

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,

the prospects for paleogenomics are much broader.

Genomics offers the opportunity of identifying genes that are responsible for the evolution of key shared characters of organisms, or synapomorphies, which are ultimately used to reconstruct the tree of life. Paleogenomics thus allows for both the geologic and genetic fossil records to shed light on the origin and subsequent evolution through time of key genes and the key synapomorphies that they encode.

Echinoderm Paleogenomics and the Stereom Skeleton

The initial appearance of biomineralized skeletal tissues in the fossil record (4), just before the beginning of the Cambrian ~542 million years ago (Ma), coincides with the start of the rapid increase in the diversity of metazoans termed the Cambrian explosion (5). Later in the Early Cambrian, by 520 Ma, a variety of biomineralized skeletal structures had appeared. Among the most distinctive is a major echinoderm synapomorphy: the unique endoskeletal tissue called stereom.

Stereom is composed of calcite organized into a meshlike structure (Fig. 1, A to D), the pores of which in life are populated with dermal cells and fibers (6). Much is known about this structure from studies of representatives of the approximately 7000 species that constitute the five clades of living echinoderms (crinoids or sea lilies, ophiuroids or brittle stars, asteroids or sea stars, holothuroids or sea cucumbers, and echinoids or sea urchins), all of which produce stereom endoskeletons (6). Stereom forms structural elements (Fig. 1) that can be embedded in soft tissues or may be fused together to form larger compound plates, generating the various types of echinoderm skeletons (6).

Because the high-magnesium calcite of which stereom is constructed is stable, echinoderm skeletons are very durable during the process of fossil preservation, and this has led to an abundant and well-understood echinoderm fossil record (6, 7).

Echinoderm Phylogeny and the History of Stereom

Echinoderms, one of the three major phyla of the Deuterostomia, make their first appearance in the fossil record during the Early Cambrian, about 520 Ma, but the most primitive echinoderms are the stylophorans, a bizarre group first recorded in the fossil record in the Middle Cambrian (~510 Ma) (8) (Figs. 2 and 3A). Stylophorans are

recognized as echinoderms because they possess stereom (Fig. 1, C and D). Another major echinoderm synapomorphy is the water vascular system (8, 9), a closed circulatory system that uses ambient seawater to provide the hydraulic force necessary to extend the tube feet of living forms. The water vascular system first made its appearance in another group of Cambrian forms, the solutes (Figs. 2 and 3B). Both stylophorans and solutes are “stem-group” echinoderms (8) because they are more closely related to living echinoderms than they are to living hemichordates, the closest living relatives of modern echinoderms, but they are not descended from the last common ancestor of the living echinoderms (Fig. 2).

Other stem-group echinoderms that produced both a biomineralized stereom skeleton and plate morphologies indicating the presence of a water vascular system appear in the stratigraphic record from about 520 Ma (10). These include the helioplacoids, eocrinoids, and edrioasteroids (Fig. 3, C to E). Pentamer symmetry of the adult body, a highly characteristic echinoderm synapomorphy of crown-group echinoderms, makes its initial appearance in the edrioasteroids and the eocrinoids (10).

Crown-group echinoderm fossils (Fig. 2) occur in the earliest Ordovician at about 485 Ma, in the form of primitive crinoids, another immobile filter-feeding group (11) (Figs. 2 and 3F). The remaining crown-group echinoderms have mobile life habits on and in the seafloor and as a group are termed eleutherozoans (asteroids, ophiuroids, holothuroids, and echinoids) (Fig. 3, G and H). Eleutherozoans with biomineralized stereom ossicles also first appear in the earliest Ordovician (~450 Ma) (13) (Fig. 3H). The modern forms trace their roots back to the Late Permian, when the first cidaroid echinoids (“pencil-spined”

sea urchins) appeared (14). Perhaps only two echinoid lineages survived the end-Permian mass extinction ~252 Ma. *Strongylocentrotus purpuratus*, the modern sea urchin whose genome sequence is now available, is a regular euechinoid.

Echinoderm Biomineralization: Cell Biology and Genes

The process by which the biomineralized stereom skeleton is formed in echinoids is coming to light, through combined approaches of cell, molecular, and developmental biology (15). First, embryonic mesenchymal cells secrete the earliest portions of the skeleton to appear. The larval spicules (Fig. 1E) and additional independent sites in the larva are the starting points for the adult plates. The biomineral is composed of calcite (CaCO₃) containing 5% MgCO₃. It is

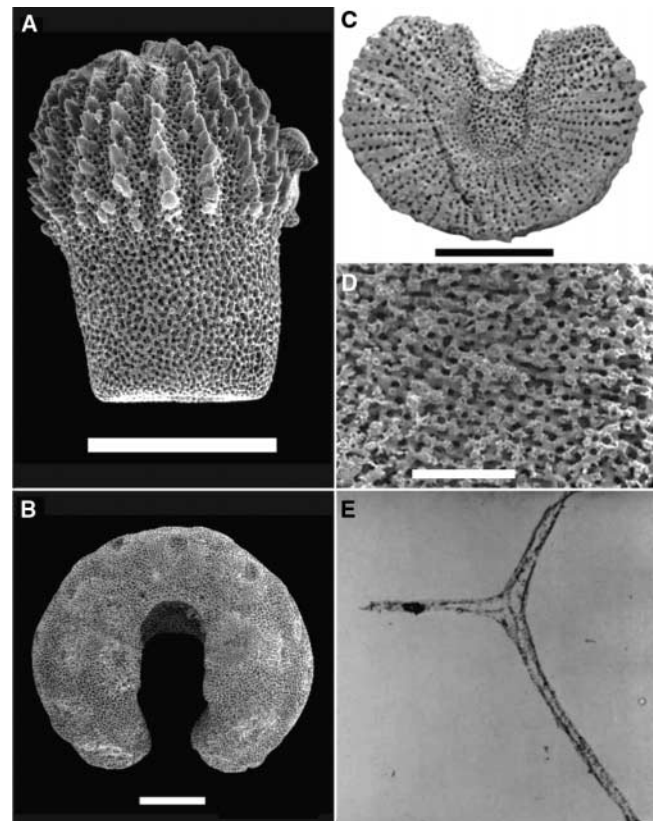


Fig. 1. Stereom formation in modern and fossil echinoderms. (A) A spine of the modern sea star *Asterias* with soft tissue removed, constructed of meshlike stereom. (B) An ocular plate of the modern sea star *Asterias* with soft tissue removed, constructed of meshlike stereom. (C) Median cross section of the stylocone from the Middle Cambrian stylophoran *Ceratocystis*, showing stereom construction. (D) Detail of stereom from the inner face in external view of the stylocone from the Middle Cambrian stylophoran *Ceratocystis*. (E) Larval spicule of *S. purpuratus* with biomineralized stereom (calcite) dissolved away, showing the distribution of spicule matrix proteins. [(A) and (B) are used with permission from C. Sumrall; (C) and (D) are reprinted by permission from Macmillan Publishers Ltd. (*Nature*) (11), copyright (2005); (E) is reprinted from (27), copyright 1983, with permission from Elsevier.] Scale bars in (A) to (C), 500 μ m; scale bar in (D), 100 μ m; magnification of (E) is \times 3000.

¹Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA. ²Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA. ³Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA.

*To whom correspondence should be addressed. E-mail: dbottjer@usc.edu

Sea Urchin Genome

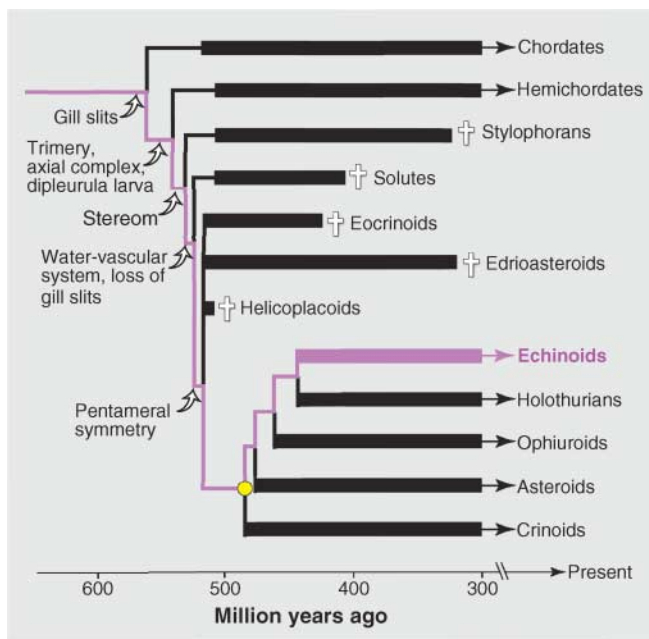


Fig. 2. Evolutionary history of the major echinoderm groups. Deuterostomia consists of three major groups: the chordates, hemichordates, and echinoderms, all with fossil representatives in the Cambrian. Cambrian echinoderms are recognized by the possession of stereom, but the phylogenetically most basal groups (such as stylophorans) lack the water vascular system, are highly asymmetrical, and possess gill slits. Pentamerism is seen in two major Early Cambrian lineages, the edrioasteroids and eocrinoids; a third Early Cambrian taxon, the helicoplacoids, have an unusual threefold symmetry thought to be derived from the ancestral pentamerism arrangement (10). All stem-group echinoderm lineages became extinct by the Carboniferous (indicated with crosses). Crown-group echinoderms, indicated by the yellow circle, consist of the five major extant lineages in addition to numerous extinct lineages not shown. Most class-level crown groups first appear in the latest Paleozoic–early Mesozoic, including echinoids. The lineage leading to echinoids, and hence to *S. purpuratus*, is indicated in purple. Known stratigraphic ranges are shown with thick lines, and inferred range extensions are shown with thin lines.

secreted into an extracellular space, probably sequestered from the surrounding environment, initially as amorphous calcium carbonate, which then undergoes a regulated transition to the crystalline form (16).

Occluded within the calcite is an organic matrix of proteins that make up about 0.1% of the mass, and the birefringent optical properties of the skeletal elements result from the regular alignment of the crystals in the matrix (17). As shown in Fig. 1E (18), which portrays a skeletal element after demineralization, the triadial physical form is a property of the matrix proteins, which originally were deposited with the biomineral. Additionally, there is an envelope of proteins around the mineral portion. Initial surveys indicate that a large number of separable proteins or protein derivatives is associated with the mineral (15). Something is known of the structure and deployment of seven of these proteins, and four have been studied in detail, namely SM50, SM30, SM37, and PM27 (15).

small clusters and thus are likely to be the related products of local gene family expansions. For example, four related SM30 genes were found to be arranged in tandem on a single assembly scaffold (21). In addition, the SM37 gene is closely linked to the SM50 gene (21–23).

The *S. purpuratus* sequence also contains homologs of many of the signaling molecules and extracellular matrix proteins involved in vertebrate biomineralization (21). But in contrast to these, the sea urchin C-lectin spicule matrix proteins share little similarity with the well-characterized vertebrate skeletogenic proteins. Nor are they similar to any sequences present in current databases of expressed sequence tags (ESTs) from the hemichordate *Saccoglossus kowalewskyi* or the cephalochordate *Branchiostoma floridae*. With respect to any other known genome, the spicule matrix proteins of echinoids are encoded by a clade-specific set of genes. This may be true for echinoderms in general, but there is too little sequence data from other classes to

make the conclusion definitive. The stereom structure is so similar among the classes that it would be remarkable if these proteins were not a character of the phylum.

The molecular and cell biology of stereom biomineralization in the sea urchin offers a fascinating glimpse into the genetic underpinnings of an echinoderm synapomorphy that arose in the Early Cambrian. A suite of identifiable unique genes (except that they have in common a domain encoding a calcium-dependent lectin) evolved to construct the unique biomineral structure of the stereom. The basic pattern of fenestration in the stereom (Fig. 1) is the property of a single differentiated cell type, defined by the expression of a battery of matrix genes, which first appeared in echinoderms at least 520 Ma.

Little is known, though, about the exact functions of these proteins except that interference with the expression of SM50 inhibits spicule formation in *S. purpuratus* embryos (20). The sea urchin genome project revealed the seven known spicule matrix genes and eight new ones as well (21). Furthermore, the genome sequence provides the opportunity to observe the arrangement of the spicule matrix protein genes, and the results illuminate an aspect of their evolution.

These genes occur in small clusters and thus are likely to be the related products of local gene family expansions. For example, four related SM30 genes were found to be arranged in tandem on a single assembly scaffold (21). In addition, the SM37 gene is closely linked to the SM50 gene (21–23).

The multiplicity of the spicule matrix proteins is reflected by structural and functional variety within this sample, though all have a C-type lectin domain, which is a calcium-dependent carbohydrate binding motif. SM30 and SM37 are glycosylated. SM30 is known to occupy the occluded protein compartment, whereas SM50 and PM27 are found occluded and in the extracellular matrix around the spicule. SM50 contains an unusual proline- and glycine-rich repeat sequence similar to the pericardin repeat motif (19).

Discussion

Paleogenomics adds a genomic dimension to the paleontological description of synapomorphies. Stereom is an iconic synapomorphy for echinoderms, much as bone is for the vertebrates, and it was the first to arise in the divergence of echinoderms from the other major deuterostome lineages. Thus, we hypothesize that the evolution of the spicule matrix genes occurred after the Precambrian-Cambrian boundary (542 Ma), but before the time in the Early Cambrian when stereom-containing fossils first appear (~520 Ma) (Fig. 2). The specific prediction is that a unique echinoderm synapomorphy, a definitive property not shared with phylogenetic sister groups, will be the genomic constituents of the calcite/stereom differentiation gene battery (Table 1). That is, echinoderms will in general share variants of the same biomineralization genes and use the same transcriptional regulatory controllers of these genes. This is of course open to experimental verification by comparative molecular analysis of biomineralization in modern forms of echinoderm. Were it found that the genetic repertoire used to produce stereom in the diverse echinoderm classes is indeed similar, then it would be indisputable that their stem-group ancestors used the same genetic apparatus.

Discussion

Paleogenomics adds a genomic dimension to the paleontological description of synapomorphies. Stereom is an iconic synapomorphy for echinoderms, much as bone is for the vertebrates, and it was the first to arise in the divergence of echinoderms from the other major deuterostome lineages. Thus, we hypothesize that the evolution of the spicule matrix genes occurred after the Precambrian-Cambrian boundary (542 Ma), but before the time in the Early Cambrian when stereom-containing fossils first appear (~520 Ma) (Fig. 2). The specific prediction is that a unique echinoderm synapomorphy, a definitive property not shared with phylogenetic sister groups, will be the genomic constituents of the calcite/stereom differentiation gene battery (Table 1). That is, echinoderms will in general share variants of the same biomineralization genes and use the same transcriptional regulatory controllers of these genes. This is of course open to experimental verification by comparative molecular analysis of biomineralization in modern forms of echinoderm. Were it found that the genetic repertoire used to produce stereom in the diverse echinoderm classes is indeed similar, then it would be indisputable that their stem-group ancestors used the same genetic apparatus.

Paleogenomics is a knife that cuts two ways: We gain insights not only into the genes that built the structures of our fossils, but also into the evolutionary origin of the gene networks that operate in the construction of modern animal body parts. In this case, we propose that the specific stereom matrix gene battery (that is, the variety of structural functions encoded in its diverse proteins, plus its regulatory controls) must have been assembled as such in Early Cambrian time. It has remained a feature of echinoderm genomes ever since. Something is already known of the regulatory network apparatus controlling spicule matrix protein expression in the *S. purpuratus* embryo. The differentiation genes of the biomineralization gene battery of this embryo are

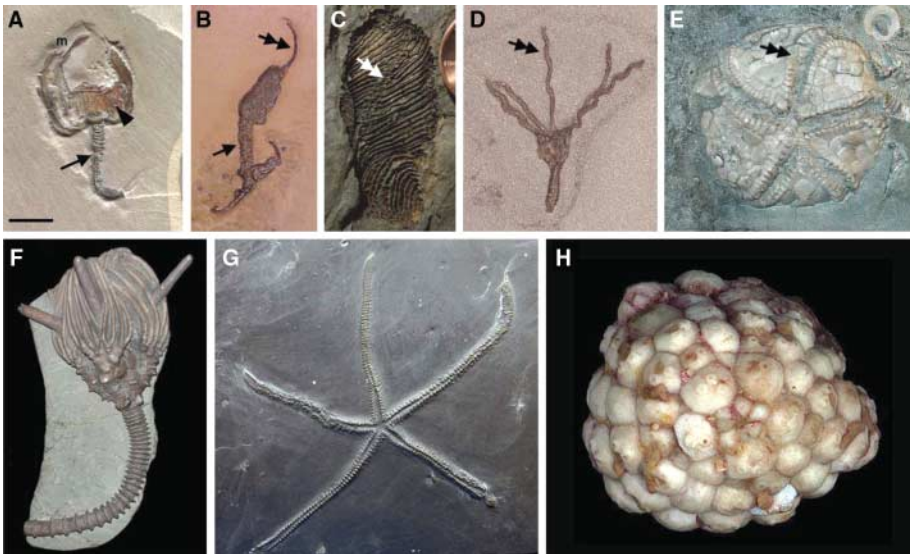


Fig. 3. Stem-group (A to E) and crown-group (F to H) echinoderms. (A) The stylophoran *Cothurnocystis bifida* (Middle Cambrian, Utah, USA). The putative gill skeletons as viewed from the back side are indicated with an arrowhead. M is the putative mouth. The arrow indicates the posterior appendage. (B) The solute *Coleicarpus sprinklei* (Middle Cambrian, Utah, USA). The arrow indicates the posterior appendage, and the double arrow points to the single ambulacrum. (C) The helicoplacoid *Helicoplacus* (Early Cambrian, California, USA). The double arrow points to one of the ambulacral grooves. (D) The eocrinoid *Gogia spiralis* (Middle Cambrian, Utah, USA). The double arrow points to one of the five arms. (E) The edrioasteroid *Edriophus bigsbyi* (Ordovician, Ontario, Canada). It displays conspicuous pentamerous symmetry; one of the arms is indicated by the double arrow. (F) The crinoid *Dorycrinus mississippiensis* (Mississippian, Indiana, USA). (G) The asteroid *Furcaster palaeozoicus* (Devonian, Budenbach, Germany). (H) The echinoid *Bothriocidaris* (Ordovician, Estonia) [reprinted with permission from A. B. Smith, from (30)]. Scale bar, 0.5 cm in (A); 0.75 cm in (B) to (E); 2 cm in (F); 1.3 cm in (G); and 0.15 cm in (H). Part of a penny is shown for scale in (C).

Table 1. Predicted features of the echinoderm biomineralization gene battery extrapolated from *S. purpuratus*.

Functions	Characteristics
Regulatory apparatus	Multiple specific regulatory genes (such as <i>Alx1</i> , <i>Ets1</i> , <i>Dr</i> , and <i>Hnf6</i>) with feed-forward input into biomineralization genes (31)
Cellular biology	Non–echinoderm-specific molecules used for secretion and motility (21)
Biomineralization genes	Echinoderm-specific genes featuring glycine- and proline-rich repeats on the same protein with C-type lectin domains (15, 21)

together regulated by a specific small set of transcriptional control genes (24, 25).

Paleogenomic approaches can be extended to other clades for which there is both a sequenced genome and a well-preserved fossil record. The objective is the identification of clade-specific gene batteries that encode clade-specific features of the body plan. What emerges will add a time dimension to specific parts of the underlying gene regulatory networks. Thus we will be able to “age-date” portions of the functional genome to determine parts of genomes that are relatively young, in contrast to others that are extraordinarily old.

It is unusual in the consideration of body plan evolution to be able to cite a phylum-specific set of structural or differentiation genes, of which it can literally be said that their evolution underlay

a phylum-specific morphological feature. The organization of body plans obviously depends in general on the regulatory control of the developmental process, which in turn depends at the genomic level on the organization of developmental gene regulatory networks. In general, therefore, the evolution of diversity of body plans depends on changes in the architecture of gene regulatory networks. But the regulatory genes constituting these networks, which encode transcription factors and intercellular signaling components, are notoriously not clade-specific: They are largely pan-bilaterian (26–28), if not pan-metazoan. Similarly, the downstream differentiation genes that produce the proteins from which major components of the body plan are constructed are often not clade-specific either. For example, muscles, nervous systems, and

hearts use many orthologous genes across the Bilateria. Clade-specific sets of genes that are often noted in animal genome sequences, such as genes of the immune system or smell receptor genes, are not very likely to produce signatures in the fossil record.

In temporal and historical aspects, as well as in architectural and functional terms, genetic systems for the control of development are internally inhomogeneous (27). Some subcircuits are very ancient and have changed little since their early evolution hundreds of millions of years ago; others are of more recent origin and have arisen in given evolutionary branches. This view is, of course, inconsistent with the microevolutionary presumption of temporal uniformity in evolutionary processes. The broad objectives of paleogenomics are convergent with those of “regulatory phylogenetics” [for example, gene regulatory network comparisons (29)], in that both result in the association of given genetic components with specific time-resolved evolutionary nodes. A distinction, as in the example described in this paper, is the direct relation between genes producing a structure and the fossil record, rather than the indirect relation between the regulatory genes and the body plan. It is satisfying to be able to apply genomic data directly to the origins of a character that arose over half a billion years ago and is found in animals present on Earth today.

References and Notes

1. D. Birnbaum, F. Coulter, M.-J. Pebusque, P. Pontarotti, *J. Exp. Zool.* **288**, 21 (2000).
2. M. Blanchette, E. D. Green, W. Miller, D. Haussler, *Genet. Res.* **14**, 2412 10.1101/gr.2800104 (2004).
3. H. N. Poinar *et al.*, *Science* **311**, 392 (2006).
4. J. P. Grotzinger, W. A. Watters, A. H. Knoll, *Paleobiology* **26**, 334 (2000).
5. K. J. Peterson, M. A. McPeck, D. A. D. Evans, *Paleobiology* **31** (suppl.), 36 (2005).
6. A. B. Smith, in *Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends*, vol. 1, J. G. Carter, Ed. (Van Nostrand, New York, 1990), pp. 413–443.
7. E. Flügel, *Microfacies of Carbonate Rocks* (Springer-Verlag, Berlin, 2004).
8. A. B. Smith, *Geol. J.* **40**, 255 (2005).
9. S. Clausen, A. B. Smith, *Nature* **438**, 351 (2005).
10. J. Sprinkle, B. C. Wilbur, *Geol. J.* **40**, 281 (2005).
11. T. E. Guensburg, J. Sprinkle, *Geology* **29**, 131 (2001).
12. D. B. Blake, T. E. Guensburg, *J. Paleontol.* **79**, 395 (2005).
13. A. B. Smith, J. J. Savill, *Trans. R. Soc. Edinb. Earth Sci.* **92**, 137 (2001).
14. A. B. Smith, N. T. J. Hollingworth, *Proc. Yorkshire Geol. Soc.* **48**, 47 (1990).
15. F. H. Wilt, *Dev. Biol.* **280**, 15 (2005).
16. E. Beniash, J. Aizenberg, L. Addadi, S. Weiner, *Proc. R. Soc. London Ser. B* **264**, 461 (1997).
17. A. Berman *et al.*, *Science* **250**, 664 (1990).
18. S. Benson, E. M. E. Jones, N. Crise-Benson, F. Wilt, *Exp. Cell Res.* **148**, 249 (1983).
19. Y. Katoh-Fukui *et al.*, *Dev. Biol.* **145**, 201 (1991).
20. M. Peled-Kamar, P. Hamilton, F. H. Wilt, *Exp. Cell Res.* **272**, 56 (2002).
21. B. T. Livingston *et al.*, *Dev. Biol.* 10.1016/j.ydbio.2006.07.047 (2006).
22. Y.-H. Lee, R. J. Britten, E. H. Davidson, *Dev. Growth Differ.* **41**, 303 (1999).

23. M. R. Illies, M. T. Peeler, A. M. Dechtiaruk, C. A. Etnessohn, *Dev. Genes Evol.* **212**, 419 (2002).
24. P. Oliveri, E. H. Davidson, *Curr. Opin. Genet. Dev.* **14**, 351 (2004).
25. G. Amore, E. H. Davidson, *Dev. Biol.* **293**, 555 (2006).
26. V. F. Hinman, A. T. Nguyen, R. A. Cameron, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13356 (2003).
27. D. H. Erwin, E. H. Davidson, *Development* **129**, 3021 (2002).
28. E. H. Davidson, D. H. Erwin, *Science* **311**, 796 (2006).
29. E. H. Davidson, *The Regulatory Genome. Gene Regulatory Networks in Development and Evolution* (Academic Press/Elsevier, San Diego, CA, 2006).
30. The Echinoid Directory (www.nhm.ac.uk/research-curation/projects/echinoid-directory).
31. G. Amore, E. H. Davidson, *Dev. Biol.* **293**, 555 (2006).
32. This work was partially supported by NSF grant IOB-0212869 (to R.A.C.), NIH grant RR-15044 (to E.H.D.), and the Caltech Beckman Institute. D.J.B. is supported by NASA, NSF, and the University of Southern California; K.J.P. is supported by NSF, NASA-Ames, and Dartmouth College.

10.1126/science.1132310

REPORT

The Transcriptome of the Sea Urchin Embryo

Manoj P. Samanta,¹ Waraporn Tongprasit,^{2,3} Sorin Istrail,^{4,5} R. Andrew Cameron,⁵ Qiang Tu,⁵ Eric H. Davidson,⁵ Viktor Stolc^{2*}

The sea urchin *Strongylocentrotus purpuratus* is a model organism for study of the genomic control circuitry underlying embryonic development. We examined the complete repertoire of genes expressed in the *S. purpuratus* embryo, up to late gastrula stage, by means of high-resolution custom tiling arrays covering the whole genome. We detected complete spliced structures even for genes known to be expressed at low levels in only a few cells. At least 11,000 to 12,000 genes are used in embryogenesis. These include most of the genes encoding transcription factors and signaling proteins, as well as some classes of general cytoskeletal and metabolic proteins, but only a minor fraction of genes encoding immune functions and sensory receptors. Thousands of small asymmetric transcripts of unknown function were also detected in intergenic regions throughout the genome. The tiling array data were used to correct and authenticate several thousand gene models during the genome annotation process.

Embryogenesis in the sea urchin occurs rapidly and is relatively simple in form (1). By 2 days after fertilization, when the embryo is in the late gastrula stage, there are about 800 cells and 10 to 15 cell types. Thus, genes expressed in individual cell types or territories represent a larger fraction of the total number of transcripts than do genes expressed in adult organs of vertebrates or in more complex embryos such as that of *Drosophila*. Earlier studies have provided extensive quantitative evidence on transcript prevalence for sea urchin embryos, both for populations of mRNA (and nuclear RNA) and for many individual transcripts, measured by quantitative polymerase chain reaction (QPCR) (2–4). The genome sequence of *Strongylocentrotus purpuratus* (5) enabled these advantages to be exploited for a whole-genome tiling array analysis of the embryonic transcriptome.

Transcriptome analysis by whole-genome tiling array (6–9) has three advantages relative to standard microarray analysis with oligonucleotide probes constructed on the basis of known or predicted protein-coding genes: (i)

The genes identified are not limited a priori by the gene predictions used to design the probes and therefore are not biased in favor of more prevalent or more conserved sequences; (ii) the transcripts detected will include noncoding as well as protein-coding RNAs; and (iii) intron-exon boundaries plus untranslated regions (UTRs) are revealed. In comparison with expressed sequence tag (EST) or cDNA-based approaches, whole-genome tiling arrays offer an unbiased and complete view of the transcriptional activity of the genome in the developmental state examined and in addition display the intron and exon structures of expressed genes. In itself, tiling array data cannot assign a distant exon to its gene, but this shortcoming can be overcome by integrating tiling and EST/cDNA data for genome annotation.

Tiling array experiments have traditionally been performed only several years after genome sequencing (9). However, maskless array synthesizer technology permitted us to develop custom arrays from preliminarily assembled draft sequence. This initiative enhanced the genome project while it was still in process, by substantially reducing the gap between sequencing and comprehensive annotation of the genome.

To sample transcriptional activity throughout early sea urchin development on a single set of high-density microarrays, we prepared polyadenylated RNA from egg, early blastula (15 hours), early gastrula (30 hours), and late

gastrula stage (45 hours) embryos. Samples were mixed in equal quantities, reverse transcribed, fluorescently labeled, and hybridized. The tiling array probes were designed from the initial draft assembled sequence, which at that time was based on 6× whole-genome shotgun sequence coverage (5). A total of 10,133,868 50-nucleotide (nt) probes were selected to uniformly represent the entire sea urchin genome, maintaining an average spacing of 10 nt between consecutive probes (table S1). Repetitive sequences and simple sequence tracts were excluded. The probes were synthesized on 27 glass-based microarrays. To avoid any potential bias due to cutoff selection based on unexpressed genomic probes, we also added a set of 1000 random sequences not represented anywhere in the genome to each array. The cutoff was such that only 1% of those random probes were falsely expressed. Additionally, each array included a small (2000) identical set of genomic control probes used for normalization purposes. After hybridization, data from all arrays were normalized according to the control probes, mapped back to the latest genome sequence assembly, and mounted on a genome browser together with the optimal set of computationally derived gene models [OGS set in (5); for visual presentation of all transcriptome results as in Fig. 1A, see www.systemix.org/sea-urchin]. Details of the methods used are available in the Supporting Online Material (10), and the microarray designs and experimental data have been deposited in the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (www.ncbi.nlm.nih.gov/geo) under the accession code GSE6031.

Analysis of signals for 28 well-characterized genes (11) (table S2) showed that the array measurements were highly sensitive. When mapped against the known structure of these genes, it was apparent that transcribed regions were clearly distinguished from silent regions, and no intronic transcripts were detected. Intron-exon boundaries of expressed genes were thus clearly distinguishable (e.g., Fig. 1A, fig. S1). To establish a conservative statistical criterion of expression, we first established the background variance and chose a cutoff value about 2.5 times that of the mean background. At this value, about 1% of random control probes displayed apparently artifactual noise, e.g., single-point peaks over background surrounded by probes at the background level (as in the single-

¹Systemix Institute, Los Altos, CA 94024, USA. ²NASA Ames Genome Research Facility, Moffet Field, CA 94035, USA. ³Eloret Corporation, Sunnyvale, CA 94086, USA. ⁴Brown University, Providence, RI 02912, USA. ⁵California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: vstolc@arc.nasa.gov