Instructions for Handling PDF Galley Proofs

Dear Author:

Thank you for publishing in Science. This letter explains how to use this PDF file for transmitting corrections to your galley proofs.

This PDF file includes the following:

1. A three-page reprint order form (p. 2).
2. Galley proofs of your paper (starting on p. 5).

A separate PDF file of the copyedited version of your manuscript showing editorial changes has been, or shortly will be, e-mailed to you directly by the copy editor of your paper. Please use this in checking your galley proofs.

In order to make galley corrections:

1. All changes must be marked on the galley proofs (this file). This can be either a printed copy that is returned to us (by FAX, express mail, or a scanned copy) or you may instead use the notes or other tools in Adobe Acrobat and return this file as an email attachment.
2. Mark all of your changes clearly on the printed galley proofs (not the edited manuscript). Please keep changes to an absolute minimum. If alterations are excessive, publication of your paper may be delayed and you will be billed for additional typesetting expenses.
3. Respond to all of the copy editor’s queries listed at the end of the edited manuscript or as embedded PDF notes.
4. Check all equations and tables carefully.
5. Check that any special characters, such as Greek letters, have translated correctly.
6. Make a copy of the corrected galley proofs for yourself.
7. Return the corrected galley proofs to Science by email (a scanned copy or the file marked using Acrobat), FAX or express mail directly to the Science copy editor (their contact information appears on the first page of the edited manuscript). You can also e-mail a list of corrections to the copy editor.
8. If you would like to order reprints, return the reprint order form separately to The Sheridan Press at the address noted on the form, not to Science.

It is important that you return galley corrections promptly, within 48 hours. Please let your copy editor know immediately if there will be any delay.

If your manuscript contains color figures, they will appear within the PDF galley proofs; however, the colors and resolution may appear different from those of the final published figures. Our art department will send separate color figure proofs to you. Note that although your paper currently begins at the top of a page, it may not do so when printed in Science. Your paper’s position will be determined later when the issue is imposed.

Sincerely,
The Editors
Publication Fees/Author Color Reimbursement for Use of Color in Science

As stated in Information for Contributors and your acceptance letter, authors requesting the use of color are required to pay $650 for the first color figure and $450 for each additional figure to help defray costs related to publishing color in the Science issue. Please remember that these charges are not related to your reprint order, but may be billed together on a single invoice. Authors of solicited Reviews, Viewpoints, and Special Issue Reviews are exempt from these charges.

Reprints

Ordering Information
Authors of Research Articles, Reports, Brevia, and Technical Comments, use Form A. Authors of Reviews, Viewpoints, and Special Issue Reviews, use Form B (scroll down).

Pricing
Reprint pricing is shown in the following tables. Orders are limited to 1,000 copies per author. To convert color articles to black & white reprints, add $100 for each color figure to be converted. For articles over twelve months past publication, please contact The Sheridan Press for pricing.

Black and White Reprints

<table>
<thead>
<tr>
<th>Quantity</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 page</td>
<td>245</td>
<td>252</td>
<td>260</td>
<td>268</td>
<td>285</td>
<td>301</td>
<td>317</td>
<td>333</td>
<td>349</td>
<td>368</td>
</tr>
<tr>
<td>8 page</td>
<td>378</td>
<td>396</td>
<td>415</td>
<td>433</td>
<td>464</td>
<td>494</td>
<td>531</td>
<td>558</td>
<td>590</td>
<td>620</td>
</tr>
<tr>
<td>12 page</td>
<td>498</td>
<td>534</td>
<td>566</td>
<td>598</td>
<td>650</td>
<td>694</td>
<td>742</td>
<td>788</td>
<td>836</td>
<td>889</td>
</tr>
<tr>
<td>16 page</td>
<td>618</td>
<td>667</td>
<td>713</td>
<td>761</td>
<td>825</td>
<td>889</td>
<td>958</td>
<td>1020</td>
<td>1084</td>
<td>1151</td>
</tr>
</tbody>
</table>

Color Reprints

<table>
<thead>
<tr>
<th>Quantity</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 page</td>
<td>1719</td>
<td>1728</td>
<td>1736</td>
<td>1743</td>
<td>1756</td>
<td>1766</td>
<td>1776</td>
<td>1786</td>
<td>1796</td>
<td>1805</td>
</tr>
<tr>
<td>8 page</td>
<td>1967</td>
<td>1999</td>
<td>2032</td>
<td>2055</td>
<td>2093</td>
<td>2124</td>
<td>2159</td>
<td>2186</td>
<td>2215</td>
<td>2243</td>
</tr>
<tr>
<td>12 page</td>
<td>2558</td>
<td>2589</td>
<td>2618</td>
<td>2638</td>
<td>2678</td>
<td>2703</td>
<td>2730</td>
<td>2757</td>
<td>2783</td>
<td>2813</td>
</tr>
<tr>
<td>16 page</td>
<td>2570</td>
<td>2614</td>
<td>2657</td>
<td>2692</td>
<td>2742</td>
<td>2782</td>
<td>2826</td>
<td>2865</td>
<td>2902</td>
<td>2941</td>
</tr>
</tbody>
</table>

Air Shipping Charges
(for orders outside the continental US)

<table>
<thead>
<tr>
<th>Quantity</th>
<th>100</th>
<th>200</th>
<th>300</th>
<th>400</th>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 page</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td>65</td>
<td>75</td>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
</tr>
<tr>
<td>8 page</td>
<td>55</td>
<td>65</td>
<td>80</td>
<td>100</td>
<td>120</td>
<td>140</td>
<td>160</td>
<td>180</td>
<td>200</td>
<td>220</td>
</tr>
<tr>
<td>12 page</td>
<td>60</td>
<td>80</td>
<td>110</td>
<td>120</td>
<td>170</td>
<td>200</td>
<td>225</td>
<td>255</td>
<td>290</td>
<td>305</td>
</tr>
<tr>
<td>16 page</td>
<td>70</td>
<td>100</td>
<td>120</td>
<td>175</td>
<td>220</td>
<td>255</td>
<td>305</td>
<td>345</td>
<td>385</td>
<td>425</td>
</tr>
</tbody>
</table>
**FORM A: SCIENCE’S AUTHOR REPRINTS**

**ORDER FORM / PROFORMA INVOICE**

<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
<th>Issue date of SCIENCE</th>
<th>Page(s)</th>
</tr>
</thead>
</table>

Please check one:
- Research Article
- Report
- Brevia
- Technical Comment/Reply
- Other: ______________________

**Publication fees for use of color in article**

Publication fees will be invoiced, even if reprints are not ordered.

**Number of color figures in article:**  

<table>
<thead>
<tr>
<th>Charges for color usage ($650 for 1st color figure, $450 for each additional color figure(s)):</th>
<th>$</th>
</tr>
</thead>
</table>

**Reprints Costs**

Is there color in your paper? **Yes** / **No** (please circle)  
Do you want color in your reprints? **Yes** / **No** (please circle)

<table>
<thead>
<tr>
<th>Count of pages in article</th>
<th>Cover page (required):</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+1</td>
</tr>
</tbody>
</table>

Total pages in reprint:  

<table>
<thead>
<tr>
<th>Number of copies ordered</th>
<th>Reprint production charge (See Reprint Price Table on pg. 2)</th>
<th>$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Misc. charges:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Convert Color to Black and White ($100 per color figure)</td>
</tr>
<tr>
<td></td>
<td>Science Issue Cover ($100)</td>
</tr>
<tr>
<td></td>
<td>Outside continental U.S., please add air shipping charge:</td>
</tr>
</tbody>
</table>

2. **Total Reprint Charges:**  

<table>
<thead>
<tr>
<th></th>
<th>Total Reprint Charges &amp; Publication Fees (1 + 2):</th>
<th>$</th>
</tr>
</thead>
</table>

**Billing and Shipping Instructions**

Street address must be provided for reprint orders; P.O. box numbers are not acceptable for billing and shipping address.

Send reprints to:  

<table>
<thead>
<tr>
<th>Send reprints to:</th>
<th>Street address</th>
<th>City</th>
<th>State</th>
<th>Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Send reprints to:</th>
<th>Street address</th>
<th>City</th>
<th>State</th>
<th>Zip Code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional shipping location (Attach additional instruction sheet as necessary.)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity</td>
<td>Send reprints to:</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Additional shipping location</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill to:</td>
<td>Street address</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BILLING LOCATION (Attach additional instruction sheet as necessary.)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity</td>
<td>Send reprints to:</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total Reprint Charges &amp; Publication Fees (1 + 2):</th>
<th>$</th>
</tr>
</thead>
</table>

**Payment Method (please check payment method enclosed)**

- Personal check (checks must be paid in U.S. dollars and drawn on a U.S. bank, payable to The Sheridan Press)
- Institution purchase order #
- **MC / VISA / American Express** (Please circle) Number ___________________________ Exp date ___________  

Signature: ___________________________ Date: ___________________________

Signature is required for all payment methods. **Reprint order will not be processed without signature!**  
By signing this form, the requestor agrees to accept the responsibility for the payment of reprints.
### FORM B: REPRINTS FOR SOLICITED MANUSCRIPTS
ORDER FORM / PROFORMA INVOICE

---

**Waiver of publication fees**

< Must have prior approval from Editorial Office for waiver of Publication Fees. >

(For use with solicited Reviews, Special Issue Reviews, and solicited Viewpoints)

**Author**

**Title**

**Issue date of SCIENCE**

**Page(s)**

Please check one:

- Review
- Special Issue Review
- Viewpoint
- Other: ___________________________

---

**Reprints Costs**

<table>
<thead>
<tr>
<th>Is there color in your paper?</th>
<th>Yes / No</th>
<th>Do you want color in your reprints?</th>
<th>Yes / No</th>
</tr>
</thead>
</table>

- **Count of pages in article**
- **Cover page (required):** +1
- **Total pages in reprint**
- **Number of copies ordered**

**Reprint production charge** *(See Reprint Price Table on pg. 2)*

- $ 

**Misc. charges:**

- Convert Color to Black and White *($100 per color figure)*
- Science Issue Cover *($100)*

**Outside continental U.S., please add air shipping charge:**

Total Reprint Charges: $ 

---

**Billing and Shipping Instructions**

Street address must be provided for reprint orders; P.O. box numbers are not acceptable for billing and shipping address.

**Send reprints to:**

- **Street address**
- **City**
- **State**
- **Zip Code**
- **Phone #**
- **Fax #**

**Additional shipping location** *(Attach additional instruction sheet as necessary.)*

- **Quantity**
- **Send reprints to:**
- **Street address**
- **City**
- **State**
- **Zip Code**
- **Phone #**
- **Fax #**

**Bill to:**

- **Street address**
- **City**
- **State**
- **Zip Code**
- **Phone #**
- **Fax #**

**Payment Method** *(please check payment method enclosed)*

- **Personal check** *(checks must be paid in U.S. dollars and drawn on a U.S. bank, payable to The Sheridan Press)*
- Institution purchase order
- **MC / VISA / American Express** *(Please circle)*
  - **Number**
  - **Exp date**

**Signature:** ___________________________  **Date:** ___________________________

Signature is required for all payment methods. **Reprint order will not be processed without signature!**

By signing this form, the requestor agrees to accept the responsibility for the payment of reprints.
The Genome of the Sea Urchin Strongylocentrotus purpuratus

George M. Weinstock,1,2,6 Richard A. Gibbs,1,2 Erica Sodergren,1,2 Eric H. Davidson,3 R. Andrew Cameron,3 The Sea Urchin Genome Sequencing Consortium†

We report the sequence and analysis of the 814-Mb genome of the sea urchin Strongylocentrotus purpuratus, a model for developmental and systems biology. The sequencing strategy combined whole-genome shotgun and bacterial artificial chromosome (BAC) sequences. This use of BAC clones, aided by a pooling strategy, overcame difficulties associated with high heterozygosity of the genome. The genome encodes about 23,300 genes, including many previously thought to be vertebrate innovations or known only outside the deuterostomes. This echinoderm genome provides an evolutionary outgroup for the chordates and yields insights into the evolution of deuterostomes.

The genome of the sea urchin was sequenced primarily because of the remarkable usefulness of the echinoderm embryo as a research model system for modern molecular, evolutionary, and cell biology. The sea urchin is the first animal with a sequenced genome that (i) is a free-living, motile marine invertebrate; (ii) has a bilaterally organized embryo for a century and a half, for most of that time, few were aware of one of the most important characteristics of sea urchins, a character that directly enhances its significance for genomic analysis; Echinoderms (and their sister phylum, the hemichordates) are the closest known relatives of the chordates (Fig. 1 and SOM). A description of the echinoderm body plan, as well as aspects of the life-style, longevity, polymorphic gene pool, and characteristics that but a radial adult body plan; (iii) has the endoskeleton and water vascular system found only in echinoderms; (iv) has an immune system that is unique in the enormous complexity of its receptor repertoire; (v) is remarkably long-lived with life spans of Strongylocentrotus species extending to over a century (see supporting on-line material (SOM)); (vi) is highly fecund, producing millions of gametes each year; and (vii) is a pivotal component of subtidal marine ecology and an important fishery catch in several areas of the world, including the United States. Although a research model in developmental biology for a century, and a half, for most of that time, few were aware of one of the most important characteristics of sea urchins, a character that directly enhances its significance for genomic analysis; Echinoderms (and their sister phylum, the hemichordates) are the closest known relatives of the chordates (Fig. 1 and SOM). A description of the echinoderm body plan, as well as aspects of the life-style, longevity, polymorphic gene pool, and characteristics that
Sea Urchin Genome

Fig. 1. The phylogenetic position of the sea urchin relative to other model systems and ourselves. The chordates are shown on the darker blue background overlying the deuterostomes as a whole on a lighter blue background. Organisms for which genome projects have been initiated or are finished are show across the top.

Meredith Howard-Ashby,1 Sorin Istrail,2 Pei Yun Lee,3 Annamaria Locascio,4 Pedro Martínez,7,2,7 Stefan C. Materna,5 Jongmin Nam,4 Paola Oliveri,3 Francesca Rizzo,3 Joel Smith,2 DNA sequencing: Donna Muzny1,2 (leader), Erica Sodergren1,2 (leader), Richard A. Gibbs1,2 (leader), George M. Weinstock1,2 (leader), Stephanie Bell,1,2 Joseph Chacko,2 Andrew Cree,2 Stacey Curry,2 Clay Davis,2 Huyen Dinh,2 Shannon Dugan-Rocha,1,2 Jerry Fowler,1,2 Rachel Gill,1,2 Cerrissa Hamilton,1,2 Judith Hernandez,1,2 Sandra Hines,1,2 Jennifer Hoe,1,2 LaRonda Jackson,1,2 Angela Jolivet,1,2 Christie Kowar,1,2 Sandra Lee,1,2 Lori Lewis,1,2 George Minor,1,2 Margaret Morgan,1,2 Lynne V. Nazareth,1,2 Geoffrey Okwuonu,1,2 David Parker,1,2 Ling-Ling Pu,1,2 Yufeng Shen,1,2 Rachel Thorn,1,2 Rita Wright,1,2

1Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. 2Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. 3Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. 4National Institute of Dental and Craniofacial Research, NIH, Bethesda, MD 20892, USA. 5Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy. 6Department of Biology, University of California, Santa Barbara, Santa Barbara, CA 93106, USA. 7Department of Animal Science, Texas A&M University, College Station, TX 77843, USA. 8Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA. 9National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA. 10Department of Marine and Cell Biology, University of California, Berkeley, Berkeley, CA 94720–1200, USA. 11Department of Biology, Polytechnic State University, San Luis Obispo, CA 93407, USA. 12Université Pierre et Marie Curie-Paris 6, UMR 7150, Equipe Cycle Cellulaire et Développement, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. 13CNRS, UMR 7150, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. 14Université Pierre et Marie Curie-Paris 6, UMR7628, Banyuls-sur-Mer, F-66650, France. 15University of Toronto, Toronto, Ontario, Canada, M4N 3M5. 16Department of Biotechnology, University of Southern California, Los Angeles, CA 90089–0740, USA. 17Department of Biology, University of Central Florida, Orlando, FL 32816–3368, USA. 18Department of Biological Sciences, Dartmouth College, Hanover, NH 03755, USA. 19Center for Computational Regulatory Genomics, Beckman Institute, California Institute of Technology, Pasadena, CA 91125, USA. 20Department of Biology and Biochemistry, University of Houston, Houston, TX 77204, USA. 21Sunnybrook Research Institute and Department of Cancer Research, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada, V6A 3N4. 22Department of Biology, George Washington University, Washington, DC 20052, USA. 23Royal Swedish Academy of Sciences, Kristineberg Marine Research Station, Fiskebäckskil, 450 34, Sweden. 24Marine Biology, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92039–0202, USA. 25Department of Molecular and Cellular Biology and Biochemistry, Brown University Providence, RI 02912, USA. 26Department of Biology and Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA. 27Department of Animal Science, Texas A&M University, College Station, TX 77843, USA. 28National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894, USA. 29Department of Ecology, Evolution, and Marine Biology, University of California Santa Barbara, Santa Barbara, CA 93106, USA. 30National Center for Biotechnology Information, NIH, Bethesda, MD 20894, USA. 31Penn Genomics Institute, University of Pennsylvania, Philadelphia, PA 19104, USA. 32Evolution and Development Group, MaxPlanck Institut fuer Molekulare Genetik, 14195 Berlin, Germany. 33Royal Holloway, University of London, Egham, Surrey TW20 0EX, UK. 34Center for Cancer Research, MIT, Cambridge, MA 02139, USA. 35Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA 94720–1200, USA. 36Department of Biology, California Polytechnic State University, San Luis Obispo, CA 93407, USA. 37Université Pierre et Marie Curie-Paris 6, UMR 7150, Equipe Cycle Cellulaire et Développement, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. 38CNRS, UMR 7150, Station Biologique de Roscoff, 29682 Roscoff Cedex, France. 39CNRS, UMR7628, Banyuls-sur-Mer, F-66650, France. 40Université Pierre et Marie Curie-Paris 6, UMR7628, Banyuls-sur-Mer, F-66650, France. 41Center for Bioinformatics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. 42Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. 43Tethys Research, LLC, 2115 Union Street, Bangor, Maine 04401, USA. 44Department of Molecular, Cellular, and Developmental Biology, University of California, Berkeley, CA 94720, USA. 45Department of Computational Molecular Biology, and Computer Science Department, Brown University, Providence, RI 02912, USA. 46Genome Research Facility, National Aeronautics and Space Administration, Ames Research Center, Moffet Field, CA 94035, USA. 47Systemix Institute, Cupertino, CA 95014, USA. 48Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia, Canada, V5A 1S6. 49Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada, V5A 1S6. 50Department of Biology, Center for Neuroscience, Uppsala University, Uppsala, Sweden. 51Department of Cellular and Molecular Biophysics, National Institute of Child Health and Development, NIH, Bethesda, MD 20895, USA. 52Department of Developmental Unit, EMBL, 69117 Heidelberg, Germany. 53Computational Unit, EMBL, 69117 Heidelberg, Germany. 54Biotechnology Institute, Universidad Nacional Autónoma de Mexico (UNAM), Cuernavaca, Morelos, Mexico 62250. 55Department of Biology, University of Patras, Patras, Greece. 56Laboratory of Biología del Desarrollo (UMR 7009), CNRS and Université Pierre et Marie Curie (Paris 6), Observatoire Oceánologique de Villefranche-sur-Mer, France. 57Department of Cell Biology and Anatomy, Louisiana State University Health Sciences Center, New Orleans, LA 70112, USA. 58Department of Biology, University of Victoria, Victoria, BC, Canada, V8W 2Y2. 59Department of Neuroscience, Uppsala University, Uppsala, Sweden. 60Department of Cell Biology and Anatomy, University of California San Diego, La Jolla, CA 92037, USA. 61Department of Genetics, George Washington University, Washington, DC 20052, USA. 62Royal Swedish Academy of Sciences, Kristineberg Marine Research Station, Fiskebäckskil, 450 34, Sweden. 63Department of Biological Sciences, University of Houston, Houston, TX 77204, USA. 64Department of Cancer Biology, British Columbia Cancer Agency, Vancouver, BC, Canada, V5Z 4E6. 65Department of Biology and the Institute of Systems Research, University of Maryland, College Park, MD 20742, USA. 66Department of Cellular and Molecular Biology, National Institute on Aging, NIH, Baltimore, MD 21224, USA. 67Department of Biological Sciences, Macquarie University, Sydney NSW 2109, Australia. 68Department of Marine Biotechnology, UMBI, Columbus Center, Baltimore, MD 21202, USA. 69Department of Cell Biology and Anatomy, Louisiana State University Health Sciences Center, New Orleans, LA 70112, USA. 70Department of Biology, University of Texas, M. D. Anderson Cancer Center, Charlestown, MA 02129, USA. 71Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. 72Department of Genetics, Universitat de Barcelona, 08028–Barcelona, Spain. 73Instituto Catalán de Recerca i Estudis Avancats (ICREA), Barcelona, Spain. 74Institut Jacques Monod, CNR-UMR 7592, 75005 Paris, France. 75Consejo Nacional de Investigaciones Científicas y de la Competencia, Madrid, Spain. 76Department of Developmental Biology, Albert Einstein College of Medicine, Palermo, 90146 Palermo, Italy. 77Department of Biologie du Développement (UMR 7009), CNRS and Université Pierre et Marie Curie (Paris 6), Observatoire Océanologique de Villefranche-sur-Mer, France. 78Department of Biology, University of Patras, Patras, Greece. 79Department of Molecular and Cellular Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA. 80Department of Genetics, Universitat de Barcelona, 08028–Barcelona, Spain. 81Department of Biotechnology, University of Massachusetts at Amherst, Amherst, MA 01003, USA. 82Present address: Massachusetts General Hospital Cancer Center, Charlestown, MA 02129, USA.
make sea urchin so valuable as a research organism, are presented in the SOM.

The last common ancestors at the branch points shown in Fig. 1 are of Precambrian antiquity (>540 million years ago (Mya)), according to protein molecular phylogeny, and give the appearance of animals belonging to basal clades of most current phyla in the Lower Cambrian fossil assemblages dating to 520 Mya. Cambrian echinoderms came in many distinct forms, but from their first appearance, the fossil record illustrates certain distinctive features that are still present: their water vascular system, including rows of tube feet protruding through holes in the ambulacral grooves and their calcite endoskeleton (mainly, a certain form of CaCO₃), which displays the specific three-dimensional structure known as “stereom.” The species sequenced, Strongylocentrotus purpuratus, commonly known as the “California purple sea urchin” is a representative of the thin-spined “modern” group of regularly developing sea urchins (echinoids). These evolved to become the dominant echinoid form after the great Permian-Triassic extinction 250 million years ago.

We present here a description of the S. purpuratus genome and gene products. The genome provides a wealth of discoveries about the biology of the sea urchin, Echinodermata, and the deuterostomes. Among the key findings are the following.

- The sea urchin is estimated to have 23,300 genes with representatives of nearly all vertebrate gene families, although often the families are not as large as in vertebrates.
- Some genes thought to be vertebrate-specific were found in the sea urchin (deuterostome-specific); others were identified in sea urchin but not the chordate lineage, which suggests loss in the vertebrates.
- Expansion of some gene families occurred apparently independently in the sea urchin and vertebrates.
- The sea urchin has a diverse and sophisticated immune system mediated by an astonishingly large repertoire of innate pathogen recognition proteins.
  - An extensive defense was identified.
  - The sea urchin has orthologs for genes associated with vision, hearing, balance, and chemosensation in vertebrates, which suggests additional sensory capabilities than previously thought.
  - Distinct genes for biomineralization exist in the sea urchin and vertebrates.
  - Orthologs of many human disease-associated genes were found in the sea urchin.

### Sequecing and Annotation of the S. purpuratus Genome

**Sequencing and assembly.** Sperm from a single male was used to prepare DNA for all libraries (tables S1 and S2) and whole-genome shotgun (WGS) sequencing. The overall approach was based on the “combined strategy” used for the rat genome (1), where WGS sequencing to six times coverage was combined with two times sequence coverage of BAC clones from a minimal tiling path (MTP) (fig. S1). The use of BACS provided a framework for localizing the assembly process, which aided in the assembly of repeated sequences and solved problems associated with the high heterozygosity of the sea urchin genome, without our resorting to extremely high coverage sequencing.

Several different assemblies were produced during the course of the project (see SOM for details). The Sea Urchin Genome Project (SUGP) was the first to produce both intermediate WGS assemblies and a final combined assembly. This was especially useful, not only for the early availability of an assembly for analysis, but also because WGS contigs were used to fill gaps between BACs in the combined assembly. The pure WGS assembly was produced (v 0.5 GenBank accession number range AAGJ01000001 to AAGJ01320773; also referred to as NCBI build 1.1) and released in April 2005. The final combined BAC-WGS assembly was released in July 2006 as version (v) 2.1 and submitted to GenBank (accession number range AAGJ02000001 to AAGJ02220581).

A second innovation in the SUGP was the use of the clone-array pooled shotgun sequencing (CAPSS) strategy (2) for BAC sequencing (fig. S2). The MTP consisted of 8248 BACs, and rather than prepare separate shotgun libraries from each of these, the CAPSS strategy involved BAC shotgun sequencing from pools of clones and then deconvoluting the reads to the individual BACs. This allowed the BAC sequencing to be done in 1/5th the time and at 1/10th the cost.

The principal new challenge in the SUGP was the high heterozygosity in the outbred animal that was sequenced. It was known that single-copy DNA in the sea urchin varied by as much as 4 to 5% [single nucleotide polymorphism (SNP) plus insertion/deletion (indel)], which is much greater than human (0.5%) (3). Moreover, alignment of WGS reads to the early v 0.1 WGS assembly revealed at least one SNP per 100 bases, as well as a comparable frequency of indel variants. This average frequency of a mismatch per 50 bases or higher prevented merging by the assembly module in Atlas, the Phrap assembler, and also made it difficult to know if reads were from duplicated but diverged sections of the genome or heterozygous homologs. This challenge was met by adding components to Atlas to handle local regions of heterozygosity and to take advantage of the BAC data, because each BAC sequence represented a single haplotype (see SOM). High heterozygosity has been seen in the past with the Ciona genomes (4, 5) and is likely to be the norm in the future as fewer inbred organisms are sequenced. Moreover, the CAPSS approach makes BAC sequencing more manageable for large genomes. Thus, the sea urchin project may serve as a paradigm for future difficult endeavors.

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Species, total number (percentage of total matches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPR0001190</td>
<td>Speract/scavenger receptor</td>
<td>Sp: 361 (1.79), Mm: 14 (0.08), Ci: 1 (0.01), Dm: 2 (0.02), Ce: 0 (0.00)</td>
</tr>
<tr>
<td>IPR000157</td>
<td>TIR</td>
<td>Sp: 248 (1.23), Mm: 22 (0.12), Ci: 9 (0.09), Dm: 9 (0.09), Ce: 2 (0.02)</td>
</tr>
<tr>
<td>IPR011029</td>
<td>DEATH-like</td>
<td>Sp: 172 (0.85), Mm: 8 (0.05), Ci: 19 (0.18), Dm: 5 (0.05), Ce: 1 (0.01)</td>
</tr>
<tr>
<td>IPR007111</td>
<td>NACHT nuclease triphosphatase</td>
<td>Sp: 135 (0.67), Mm: 16 (0.09), Ci: 28 (0.27), Dm: 0 (0.00), Ce: 0 (0.00)</td>
</tr>
<tr>
<td>IPR007104</td>
<td>Quinoprotein amine dehydrogenase, β chain–like</td>
<td>Sp: 122 (0.60), Mm: 70 (0.4), Ci: 15 (0.15), Dm: 5 (0.05), Ce: 6 (0.05)</td>
</tr>
<tr>
<td>IPR000558</td>
<td>Histone H2B</td>
<td>Sp: 110 (0.54), Mm: 14 (0.08), Ci: 2 (0.02), Dm: 1 (0.01), Ce: 17 (0.13)</td>
</tr>
<tr>
<td>IPR001951</td>
<td>Histone H4</td>
<td>Sp: 93 (0.46), Mm: 7 (0.04), Ci: 0 (0.00), Dm: 2 (0.02), Ce: 16 (0.12)</td>
</tr>
<tr>
<td>IPR002119</td>
<td>Histone H2A</td>
<td>Sp: 87 (0.43), Mm: 24 (0.14), Ci: 2 (0.02), Dm: 5 (0.05), Ce: 19 (0.14)</td>
</tr>
<tr>
<td>IPR008042</td>
<td>Retrotransposon, Pao</td>
<td>Sp: 76 (0.38), Mm: 0 (0.00), Ci: 0 (0.00), Dm: 0 (0.00), Ce: 6 (0.05)</td>
</tr>
<tr>
<td>IPR000164</td>
<td>Histone H3</td>
<td>Sp: 72 (0.36), Mm: 17 (0.10), Ci: 5 (0.05), Dm: 4 (0.04), Ce: 22 (0.17)</td>
</tr>
</tbody>
</table>
Combining the BAC-derived sequence with the WGS sequence generated a high-quality draft with 4 to 5% redundancy and covering more than 90% of the genome, and it sequenced only about eight times coverage (table S2). The assembly size of 914 Mb is in good agreement with the previous estimate of genome size, 800 Mb ± 5% (6). The assembly is a mosaic of the two haplotypes, but it was possible to determine the phase of the BACs on the basis of how many mismatches neighboring BACs had in their overlap regions. This information will be used to create a future version of the genome in which the individual haplotypes are resolved.

**Gene predictions.** The v 0.5 WGS assembly displayed sufficient sequence continuity (a contig N50 of 9.1 kb) and higher-order organization (a scaffold N50 of 65.6 kb) to allow gene predictions to be produced and the annotation process to begin even while the BAC component was being sequenced. We generated an official gene set (OGS), consisting of ~28,900 gene models, by merging four different sets of gene predictions with the GLEAN program (7) (see SOM for details). One of these gene sets, produced from the Ensembl gene prediction software, was created for both v 0.5 and v 2.0 assemblies.

To estimate the number of genes in the *S. purpuratus* genome, we started with the 28,900 gene models in the OGS and reduced this by the 5% redundancy found by mapping to the v 2.0 assembly, then increased it by a few percent for the new genes observed in the Ensembl set from the v 2.0 assembly compared with v 0.5. From manual analysis of well-characterized gene sets (e.g., ciliary, cell cycle control, and RNA metabolism genes), we estimated that, in addition to redundancy, another 25% of the genes in the OGS were fragments, pseudogenes, or otherwise not valid. Finally, microarray analysis (see below) showed 10% of the transcriptionally active regions (long open reading frames, not small RNAs) were not represented by genes in the OGS. Taken together, this analysis gave an estimate of about 23,300 genes for *S. purpuratus*. Information on all annotated genes can be found at (8).

The overall trends in gene structure were similar to that seen in the human genome. The statistics of the Ensembl predictions from the WGS assembly revealed an average of 8.3 exons and 7.3 introns per transcript (see SOM). The average gene length was 7.7 kb with an average primary transcript length of 8.9 kb. A broad distribution of all exon lengths peaked at around 100 to 115 nucleotides, whereas that for introns at around 750 nucleotides. The smaller average intron size relative to humans’ was consistent with the idea that intron size is correlated with genome size.

**Annotation process.** Manual annotation and analysis of the OGS was performed by a group of over 200 international volunteers, primarily from the sea urchin research community. To facilitate and to centralize the annotation efforts, an annotation database and a shared Web browser, Gencode (9), were established at the BCM-HGSC. These tools enabled integrated and collaborative analysis of both precomputed and experimental information (see SOM). A variety of precomputed information for each predicted gene model was made available in the database together with supporting information made available to the annotators in the browser, such as expressed sequence tag (EST) data, the four unmerged gene prediction sets, and

---

### Table 2. Distribution among sequenced animal genomes of various Pfam domains associated with selected aspects of eukaryotic cell physiology.

<table>
<thead>
<tr>
<th>Process</th>
<th>Domain</th>
<th>Pfam no.</th>
<th>Sp</th>
<th>Hs</th>
<th>Dm</th>
<th>Ce</th>
<th>Nv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell cycle control</td>
<td>Cyclin_N</td>
<td>PF00134</td>
<td>15 (17)</td>
<td>21</td>
<td>11</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF02984</td>
<td>7 (8)</td>
<td>12</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF02931</td>
<td>3 (5)</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01858</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01857</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00870</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00887</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01466</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>Histone metabolism</td>
<td>Histone*</td>
<td>PF00125</td>
<td>49</td>
<td>75</td>
<td>8</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00538</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF03066</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00956</td>
<td>2</td>
<td>24</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00850</td>
<td>8</td>
<td>11</td>
<td>5</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF08123</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RNA metabolism</td>
<td>RRM_1</td>
<td>PF00767</td>
<td>140 (178)</td>
<td>245</td>
<td>126</td>
<td>99</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00567</td>
<td>15</td>
<td>13</td>
<td>15</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00270</td>
<td>73 (125)</td>
<td>78</td>
<td>56</td>
<td>65</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01423</td>
<td>17</td>
<td>21</td>
<td>17</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00133</td>
<td>28 (31)</td>
<td>36</td>
<td>28</td>
<td>28</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00035</td>
<td>14 (15)</td>
<td>21</td>
<td>14</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01612</td>
<td>13 (15)</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00929</td>
<td>9</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Apoptosis†</td>
<td>Caspase</td>
<td>PF00653</td>
<td>31 (33)</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00452</td>
<td>10</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00020</td>
<td>8 (9)</td>
<td>8</td>
<td>1 (no DD)</td>
<td>1 (no DD)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF05729</td>
<td>129 (145)</td>
<td>18</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00931</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF05311</td>
<td>47 (101)</td>
<td>30</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF01335</td>
<td>4 (5)</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PF00619</td>
<td>5 (10)</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

---

*Numbers of histone genes refer to distinct core or linker histone genes, as opposed to total gene number as a result of large tandemly repeated arrays (e.g., ~400 clusters of early histone arrays in sea urchin, 100 copies of a tandem array in Drosophila, with each array containing a gene for the four core and one H1 histone). †Numbers for Hs, Dm, and Ce obtained from (52).}
transcription data from whole-genome tiling microarray with embryonic RNA (see below) (10). Additional resources available to the community are listed in table S4.

Over 9000 gene models were manually curated by the consortium with 159 novel models (gene models not represented in the OGS) added to the official set. If we assume no bias in the curated gene models, the number of novel models added may imply that the official set contains >98% of the protein-coding genes.

**Genomes features.** A window on the genetic landscape is scaffold-centric in S. purpuratus, because linkage and cytogenetic maps are not available. The 36.9% GC content of the genome is uniformly low because assessment of the average GC content by domains is consistent (36.8%), and the distribution is tight (see SOM). Genes from the OGS show no tendency to occupy regions of higher- or lower-than-average GC content. In fact, nearly all genes lie in regions of 35 to 39% GC.

**The Echinoderm Genome in the Context of Metazoan Evolution**

The sea urchin genetic tool kit lends evolutionary perspective to the gene catalogs that characterize the superclades of the bilaterian animals. The distribution of highly conserved protein domains and sequence motifs provides a view of the expansion and contraction of gene families, as well as an insight into changes in protein function. Examples are enumerated in Table 1, which presents a global overview of gene variety obtained by comparing sequences identified in Interpro, and Table 2, which shows the distribution of specific Pfam database domains associated with selected aspects of cell physiology, including sequences identified in the cnidianan Nematoscelis vectensis (11). The Interpro data suggest that about one-third of the 50 most prevalent domains in the sea urchin gene models are not in the 50 most abundant families in the other representative genomes (mouse, tunicate, fruit fly, and nematode), and thus, they constitute expansions that are specific at least to sea urchins, if not to the complex of echinoderms and hemichordates. Two of the most abundant domains make up 3% of the total and mark genes that are involved in the innate immune response. Others define proteins associated with apoptosis and cell death regulation, as well as proteins that serve as downstream effectors in the Toll-interleukin 1 (IL-1) receptor (TIR) cascade. The quinoprotein amine dehydrogenase domain seen in the sea urchin set is 10 times as abundant as in other representative genomes and may be used in the systems of quinone-containing pigments known to occur in these marine animals. The large number of nucclesomal histone domains found agrees with the long-established sea urchin-specific expansion of histone genes. In summary, the distribution of proteins among these conserved families shows the trend of expansion and shrinkage of the preexisting protein families, rather than frequent gene innovation or loss. Gene family sizes in the sea urchin are more closely correlated with what is seen in deuterostomes than what is seen in the protostomes.

Of equal interest are the sorts of proteins not found in sea urchins. The sea urchin gene set shares with other bilaterian gene models about 4000 domains, whereas 1375 domains from other bilaterian genomes are not found in the sea urchin set. In agreement with the lack of morphological evidence of gap junctions in sea urchins, there are no gap junction proteins (connexins, pannexins, or innexins) Also missing are several protein domains unique to insects, such as insect cuticle protein, chitin-binding protein, and several pheromone– or odorant-binding proteins, as well as a vertebrate invention—the Krüppel-associated box or KRAB domain, a repressor domain in zinc finger transcription factors (12).

Finally, searches for specific subfamilies of G protein–coupled receptors (GPCRs) that are known as chemosensory and/or odorant receptors in distinct bilaterian phyla failed to detect clear representatives in the sea urchin genome. However, this failure more likely reflects the fast evolution of these receptors, rather than a lack of chemoreceptor molecules, because the sea urchin genome encodes close to 900 GPCRs of the same superfamilies (rhodopsin-type GPCRs), several of which are expressed in sensory structures (13). A conservative way to compare gene sets is to count the strict orthologs that give reciprocal BLAST matches. Genes that are genuine orthologs are likely to yield each other as a best hit. Comparison of sea urchin, fruit fly, nematode, ascidian, mouse, and human gene sets (Fig. 2) indicates that the greatest number of reciprocal best matches is observed between mouse and human, which reflects their close relation. The numbers of presumed orthologous genes between the ascidian and the two mammals are about equal, but are less than the number counted between these species and the sea urchin. The difference is consistent with the lower gene number and reduced genome size in the urochordates (4).

The number of reciprocal pairs for sea urchin and mouse is about 1.5 times the matches between proteins in sea urchin and fruit fly. The number of nematode proteins matching either sea urchin or fruit fly is even lower. This is likely the result of the more rapid sequence changes in the nematode compared with the other species used in this analysis. More than 75% of the genes that are shared by sea urchin and fruit fly are also shared between sea urchin and mouse. Thus, these genes constitute a set of genes common to the bilaterians, whereas the additional sea urchin–mouse pairs are unique to the deuterostomes.

The sea urchin genome consequently provides evidence for the now extremely robust concept of the deuterostome superclade. A 1908 concept that originated in the form of embryos of dissimilar species (14) is demonstrated by genomic comparisons.

**Developmental genomics.** In the 1980s, the sea urchin embryo became the focus of cis-regulatory analyses of embryonic gene expression, and there was a great expansion of molecular explorations of the developmental cell biology, signaling interactions, and regulatory control systems of the embryo. Analysis of the entire genome facilitated the first large-scale correlation of the gene regulatory network for development, which represents the genomic control circuitry for specification of the endoderm and mesoderm of this embryo (15–17) with the encoded potential of the sea urchin.

The embryo transcriptome and regulome. As noted earlier, embryogenesis is cleanly separated from adult body plan formation, in developmental process and in time, and therefore, it is possible to estimate the genetic repertoire specifically required for formation of a simple embryo (11). Pooled mRNA preparations from four stages of development, up to the mid-late gastrula stage (48 hours), were hybridized with a whole-genome tiling array. Expression of about 12,000 to 13,000 genes was seen during this early period, indicating that ~52% of the entire protein-coding capacity of the sea urchin genome is expressed during development to the mid-late gastrula stage. An additional set of microarray experiments extended the interrogation of embryonic expression to the 3-day pluteus larva stage (see SOM) (18). The DNA binding domains of transcription factor families are conserved across the Bilateria, and these protein domain motifs were used to extract the sea urchin homologs (see SOM). For each identified gene, if data were not already available, probes were built from the genome sequence and used to measure transcript concentration by quantitative polymerase chain reaction with a time series of embryo mRNAs, as well as to determine spatial expression by whole-mount in situ hybridization.
All bilaterian transcription factor families were represented in the sea urchin with a few rare exceptions (see below), so the sea urchin data strongly substantiate the concept of a pan-bilaterian regulatory tool kit (19) or “regulome.” We found that 80% of the whole sea urchin regulome (except the zinc finger genes) was expressed by 48 hours of embryogenesis (20), an even greater genetic investment than the 52% total gene use in the same embryo.

Signal transduction pathways. More than 1200 genes involved in signal transduction were identified. Comparative analysis highlights include the protein kinases that mediate the majority of signaling and coordination of complex pathways in eukaryotes. The S. purpuratus genome has 353 protein kinases, intermediate between the core vertebrate set of 510 and the fruit fly and nematode conserved sets of ~230. Fine-scale classification and comparison with annotated kinomes (21, 22) reveals a remarkable parsimony. Indeed, with only 68% of the total number of human kinases, the sea urchin has members of 97% of the human kinase subfamilies, lacking just four of those subfamilies (Axl, FastK, H11, and NKF3), whereas Drosophila lacks 20 and nematodes 32 (Fig. 3) (23). Most sea urchin kinase subfamilies have just a single member, although many are expanded in vertebrates; thus, the sea urchin kinase is largely nonredundant. A small number of kinases were more similar to insect than to vertebrate homologs (including the Titin homolog projection, the Syk-like tyrosine kinase Shark, and several guanylate cyclases), which indicated for the first time the loss of kinase classes in vertebrates (23). Expression profiling showed that 87% of the signaling kinases and 80% of the 91 phosphatases were expressed in the embryo (23, 24), which emphasized the importance of signaling pathways in embryonic development.

The small guanosine triphosphatases (GTPases) function as molecular switches in signal transduction, nuclear import and export, lipid metabolism, and vesicle docking. Vertebrate GTPase families were expanded after their divergence from echinoderms, in part by whole-genome duplications (25–27). The sea urchin genome did not undergo a whole-genome duplication, yet phylogenies for four Ras GTPase families (Ras, Rho, Rab, and Arf) revealed that local gene duplications occurred (Fig. 4), which ultimately resulted in a comparable number of monomeric GTPases in the human and sea urchin genomes (28). Thus, expansion of each family in vertebrates and echinoderms was achieved by distinct mechanisms (gene-specific versus whole genome duplication). More than 90% of the small GTPases are expressed during sea urchin embryogenesis, which suggests that the complexity of signaling through GTPases is comparable between sea urchins and vertebrates.

The Wnt family of secreted signaling molecules plays a central role in specification and patterning during embryonic development. Phylogenetic analyses from cnidarian to human indicate that of the 13 known Wnt subfamilies, S. purpuratus has 11, missing Wnt2 and Wnt11 homologs (Fig. 5). S. purpuratus has WntA, previously reported as being absent from deuterostomes (29). Of 126 genes described as components of the Wnt signal transduction machinery, homologs of ~90% were present in the sea urchin genome, which indicated a high level of conservation of all three Wnt pathways (30). However, of 94 Wnt transcriptional target genes reported in the literature, mostly from vertebrates (31), only 53% were found with high confidence in the sea urchin genome (Fig. 6).
absent Wnt targets include vertebrate adhesion molecules, which were frequently missing from the sea urchin genome (32), as well as signaling receptors, which are more divergent and thus more difficult to identify. In contrast, most transcription factor targets of the Wnt pathway are present in the genome, which reflects a higher degree of conservation of transcription factor families (20). Taken together, the genomic analysis of signal transduction components indicates that sea urchins have signaling machinery strikingly comparable to that of vertebrates, often without the complexity that arises from genetic redundancy.

Sea Urchin Biology
Analysis of the genome allows understanding of parts of the organism that have not been well studied. Several examples of this follow with further details in the SOM. Additional areas such as intermediary metabolism, metalloproteases, ciliary structure, fertilization, and germ-line specification are presented in the SOM.

Fig. 5. Survey of the Wnt family of secreted signaling molecules in selected metazoans. Each square indicates a single Wnt gene identified either through genome analyses or independent studies, and squares with a question mark underline uncertainty of the orthology. Letter X’s represent absence of members of that subfamily in the corresponding annotated genome; empty spaces have been left for species for which genomic databases are not yet available. [From (30)]

Fig. 6. Presence of Wnt signaling machinery components (A) and target genes (B) in the S. purpuratus genome. (A) The 126 genes involved in the transduction of the Wnt signals have been separated into four categories from the extracellular compartment to the nucleus. Sea urchin homologs are identified by the lighter shade (indicated by both the number and the percentage of homologs that were identified within the chart); the total number of known genes is indicated in the chart legend. (B) The 93 reported Wnt targets have been divided into three categories: signaling molecules, transcription factors, and cell adhesion molecules. Colors and numbers are as in (A).

The complement system. The complement system of vertebrates is a complex array of soluble serum proteins and cellular receptors arranged into three activation pathways (classical, lectin, and alternative) that converge and activate the terminal or lytic pathway. This system opsonizes pathogenic cells for phagocytosis.
Sea Urchin Genome

and sometimes activates the terminal pathway, which leads to pathogen destruction. An invertebrate complement system was first identified in the sea urchin [for reviews, see (39, 40)], and the analysis of the genome sequence presented a more complete picture of this important immune effector system. In chordates, collectins initiate the lectin cascade through members of the mannose-binding protein (MBP)–associated protease (MASP)/C1r/C1s family. Several genes encoding collectins, C1q and MBP, have been predicted (39) and were present in the genome; however, members of the MASP/C1r/C1s family were not identified. There was no evidence for the classical pathway, which links the complement cascade with immunoglobulin recognition in jawed vertebrates. The alternative pathway is linked with immunoglobulin recognition and sometimes activates the terminal pathway, the major activity of this system is expected to be opsonization.

Homologs of immune regulatory proteins. Cytokines are key regulators of intercellular communication involving immune cells, acting to coordinate vertebrate immune systems. Genes encoding cytokines and their receptors often evolve at a rapid pace, and most families are known only from vertebrate systems. Although members of many cytokine, chemokine, and receptor families were not identified in the sea urchin genome, a number of important immune signaling homologs were present. These included members of the tumor necrosis factor (TNF) ligand and receptor superfamilies, an IL-1 receptor and accessory proteins, two IL-17 receptor-like genes and 30 IL-17 family ligands, and nine macrophage inhibitory factor (MIF)-like genes. Receptor tyrosine kinases (RTKs) included those that bind important growth factors that regulate cell proliferation in vertebrate hematopoietic systems. Of particular note, from the sea urchin genome, two vascular endothelial growth factor (VEGF) receptor–like genes and a Tie1/2 receptor, all of which were expressed in adult coelomocytes. Many of these genes are homologs of important inflammatory regulators and growth factors in higher vertebrates, and these sea urchin homologs may have similar functions in regulating coelomocyte differentiation and recruitment.

Representatives of nearly all subclasses of important vertebrate hematopoietic and immune transcription factors were present in the sea urchin genome. Notably, the genome contained homologs of immune transcription factors that had not been identified previously outside of chordates, including PU.1/SpiB/SpiC, a member of the Ets subfamily, and a zinc finger gene with similarity to the Ikaros subfamily. Transcript prevalence measurements showed that PU.1, the Ikaros-like gene and homologs of Gata1/2/3, E2A/HEB/ITF2, and Scarecrow-like (SCL) were all expressed at substantial levels in coelomocytes (41). This was consistent with the presence of conserved mechanisms of regulating gene expression among sea urchin coelomocytes and vertebrate blood cells.

ABC transporters. Many chemicals are kept out of cells by efflux proteins known as ATP-binding cassette (ABC) or multidrug efflux transporters. *S. purpuratus* has 65 ABC transporters genes in the eight major subfamilies of these genes (ABC A to H; (42)]. The ABC family of multidrug transporters is about 25% larger than in other deuterostome genomes with at least 30 genes in this family (nearly half of the sea urchin ABC transporters), and 25 of these 30 genes showed substantial mRNA expression in eggs, embryos, or larvae. Much of the expansion is in the Sp-ABC5 and Sp-ABC9 families, whereas orthologs of the vertebrate gene ABC2C (also called MRP2) are absent. Because the ABCB family is known to generally transport more hydrophilic compounds than other transporter families, such as the ABCC genes, sea urchins may have increased need for transport of these compounds. ABCB efflux activity has been described in sea urchin embryos and, consistent with the genomic expansion of the ABCB family, the major activity in early embryos ensues from an ABCB-like efflux mechanism.

Flavoprotein monooxygenase (FMO) and cytochrome P-450 (CYP). Enzymes in the CYP1, CYP2, CYP3, and CYP4 families carry out oxidative biotransformation of chemicals to more hydrophilic products. The sea urchin has 120 CYP genes, and those related to CYP gene families 1 to 4 constitute 80% of the total, which suggests that there has been selective pressure to expand functionality in these gene families (42). Eleven CYP1-like genes are present in the sea urchin genome, more than twice the number in chordates. CYP2-like and CYP3-like genes are also present at greater numbers than in other deuterostomes. In addition to the CYPs in families 1 to 4, the sea urchin genome contains homologs of proteins involved in developmental patterning (CYP26), cholesterol synthesis (CYP51), and metabolism (CYP27, CYP46). Homologs of some CYPs with endogenous functions in vertebrates were missing; however, (CYP19, androgen aromatase; CYP8, prostacyclin synthase; CYP11, pregnenolone synthase; CYP7, cholesterol-7a-hydroxylase). These CYP genes in concert with additional expanded defensive gene families represent a large diversification of these gene families by the sea urchin relative to mammals (42). Oxidative defense and metal-complexing proteins. The metal-complexing proteins include three metallothionein genes and three homologs of phytochelatin synthase genes. Genes for antioxidant proteins include three superoxide dismutase (SOD) genes and a gene encoding ovoperoxidase (an unusual peroxidase with SOD-like activity), along with one catalase, four glutathione peroxidase, and at least three thioredoxin peroxidase genes. Reactive oxygen detoxification genes may be especially important in conferring the long life-span of sea urchins, because oxidative damage is thought to be a major factor in aging.

Diversity and conservation in xenobiotic signaling. The diversity of genes encoding xenobiotic-sensing transcription factors that
Table 3. Genomic insights into sea urchin neurobiology.

<table>
<thead>
<tr>
<th>Neural process</th>
<th>Revelations from the genome</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural development</td>
<td>Neurogenic ectoderm is specified in early embryonic development.</td>
<td>Sp-Achaete-scute, Sp-homeobrain, Sp-Rx (retinal anterior homeobox), Sp-Zic2</td>
</tr>
<tr>
<td>Synapse structure and function</td>
<td>Echinoderm synapses are structurally unusual, despite the presence of many genes encoding proteins involved in synapse function.</td>
<td>Sp-Neurilognin, Sp-neurexin, Sp-agtin, Sp-MUSK, Sp-thrombospondin, Sp-Rim2, Sp-Rab3, exocyst complex, Snares, SM, synaptotagmins</td>
</tr>
<tr>
<td>Electrical signaling and coupling</td>
<td>Neurons have ion channel proteins, but lack electrical coupling via gap junctions</td>
<td>Voltage-gated K⁺, Ca⁺⁺, and Na⁺ channels, but no connexins or pannexins/innexins</td>
</tr>
<tr>
<td>Neurotransmitter/neuromodulatory diversity</td>
<td>Neurons use the same neurotransmitters as vertebrates, but lack melatonin and adrenaline</td>
<td>Enzymes involved in synthesis, transport, reception, and hydrolysis of serotonin, dopamine, noradrenaline, γ-aminobutyric acid (GABA), histamine, acetylcholine, glycine, and nitric oxide</td>
</tr>
<tr>
<td>GPCR signaling</td>
<td>Identification of GPCRs that are unique to chordates and identification of expanded GPCR families</td>
<td>Orthologs of vertebrate cannabinoid, lysophospholipid, and melanocortin receptors are absent; 162 secretin-receptor-like genes</td>
</tr>
<tr>
<td>Peptide signaling</td>
<td>G-protein coupled peptide receptors indicate diversity in peptide signaling systems, but only a few sea urchin neuropeptides or peptide hormones identified</td>
<td>37 G protein-coupled peptide receptors. Precursors for SALMfamides, NGFFlamide, and a vasotocin-like peptide</td>
</tr>
<tr>
<td>Neurotrophs</td>
<td>Neurotrophin and neurotrophin receptors are not unique to chordates.</td>
<td>Sp-Neurotrophin, Sp-Trk, Sp-p75NTR, ependymins</td>
</tr>
<tr>
<td>Insulin and IGFs</td>
<td>More similar to vertebrate forms than invertebrate insulin-like molecules</td>
<td>Sp-IGF1, SpIGF2</td>
</tr>
<tr>
<td>Chemosensory functions</td>
<td>A large family of predicted chemoreceptor genes, some expressed in tube feet or pedicellaria, indicates a complex chemosensory system.</td>
<td>Over 600 genes encoding putative G protein-coupled chemoreceptors, many tandemly repeated and lacking introns</td>
</tr>
<tr>
<td>Photoreception functions</td>
<td>Genes associated with photoreception are expressed in tube feet.</td>
<td>Phototrohodopsins, Sp-Pax6, retinal transcription factors</td>
</tr>
<tr>
<td>Mechanosensory functions</td>
<td>Orthologs of vertebrate mechanosensory genes are present.</td>
<td>Sp-Usherin, Sp-VLGR-1, Sp-cadherins, Sp-myosin 7, Sp-myosin 15, Sp-harmonin, Sp-whirlin, Sp-NBC, Sp-TrpA1</td>
</tr>
</tbody>
</table>

regulate biotransformation enzymes and transporters was similar to other invertebrate genomes, but in most cases lower than vertebrates. For example, the sea urchin genome encoded a single predicted CNC-bZIP protein homologous to the four human CNC-bZIP proteins involved in the response to oxidative stress. There were two sea urchin homologs of the aryl hydrocarbon receptor (AHR), which in vertebrates mediates the transcriptional response to polynuclear and halogenated aromatic hydrocarbons and, in both protostomes and deuterostomes, also regulates specific developmental processes (43–45). One of the sea urchin AHR homologs was more closely related to the vertebrate AHR; the other shared greatest sequence identity with the Drasophila AHR homolog spineless. Sea urchins also had two genes encoding hypoxia-inducible factors (HIFα subunits), which regulate adaptive responses to hypoxia, and a gene encoding ARNT, a PAS protein that is a dimerization partner for both AHRs and HIFs.

Stronglylocentrotus purpuratus has 32 nuclear receptor (NR) genes (20), two-thirds the number in humans, including several with potential roles in chemical defense (42). The sea urchin genome also contains two peroxisome proliferator-activated receptor (PPAR, NR1C) homologs and an NR1H gene coorthologous to both liver X receptor (LXR) and farnesoid X receptor (FXR) (42). Genes homologous to the vertebrate xenobiotic sensor NR1H genes [pregnan X receptor, PXR; constitutive androstane receptor, CAR (46)] are absent, although three NR1H-related genes were found, which possibly form a new subfamily of genes involved in xenobiotic sensing. Many of these defense genes are expressed during development (11, 42), which suggests that they have dual roles in chemical defense and in developmental signaling. In several cases (CYPs, AHR, NF-E2), the evolution of pathways for chemical defense may have involved recruitment from developmental signaling pathways (42).

Nervous system. The echinoderm nervous system is the least well studied of all the major metazoan phyla. For a number of technical reasons, the structure and function of echinoderm nerves have been neglected. Analysis of the sea urchin genome has enabled an unprecedented glimpse into the neural and sensory functions and has revealed several novel molecular approaches to the study of echinoderm nervous systems (Table 3).

The nervous systems of echinoderm larvae and adults are dispersed, but they are not simple nerve nets, a feature that distinguishes them from other deuterostomes (47). Adult sea urchins have thousands of appendages, each with sensory neurons, ganglia, and motor neurons arranged in local reflex arcs. These peripheral appendages are connected to each other and to radial nerves, which provide overall control and coordination (47, 48).

Nearly all of the genes encoding known neurogenic transcription factors are present in the sea urchin genome, and several are expressed in neurogenic domains before gastrulation, which indicates that they may operate near the top of a conserved neural gene regulatory network (47). Axon guidance molecules known from other metazoans are also expressed in the
Sea Urchin Genome

developing embryo. Unexpectedly, genes encoding the neurotrophin-Trk receptor system are present in sea urchin that were thought to be vertebrate-specific because they were not found in Ciona, which suggests a deuterostome origin and a potential loss in urochordates.

The genes required to construct neurons and to transmit signals are present, but the repertoire of neural genes and the initial characterization of expression of a number of them led to unexpected and surprising conclusions. There appear to be no genes encoding gap junction proteins, which suggests that communication among neurons depends on chemical synapses without tonic coupling. Structurally, the synapses of echinoderms are unusual because there are no direct synaptic contacts (49). The repertoire of sea urchin neurotransmitters is large, but melanin and adrenalin are lacking, as they are in ascidians (4, 47). Cannabinoid, lysophospholipid, and melanocortin receptors are not present in urchins, but orthologs were found in ascidians (4, 47). In contrast, some sets of genes thought to be chordate-specific have sea urchin orthologs, for example, insulin and insulin-like growth factors (IGFs) that are more similar to their chordate counterparts than those of other invertebrates (47). Overall, the genome contains representatives of all five large superfamilies of GPCRs, including those that mediate signals from neuropeptides and peptide hormones. Both the secretin and rhodopsin superfamilies display marked lineage-specific expansions (14, 47).

Sensory systems. There were 200 to 700 putative chemosensory genes that formed large clusters and lacked introns, which are features of chemosensory genes in vertebrates, but not in Ciona. Many of these genes encoded amino acid motifs that were characteristic of vertebrate chemosensory and odorant receptors (14, 47). Sea urchins had an elaborate collection of photoreceptor genes that quite surprisingly appeared to be expressed in tube feet (14, 47). These included many genes encoding transcription factors regulating retinal development and a photorhodopsin gene.

Human Usher syndromes are genetic diseases affecting hearing, balance, and retinitis pigmentosa (retinal photoreceptor degeneration). Most of the genes involved have been identified, and they encode a set of membrane and cytoskeletal proteins that form an interacting network that controls the arrangement of mechanosensory stereocilia in hair cells of the mammalian ear. Many or all of the proteins play some roles in photoreceptor organization and/or maintenance. Orthologs of virtually the entire set of membrane and cytoskeletal proteins of the Usher syndrome network were found in the sea urchin genome. These include the very large membrane proteins, usherin and VLGR-1 and large cadherins (Cadh23 and possibly Pcad15), all of which participate in forming links between stereocilia in mammalian hair cells, as well as myosin 7 and 15, two PDZ proteins (harminon and whirlin) and another adaptor protein (SANS), which participate in linking these membrane proteins to the cytoskeleton. In addition, two membrane transporters, NBC (a candidate Usher syndrome target known to interact with harmonin) and TrpA1 (the mechanosensory channel connected to the tip links containing cadherin 23), have orthologs in the sea urchin genome. Sea urchins do not have ears or eyes, so they must deploy these proteins in other sensory processes. Sea urchins respond to light, touch, and displacement and probably use some of same sensory genes used by vertebrates.

The echinoderm adhesome. The *S. purpuratus* genome contained representatives of all the standard metazoan adhesion receptors (table S7), but the emphasis on different classes of receptors differed substantially from that used by vertebrates. The integrin family was intermediate in size between those of protostomes and vertebrates—several chordate-specific expansions of the integrin repertoire were absent, and there were some expansions (so far) unique to echinoderms. The cadherin repertoire was also small relative to vertebrates (a dozen or so instead of over a hundred), and many chordate-specific expansions were missing. Specialized large cadherins shared by protostomes and vertebrates were present, as well as some specialized large cadherins previously thought to be chordate-specific, but overall, the cadherin repertoire was more invertebrate than vertebrate in character. Sea urchins lacked the integrins and cadherins that link to intermediate filaments in vertebrates.

In contrast, sea urchins had large repertoires of adhesion molecules containing immunoglobulin superfamily, fibronectin type 3 repeat (FN3), epidermal growth factor (EGF), and LRR repeats. In addition to the expansion of TLRs and NLRs mentioned above, there are large expansions of other LRR receptor families, including GPCRs (32). The key neural adhesion systems involved in regulating axonal outgrowth were present (netrin/Unc5/DCC; Slt/Robo; and semaphorins and/or plexins), as were adhesion molecules involved in synaptogenesis (Agrin and/or MUSK; and neuroligins and/or neuroligins). This was not surprising because these molecules were known in both protostomes and vertebrates. However, some of them were expressed in sea urchin embryos before there are any neurons, suggesting that they may have other roles as well.

The basic metazoan basement membrane extracellular matrix (ECM) tool kit was present—two alpha-IV collagen genes, perlecán, laminin subunits, nidogen, and collagen XV/XVIII. There did not appear to be much, if any, expansion of these gene families, as is found in vertebrates, which suggests that there is less diversity among basement membranes. Quite a few ECM proteins present in chordates, but not protostomes, were also missing in sea urchins, including fibronectins, tenascins, von Willebrand factor, vitronectin, most vertebrate-type matrix proteoglycans, and complex VWA/FN3 collagens among others (32). Absence of these genes may be related to the absences of neural crest migration, a high shear endothelial-lined vasculature and, of course, cartilage and bone.

In addition to the components of Usher syndromes mentioned above, it was surprising to find a clear ortholog of reelin, a large ECM protein involved in establishing the layered organization of neurons in the vertebrate cerebral cortex. Reelin is mutated in the *reeler* mouse, and mutations in the *reeler* gene in humans have been associated with Norman-Roberts-type lissencephaly syndrome. Reelin has a unique domain composition and organization (Reeler, EGF, BNR) that has not been found outside chordates, but the sea urchin genome included a very good homolog of reelin. Receptors for reelin are believed to include low-density lipoprotein receptor-related proteins (LRPs), and there are a number of these receptors in *S. purpuratus* although it is as yet unclear whether they are reelin receptors, lipoprotein receptors, or something else. Similar receptors are also involved in human disease (atherosclerosis).

**Biomineralization genes.** Among the deuterostomes, only echinoderms and vertebrates produce extensive skeletons. The possible evolutionary relations between biomineralization processes in these two groups have been controversial. Analysis of the *S. purpuratus* genome revealed major differences in the proteins that mediate biomineralization in echinoderms and vertebrates (50). First, there were few sea urchin counterparts of extracellular proteins that mediate biomolecular deposition in vertebrates. For example, in vertebrates, an important class of proteins involved in biomineralization is the family of secreted, calcium-binding phosphoproteins, or SCPPs. Sea urchins did not have counterparts of SCPP genes, which supports the hypothesis that this family arose via a series of gene duplications after the echinoderm-chordate divergence (51). Second, almost all of the proteins that have been directly implicated in the control of biomineralization in sea urchins were specific to that clade. The echinoderm skeleton consists of magnesium calcite (as distinct from the calcium phosphate skeletons of vertebrates) in which is occluded many secreted matrix proteins. The sea urchin spicule matrix proteins were encoded by a family of 16 genes that are organized in small clusters and likely are proliferated by gene duplication. Counterparts of sea urchin spicule matrix genes were not found in vertebrates, amphibious, or ascidians. Likewise, other genes that have been implicated in biomineralization in sea urchins, including genes that encode the transmembrane protein P16 and MSP130, a glycosylphosphatidylinositol-linked
glycoprotein, were members of small clusters of closely related genes without apparent homologs in other deuterostomes. The members of all three of these sea urchin–specific gene families were expressed specifically by the biomecoral-forming cells of the embryo, the primary mesoblast-chyme cells (see [50]). As a whole, these findings highlighted substantial differences in the primary sequences of the proteins that mediate biomineralization in echinoderms and vertebrates.

Cytoskeletal genes. In addition to identifying gene families for all previously known S. purpuratus actins and tubulins, 18- and 20-tubulin genes were found [52]. Newly identified motor protein gene families included members of four more classes of myosin, and eight more families of kinesins. The first dynein cloned and sequenced was from sea urchin, and although most S. purpuratus dynein heavy chain genes mapped one-to-one to mammalian homologs, Sp-DNAH9 mapped one-to-three, as it was equidistant between the closely similar mammalian genes DNAH9, DNA11, and DNAH17 [53].

Conclusions

Our estimate of 23,000 genes is similar to estimates for vertebrates, despite the fact that two whole-genome duplications are believed to have occurred in the chordate lineage after divergence from echinoderms (25–27). From the analysis presented here, it seems likely that many mechanisms shaped the final genetic content of these genomes. On the one hand, there are cases of gene families that are expanded in vertebrates compared with sea urchin, including examples of the expected 4:1 ratio from two duplications [16]. However other patterns are also found.

The nuclear receptor family is only slightly reduced in sea urchin compared with that of humans, which suggests gene loss followed the vertebrate duplications. The unprecedented expansions of innate immune system diversity contrast sharply with the much smaller sets of counterparts that are present in the sequenced genomes of protostomes, Ciona, and vertebrates, an example of independent expansion in the sea urchin, whereas the GTPases described here have expanded in sea urchin to about the same numbers as in vertebrates. Thus, whereas the duplications of the chordate lineage were a contributor to the increased complexity of vertebrates, regional expansions clearly play a large role in the evolution of these animals.

The refinement of the inventory of vertebrate-specific or protostome-specific genes likewise benefits from the sea urchin genome. Many more human genes have shared ancestry across the deuterostomes, and in fact, bilaterian genes are more broadly shared than had been inferred from comparison of the previously limited genome sequences. The new biological niche sampled by the sea urchin genome provides not only a clearer view of the deuterostome and bilaterian ancestor, but has also provided a number of surprises. The finding of sea urchin homologs for sensory proteins related to vision and hearing in humans may lead to interesting new concepts of perception, and the extraordinary organization of the sea urchin immune system is different from any animal yet studied. From a practical standpoint, the sea urchin may be a treasure trove. Because of the many pathways shared by sea urchin and human, the sea urchin genome includes a large number of human disease gene orthologs. Many of the gene families described in the preceding sections fall into this category (see tables S7 and S8) and cover a surprising diversity of systems such as nervous, endocrine, and blood systems, as well as muscle and skeleton, as exemplified by the Huntington and muscular dystrophy genes.

Continued exploration of the sea urchin immune system is expected to uncover additional variations for protection against pathogens. The immense diversity of pathogen-binding motifs encoded in the sea urchin genome provides an invaluable resource for antimicrobial applications and the identification of new deuterostome immune functions with direct relevance to human health. These exciting possibilities show that much biodiversity is yet to be uncovered by sampling additional evolutionary branches of the tree of life.

References and Notes

5. J. P. Vasten et al., Genome Res. 15, 1127 (2005).
31. The Wnt homepage (www.stanford.edu/~rnusse/).
Sea Urchin Genome

(NCI) (MIT Cancer Center core grant P30-CA14053); D. McClay, NIH grants GM61464, HD039948, and HD14483; V. D. Vacquier (group leader), G. W. Moy, H. J. Gunaratne, M. Kinukawa, M. Nomura, A. T. Neill, and Y.-H. Su, NIH grant R37-HD12896; R. D. Burke, NSERC and CIHR; L. M. Angerer, National Institute of Dental and Craniofacial Research (NIDCR), R. C. Angerer (NIDCR), Z. Wei (NIDCR), G. Humphrey, National Institute of Child Health and Human Development (NICHD), M. Landrum, National Center for Biotechnology Information (NCBI), P. Kitts (NCBI), K. Pruitt (NCBI), V. Sapojnikov (NCBI), A. Souvorov (NCBI), W. Hiavina (NCBI), S. Fugmann, National Institute on Aging (NIA), M. Dean, National Cancer Institute—Frederick (NCI/NCIFCRF) Intramural Research Program of the NIH; P. Cormier, Association pour la Recherche contre le Cancer (ARC), France, grants 4247 and 3507 to P.C., Ligue Nationale contre le Cancer to P.C., Conseil Régional de Bretagne and Conseil Général du Finistère; W. H. Klein, National Eye Institute, NIH grant EY11930, NICHD HD66219, and the Robert A. Welch Foundation (G-0010); N. Adams, NSF grant IBN 0417003 and the Department of the Navy, Office of Naval Research, under Award N00014-05-1-0855; D. Epel, NSF 0417225; A. Hamdoun, F32-HD47136; C. Byrum, American Heart Association grant 0420074Z; K. Walton, U.S. Army Medical Research and Materiel Command grant WBXWH-04-1-0324.

Supporting Online Material
www.sciencemag.org/cgi/content/full/VOL/ISSUE/PAGE/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