). The latter findings suggest that TLRs of this form were present in the common bilaterian ancestor and subsequently were lost in the vertebrate lineage. The expanded set of sea urchin TLRs (211 genes) consists of vertebrate-like structures, of which many appear to have been duplicated and diversified recently. Within subfamilies of these vertebrate-like genes [defined by clustering in phylogenetic analysis (Fig. 2B)], hypervariability is regionalized in particular LRRs (22). These patterns of

intergenic variation and the high prevalence of apparent pseudogenes (25 to 30%) suggest that the evolution of the sea urchin TLR genes is dynamic with a high gene turnover rate and could reflect rapidly evolving recognition specificities. By comparison, the relatively few TLR genes found in vertebrates derive from an ancient vertebrate diversification that appears to have been stabilized by selection for binding to invariant PAMPs (24).

It is unclear at present what aspects of sea urchin biology drive the differences in size and diversity of the expanded multigene families of innate receptors (we speculate on this below), but the characteristics of the TLR genes and their putative downstream signal mediators may have some bearing on their mode of function. It is likely that such a large and variable family recognizes pathogens directly rather than through intermediate molecules, as reported in insects (25). The moderate expansion of immediate downstream adaptors of TLR signaling that contain the Toll-interleukin 1 receptor (TIR) domain (four Myd88-like and 22 other cytoplasmic TIR domain adaptor genes) may serve to partition cellular responses after recognition by different classes of TLR proteins. In contrast, the lack of multiplicity of genes encoding the kinases and of transcription factors further downstream in the TLR signaling pathway resembles that observed in other species (22). This narrowed molecular complexity from the cell surface to the nucleus may mean that specificity of downstream cellular responses with respect to activation by different TLRs (if it exists) arises within the context of their restricted expression, as is the case for diversity in vertebrate adaptive systems. In certain general respects, the patterns of variation (Fig. 2B), the apparently rapid gene turnover rate, and the tandem genetic linkage of TLRs (Fig. 2C) resemble the multiplicity and diversity of the germline components of somatically variable adaptive immune receptors of vertebrates (6) and, taken together, they suggest that similar selective forces have molded their function.

Diverse TLRs are expressed by coelomocytes in the sea urchin (22). Furthermore, marked variation in the relative levels of expression is seen for different TLR subfamilies that is not strictly correlated with gene family size (fig. S1). In principle, restricted combinatorial expression of TLRs on individual immunocytes could generate a highly diverse range of individual functional specificities and, if shown to be the case, would provide one explanation for the observed patterns of TLR diversity. Combinatorial utilization within the more limited range of TLRs has been shown for mammals (26).

Some sea urchin TLR subgroup members are linked in large tandem arrays of identically oriented genes that appear to have been duplicated and diversified recently (Fig. 2C). Within this genomic context, the possibility exists for exclusive regulatory control. Both the linkage in direct tandem arrays and intergenic sequence

¹Sunnybrook Research Institute and Department of Medical Biophysics, University of Toronto, 2075 Bayview Avenue, Room 5-126B, Toronto, Ontario M4N 3M5, Canada.

²Department of Biological Sciences, George Washington University, 2023 G Street, NW, Washington, DC 20052, USA. ³Department of Pediatrics, University of South Florida (USF) College of Medicine, USF/ACH (All Children's Hospital) Children's Research Institute, St. Petersburg, FL 33701, USA. ⁴H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL 33612, USA.

*To whom correspondence should be addressed. E-mail: jrast@sri.utoronto.ca

Sea Urchin Genome

identity of the TLRs may promote gene diversification through duplication and/or deletion, gene conversion, recombination, and meiotic mispairing of alleles, followed by unequal crossovers as has been shown for plant disease resistance genes (27). The clustered genomic organization of sea urchin TLR genes resembles that seen in olfactory receptors, which exhibit clonal restriction in the absence of DNA-level rearrangement (28, 29). As innate immune systems reach higher levels of complexity, it is plausible that increased evolutionary pressure would drive the immune response toward regulation through isotype- and/or allele-restricted expression, cellular selection, and expansion, characteristics that we traditionally ascribe to adaptive immune receptors in vertebrates. The boundaries between germline-encoded innate receptors (e.g., vertebrate and insect TLRs) and the somatically variable adaptive immune receptors of vertebrates are becoming increasingly less distinct (30, 31).

Whereas the TLRs are the most readily characterized family of diversified innate receptors in sea urchin genome sequence and thus the focus of discussion here, a similar expansion is seen in other multigene families encoding immune proteins (Fig. 2A). NLR genes, which have been described previously only from vertebrates, serve as pathogen recognition receptors (PRRs) that detect cytoplasmic PAMPs (19) and are associated with immunity and autoimmune disease in the gut (32). The number and complexity of the more than 200 sea urchin NLR genes stand in distinct contrast to the ~20 NLR proteins in vertebrates. The gut is a major site of transcription of the NLRs in sea urchin (22), and gut-related immunity is likely a driving force behind expansion of this gene family. *S. purpuratus* is an herbivore, and much of its diet is kelp; various symbionts likely degrade complex carbohydrates and toxic compounds. Specific NLR-types and possibly TLR-types, as has been shown for vertebrates (33), may play a role in maintaining a balance with symbionts. Like the TLRs and NLRs, the multidomain SRCR genes of the sea urchin are expanded to unprecedented degrees (Fig. 2A).

These genes encode proteins with structural similarity to some vertebrate scavenger receptors that have been ascribed roles in innate immune recognition (34). More than 1000 SRCR domains are encoded in 218 gene models, exceeding by 10-fold the number of SRCR domains seen in humans. Diverse members of this gene family are expressed in coelomocytes and exhibit dynamic shifts in transcription after immune challenge (21).

are secreted from and localized to the surface of a subset of coelomocytes (37). The *185/333* genes represent another family of tightly linked and diverse immune-type genes (35, 38). Another large gene family that is implicated in the response of the sea urchin to immune challenge includes ~100 small C-type lectin and galectin genes. These examples, in addition to the TLRs, NLRs, and SRCRs, underscore a complex immune system in the sea urchin where large gene families, many with closely linked members, may be of significant importance.

The Origins of Vertebrate Immune Systems

Some of the most intriguing questions facing evolutionary immunology concern our limited understanding of the deuterostome underpinnings of the jawed-vertebrate immune system. The sea urchin genome, which encodes mediators of immunity that are shared with vertebrates but are absent in those protostomes for which whole-genome information is available, fills an essential gap in our recently broadened view of the immune system. As emphasized elsewhere in this issue, the overall complexity of the regulatory control networks, as well as the structures and genomic organization of their constituent elements, are highly significant in understanding the evolution of complex integrated systems such as those regulating immunity. Representatives of all important lymphocyte transcription factor subfamilies can be identified, including a deuterostome-restricted PU.1/SpiB/SpiC Ets transcription factor (a gene family that is intimately connected to blood cell functions in vertebrates) and an Ikaros/Aiolos/Helios/Eos-related gene (22). Immune signaling mediators, including a family of interleukin (IL)-17 genes, the IL receptors IL-1R and IL-17R, and tumor

necrosis factor family members that were previously known only from chordates or vertebrates, are present in the sea urchin genome (22). It seems that the gene regulatory tool kit encoded in the sea urchin is remarkably complete as compared with immunity in the jawed vertebrates, which raises new questions about alternative functions of regulatory elements that we tend to associate with the basic development and differentiation of vertebrate immunocytes.

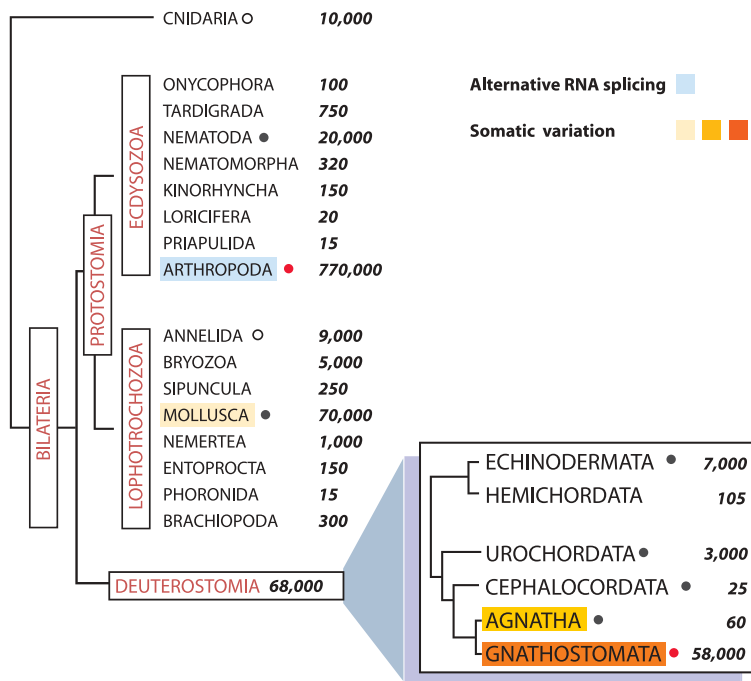


Fig. 1. A simplified phylogenetic tree depicting the general relationships of the major bilaterian phyla and chordate subphyla, highlighting select species that use different somatic mechanisms of immune receptor diversification. Red dots designate animal groups where the vast majority of immune data have been derived. Solid black dots denote taxa in which species have been the subject of extensive molecular immune research. Circles denote phyla where some molecular data are available. Color variation (see key) over specific phyla denotes the presence of a major somatic mechanism of receptor diversification in at least one representative member (6) and is not intended to be mutually exclusive. In the case of somatic variation, shade intensity indicates the level of empirically established diversity. Innate immune receptors, including TLRs, are likely present in all of the phyla. Numbers given beside taxa names are approximate estimates of species diversity and are presented to underscore the immense variety of immune mechanisms that have not yet been investigated [primarily taken from the Tree of Life Web project (44)]. Cnidarians (e.g., jellyfishes and sea anemones) are shown as an outgroup to the Bilateria. This view is not intended to represent all known species in which immune-type mediators have been identified.

There are a number of additional expanded gene families in the sea urchin genome that encode proteins with immune-related functions. The *185/333* genes were first noted because they are sharply up-regulated in response to whole bacteria and lipopolysaccharide (2, 35). Transcripts of the *185/333* genes constitute up to 6.5% of message prevalence in activated coelomocytes (36). The encoded novel proteins are highly diversified and

Rag1 and Rag2 represent the principle mediators of the somatic rearrangement process that is common to both immunoglobulin and T cell-antigen receptor gene families that effect adaptive immunity in jawed vertebrates. Whereas a number

of conventional approaches failed to identify homologs of these genes in jawless vertebrates and invertebrates, genomic analysis has identified Rag1 core region-like transposable elements and partial Rag1-like genes in a variety of invertebrates

(39). The identification of a homologous, Rag1/2-like functional gene cluster was one of the most unexpected findings from the sea urchin genome (40), as the transposon-like character of the vertebrate Rag genes suggests that they may have been acquired through a process of horizontal gene transfer at the time of the emergence of rearranging TCR and immunoglobulin gene systems in a jawed-vertebrate common ancestor. Although it is unclear at present whether or not these genes are active in immunity, it is improbable that they emerged independently in an echinoderm. The most parsimonious explanation for the distribution of Rag1/2-like clusters in two major deuterostome clades is that it represents a shared genetic feature present in a common ancestral deuterostome. Alternatively, the Rag1/2-like gene cluster may represent the independent cooption of an as yet unknown transposon that encoded both Rag1- and Rag2-like genes.

In addition to the Rag1/2-like cluster, several other components related to those that function in the somatic reorganization and diversification of immunoglobulin and TCR also have been identified, including a polymerase that is homologous to the common ancestor of terminal deoxynucleotidyl transferase (TdT) and polymerase μ . Finally, several families of immunoglobulin domain genes (a total of about 50) have been identified that are predicted to encode immunoglobulin variable-type (V) domains similar to those used by adaptive immune receptors of jawed vertebrates, and also the VCBPs, a diversified family of nonrearranging immune-type receptors in cephalochordates (31). Notably a cluster of V-type immunoglobulin genes is encoded adjacent to a large cluster of TLR genes (Scaffold_V2_74946; Fig. 2C) in the current assembly, although this will need to be independently verified (fig. S2). These V-type immunoglobulin domain structures uniformly lack canonical recombination signal sequences, which represent an integral component of DNA-mediated recombination and, thereby, the generation of a complex immune repertoire. Elucidating the function of these genes in a species where Rag1/2-like genes are present, but the process of variable-(diversity)-joining [V(D)J] segmental recombination of antigen binding receptors is absent, is potentially useful for understanding the origins of the segmental rearrangements of immunoglobulin domains in the adaptive immune receptors of jawed vertebrates.

Conclusions

The current data inform us about the evolution of immunity from multiple perspectives. First, this genome sequence significantly refines our understanding of deuterostome immunity. Immune factors previously known only from chordates and often only from vertebrates (e.g., IL-1R, IL-17, PU.1/SpiB/SpiC, NOD/NALP-like receptors) can be attributed now to the common deuterostome ancestor shared by echinoderms and chordates. Next, this genome is informative in comparison with

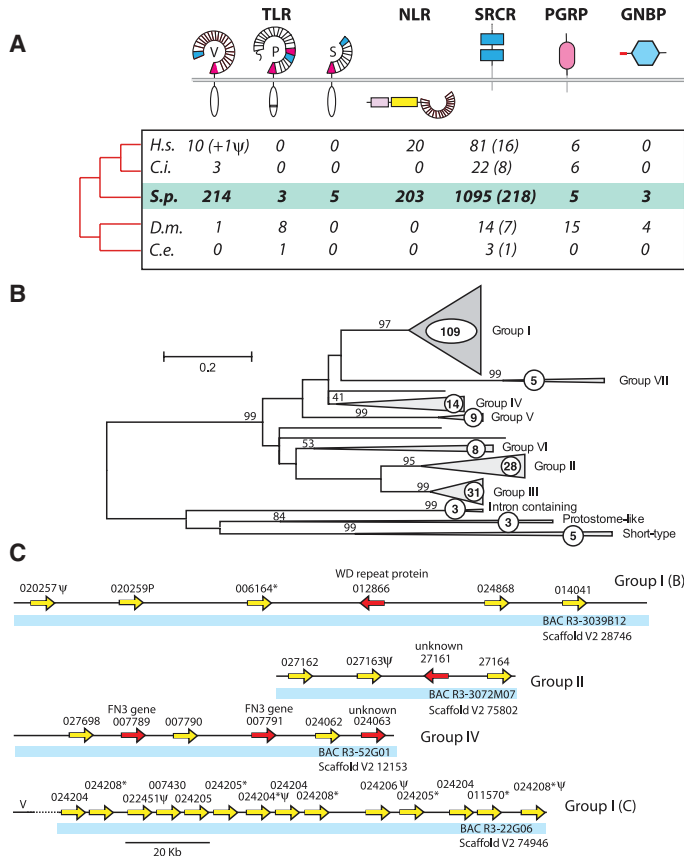


Fig. 2. Innate immune receptor multiplicity in the sea urchin genome sequence. **(A)** Comparison of gene families encoding innate immune receptors in representative animals with sequenced genomes to *S. purpuratus* (bold, hereafter designated sea urchin). For some key receptor classes, gene numbers in the sea urchin exceed those of other animals by more than an order of magnitude. Representative animals are *Homo sapiens*, *H.s.*; *C. intestinalis*, *C.i.*; *S. purpuratus*, *S.p.*; *D. melanogaster*, *D.m.*; and *C. elegans*, *C.e.* Gene families include TLRs, NLRs, SRCRs, PGRPs, and GNBP. Specifically, TLR diagrams show V, vertebrate-like, P, protostome-like; and S, short type; oval indicates TIR domain; and segmented partial circles indicate LRR regions; LRR-NT, blue; and LRR-CT, red. NLR diagram shows death family domain in pink, NACHT domain in yellow, and the LRR region, for which horizontal orientation implies cytoplasmic function. The other diagrams show multiple SRCR genes (both secreted and transmembrane), PGRP genes (PFAM: Amidase_2 domain-containing, secreted or transmembrane); and GNBP proteins (PFAM: Glyco_hydro_16-containing, secreted). For multiple SRCR genes, representative values are domain number (gene number in parentheses). For *C. intestinalis*, numbers correspond to all annotated SRCR proteins. Phylogenetic relations among species are indicated by the red cladogram at the left of the table; diagrams of molecules are not intended to imply specific structural features. **(B)** Unrooted neighbor-joining tree showing interrelations of TIR domains of TLRs in sea urchin. TLRs can be classified into three divergent classes (protostome-like, intron-containing, and short) and a large sea urchin lineage-specific family, which distributes into seven (I to VII) subgroups; numbers of member genes indicated in circles. Group I can be further subdivided [I(A) to I(E)]. Numbers beside branches indicate % bootstrap support for each subgroup. Efforts to relate vertebrate and other TLRs to the sea urchin genes result in low-confidence affinities with the divergent groups as described for other TLR comparisons (24). **(C)** Clustering of representative sea urchin TLR genes (yellow arrows) from high-confidence regions of the assembly supported by bacterial artificial chromosome (BAC) sequence (indicated by blue bar). Clusters segregate according to groups I(B) and I(C) are subgroups of group I. Gene model numbers are indicated above arrows. Model numbers with asterisks are close matches to annotated gene models and likely represent the second haplotype to that which was used to create models from the previous assembly. Red arrows indicate non-TLR genes. V indicates putative position of a V-type immunoglobulin domain cluster. Verification of cluster organization will require further independent genomic analysis. ψ signifies pseudogene. Scale is indicated in kb (kilobase pairs).

Sea Urchin Genome

protostomes as protostome-like TLRs are present in the sea urchin genome and likely were present in the common bilaterian ancestor. Another perspective is defined by those components of the sea urchin genome that are related to the basic structural units of the antigen-binding receptors, as well as to the genes encoding the molecular machinery that effects somatic diversification of immunoglobulin and TCRs in jawed vertebrates. Finally, the genome sequence reveals adaptations that appear to be specific to the sea urchin lineage. Most strikingly, the expansion of gene families encoding innate immune recognition receptors is unlike that seen in any species characterized to date. Not only are the numbers of genes increased, but they reveal distinct patterns of variation, suggesting that they function through gradations in specificity that, in turn, may reflect differences in either the pathogens they recognize and/or the manner in which they cope with nonself on a systemwide basis.

The complexity of the sea urchin innate immune receptor superfamilies may be driven by the same selective forces that mold the vertebrate adaptive system. Alternatively, this innate complexity may relate to unique aspects of sea urchin biology. It is difficult to ignore that sea urchins are particularly long-lived [*S. purpuratus* lives to >30 years, and a closely related congener has been dated to more than 100 years (41)] and that their body size is large relative to other invertebrates with sequenced genomes. Other aspects of its basic biology may also be important, including its nonreduced genome, enormous numbers of progeny, and a biphasic life history. Finally, features of its life-style, including the complex relationship it probably exhibits with symbionts, could factor in the specialization of immune mechanisms as discussed for vertebrate systems (33, 42) and for other physiological adaptations in marine organisms (43). One clear conclusion to be derived from the sea urchin genome is that the complexity of immunological mechanisms among unexplored animal phyla (Fig. 1) is likely to rival that found across the vertebrate-invertebrate (or agnathan-gnathostome) divergence.

Despite the entirely likely and intriguing links between sea urchin and vertebrate immunity, genomics only can take us so far in understanding complex regulatory and functional relations. However, the dichotomy observed in the complexity of genes encoding innate receptors within the deuterostomes provides a particularly well-defined starting point for further investigations. Clearly, the LRR proteins (TLRs and NLRs) have proven to be evolutionarily malleable in the context of sea urchin immunity. Many features of the organization and regulation of the particularly large diversified multigene families of immune receptors are consistent with potential restricted expression of individual genes in coelomocytes, which are basic characteristics of the lymphocyte- and natural killer cell-based immune systems of vertebrates (42). The experimental accessibility of the sea urchin will allow ready answers to questions of restricted

expression and the nature of the regulatory interface between the apparently ancient networks that underpin animal immunocyte specification and the more evolutionarily labile immune mechanisms that mediate their differentiated functions.

References and Notes

1. E. H. Davidson, D. H. Erwin, *Science* **311**, 796 (2006).
2. J. P. Rast, Z. Pancer, E. H. Davidson, *Curr. Top. Microbiol. Immunol.* **248**, 3 (2000).
3. C. J. Evans, V. Hartenstein, U. Banerjee, *Dev. Cell* **5**, 673 (2003).
4. P. M. Murphy, *Cell* **72**, 823 (1993).
5. A. L. Hughes, *Mol. Biol. Evol.* **14**, 1 (1997).
6. G. W. Litman, J. P. Cannon, L. J. Dishaw, *Nat. Rev. Immunol.* **5**, 866 (2005).
7. J. A. Hoffmann, *Nature* **426**, 33 (2003).
8. C. A. Janeway Jr., R. Medzhitov, *Annu. Rev. Immunol.* **20**, 197 (2002).
9. Z. Pancer *et al.*, *Nature* **430**, 174 (2004).
10. M. N. Alder *et al.*, *Science* **310**, 1970 (2005).
11. S. M. Zhang, C. M. Adema, T. B. Kepler, E. S. Loker, *Science* **305**, 251 (2004).
12. F. L. Watson *et al.*, *Science* **309**, 1874 (2005).
13. Y. Dong, H. E. Taylor, G. Dimopoulos, *PLoS Biol.* **4**, e229 (2006).
14. K. Azumi *et al.*, *Immunogenetics* **55**, 570 (2003).
15. G. K. Christophides *et al.*, *Science* **298**, 159 (2002).
16. A. C. Millet, J. J. Ewbank, *Curr. Opin. Immunol.* **16**, 4 (2004).
17. A. W. De Tomaso *et al.*, *Nature* **438**, 454 (2005).
18. Sea Urchin Genome Sequencing Consortium, *Science* **314**, 941 (2006).
19. T. A. Kufer, J. H. Fritz, D. J. Philpott, *Trends Microbiol.* **13**, 381 (2005).
20. Z. Pancer, J. P. Rast, E. H. Davidson, *Immunogenetics* **49**, 773 (1999).
21. Z. Pancer, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13156 (2000).
22. T. Hibino *et al.*, *Dev. Biol.* 10.1016/j.ydbio.2006.08.065 (2006).
23. F. L. Rock, G. Hardiman, J. C. Timans, R. A. Kastelein, J. F. Bazan, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 588 (1998).
24. J. C. Roach *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9577 (2005).
25. S. Akira, S. Uematsu, O. Takeuchi, *Cell* **124**, 783 (2006).
26. A. Ozinsky *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13766 (2000).
27. S. H. Hulbert, C. A. Webb, S. M. Smith, Q. Sun, *Annu. Rev. Phytopathol.* **39**, 285 (2001).
28. J. Li, T. Ishii, P. Feinstein, P. Mombaerts, *Nature* **428**, 393 (2004).
29. B. M. Shykind *et al.*, *Cell* **117**, 801 (2004).
30. M. F. Flajnik, L. Du Pasquier, *Trends Immunol.* **25**, 640 (2004).
31. J. P. Cannon, R. N. Haire, N. Schnitker, M. G. Mueller, G. W. Litman, *Curr. Biol.* **14**, R465 (2004).
32. J. P. Ting, D. L. Kastner, H. M. Hoffman, *Nat. Rev. Immunol.* **6**, 183 (2006).
33. S. Rakoff-Nahoum, J. Paglino, F. Eslami-Varzaneh, S. Edberg, R. Medzhitov, *Cell* **118**, 229 (2004).
34. S. Mukhopadhyay, S. Gordon, *Immunobiology* **209**, 39 (2004).
35. D. P. Terwilliger, K. M. Buckley, D. Mehta, P. G. Moorjani, L. C. Smith, *Physiol. Genomics* **26**, 134 (2006).
36. S. V. Nair, H. Del Valle, P. S. Gross, D. P. Terwilliger, L. C. Smith, *Physiol. Genomics* **22**, 33 (2005).
37. K. M. Brockton, L. C. Smith, unpublished observations.
38. V. Buckley, L. C. Smith, unpublished observations.
39. V. V. Kapitonov, J. Jurka, *PLoS Biol.* **3**, e181 (2005).
40. S. D. Fugmann, C. Messier, L. A. Novack, R. A. Cameron, J. P. Rast, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3728 (2006).
41. T. A. Ebert, J. R. Southon, *Fish. Bull. (Washington, DC)* **101**, 915 (2003).
42. Z. Pancer, M. D. Cooper, *Annu. Rev. Immunol.* **24**, 497 (2006).
43. E. S. Loker, C. M. Adema, S. M. Zhang, T. B. Kepler, *Immunol. Rev.* **198**, 10 (2004).
44. Tree of Life Web project (www.tolweb.org).
45. We thank the Human Genome Sequencing Center at the Baylor College of Medicine for assistance throughout this analysis. We greatly appreciate helpful discussions with C. Messier, Z. Pancer, S. Fugmann, D. Philpott, and S. Girardin. We thank K. M. Buckley for comments on the manuscript, B. Pryor for editorial assistance, B. Jasny for very helpful input on the organization of these findings, and two anonymous reviewers for enlightening observations. This work was made possible by support from the NSF (MCB-0424235) to L.C.S.; the Uehara Memorial Foundation to T.H.; the Canadian Institutes for Health Research (MOP74667), and National Science and Engineering Research Council of Canada (458115/211598) to J.P.R.; and the NIH (AI23338) to G.W.L.

Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5801/952/DC1

Materials and Methods

SOM Text

Figs. S1 and S2

Table S1

References

10.1126/science.1134301

REVIEW

Paleogenomics of Echinoderms

David J. Bottjer,^{1*} Eric H. Davidson,² Kevin J. Peterson,³ R. Andrew Cameron²

Paleogenomics propels the meaning of genomic studies back through hundreds of millions of years of deep time. Now that the genome of the echinoid *Strongylocentrotus purpuratus* is sequenced, the operation of its genes can be interpreted in light of the well-understood echinoderm fossil record. Characters that first appear in Early Cambrian forms are still characteristic of echinoderms today. Key genes for one of these characters, the biomineralized tissue stereom, can be identified in the *S. purpuratus* genome and are likely to be the same genes that were involved with stereom formation in the earliest echinoderms some 520 million years ago.

Paleogenomics is the addition of the component of deep time to the field of genomics (1). Initial studies have concentrated on reconstructing regions of the ancestral mam-

malian genome (2) or sequencing preserved DNA of recently extinct organisms, such as the woolly mammoth (3). Although such studies present many exciting possibilities,