[1]National Marine Fisheries Service, Systematics Laboratory, National Museum of Natural History, Washington, DC 20560, USA. [2]Department of Oceanography, University of Hawaii, 1000 Pope Road, Honolulu, HI 96822, USA. [3]ECOBIOMAR, Instituto de Investigaciones Marinas (CSIC) Vigo, Spain. [4]Marine Ecosystems Research Department, Japan Marine Science and Technology Center (JAMSTEC), 2-15 Natsushima-cho, Yokosuka-shi, Kanagawa-ken, Japan. [5]Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039–0628, USA. [6]Department of Environmental Health Science, University of South Carolina, Columbia, SC 29208, USA. [7]Department of Oceanography, Texas A&M University, College Station, TX 77843–3146, USA. [8]IFREMER/CENTOB, BP 70 Plouzané, France.

*To whom correspondence should be addressed. E-mail: vecchione.michael@nmnh.si.edu

GENETICS

# Miniature Genome in the Marine Chordate *Oikopleura dioica*

**Hee-Chan Seo,[1] Michael Kube,[2]* Rolf B. Edvardsen,[1]* Marit F. Jensen,[1] Alfred Beck,[2] Endy Spriet,[1] Gabriel Gorsky,[3] Eric M. Thompson,[1] Hans Lehrach,[2] Richard Reinhardt,[2] Daniel Chourrout[1]†**

*Oikopleura dioica* is a small pelagic chordate belonging to the class of larvaceans (Fig. 1A). Chordates include vertebrates, cephalochordates, and urochordates, and urochordates are divided into ascidians, thaliaceans, and larvaceans. Larvaceans may be derived from the most basal branch of urochordates, and for this reason, they have been targeted for studies of chordate innovations, such as the tail, the notochord, or the dorsal nervous system (*1*, *2*). Found in all oceans, *Oikopleura dioica* can be permanently cultured in the laboratory. Its very short life cycle (4 days at 20°C, <2 days at 29°C) and high female fecundity (300 oocytes on average) offer incomparable perspectives for genetic analysis among nonvertebrate chordates. Here, we show that *Oikopleura dioica* is also attractive for genome-based studies, because of unusually high gene density and gene compaction.

Flow cytometry comparison of sperm DNA content between *Oikopleura* and the ascidian *Ciona intestinalis* yielded a value of 72 ± 13 Mb for the *Oikopleura* genome. In a large-scale shotgun sequencing, 128,386 short reads were assembled in 43,990 nonredundant contigs representing a total of 32.6 Mb. These data, treated with an existing model (*3*), suggest a minimal genome size of 51 Mb. The shotgun data set also accurately matched, on average, 50% of the length of 113 expressed sequence tags (ESTs), leading to a third estimation of 65 Mb. All these results concur to indicate that *Oikopleura* has the smallest genome ever found in a chordate.

We then compared the density of genes in *Oikopleura* and in *Ciona* through a BLASTX analysis of 407 1-kb-long genomic fragments from *Oikopleura* and four cosmid sequences from *Ciona* (*4*) divided into 1-kb fragments. The BLASTX alignments at a $10^{-09}$ e-value cutoff were twice as frequent with *Oikopleura* fragments (25% compared with 12%) and twice as long for *Oikopleura* (average 45 amino acids compared with 26 amino acids per kilobase, respectively). BLASTX analysis of 113 ESTs showed that this program failed to detect half of *Oikopleura* coding sequences. All BLASTX results suggest for *Oikopleura* an average gene density of one gene every 5 kb and about 15,000 genes (*4*).

To obtain a view of overall genome organization, we randomly cloned, sequenced, and annotated a few larger genomic fragments. One of them was a 145-kb-long bacterial artificial chromosome (BAC) insert for which we predicted 33 genes (Fig. 1B). The coding nature of 28 of them was confirmed by systematic cloning/sequencing of partial cDNAs. Another 18 genes were predicted in four distinct 12- to 14-kb plasmid inserts. In most genes, many small introns were found, with 62% of all introns smaller than 50 base pairs (bp) and 6% larger than 500 bp. Predicted intergenic regions were also very short, in several cases as small as 200 bp. Short repetitive sequences were detected in the largest intergenic regions as well as in some large introns. Four well-conserved genes of the BAC insert (stomatin, retinoblastoma-binding protein 7, protein phosphatase 2A PR65 subunit, and glycine transporter 1/2) were much shorter than their best candidate human orthologs (0.8, 1.6, 2.2, and 2.1 kb for *Oikopleura* genes compared with 15, 30, 35, and 12/50 kb in human, respectively). In these four conserved genes, only five intron positions out of 24 in *Oikopleura* and 41 to 43 in human were identical.

It is still premature to establish a link between the small genome and the short generation time of *Oikopleura*, even though a correlation between longevity and genome size has been found in birds (*5*). It is also unclear whether the high gene density uncovered in *Oikopleura* is an ancestral feature of chordates or, alternatively, if the genome has been markedly reduced in this lineage of urochordates. Regardless, important intron movements must have differentially affected the vertebrate and the *Oikopleura* lineages.

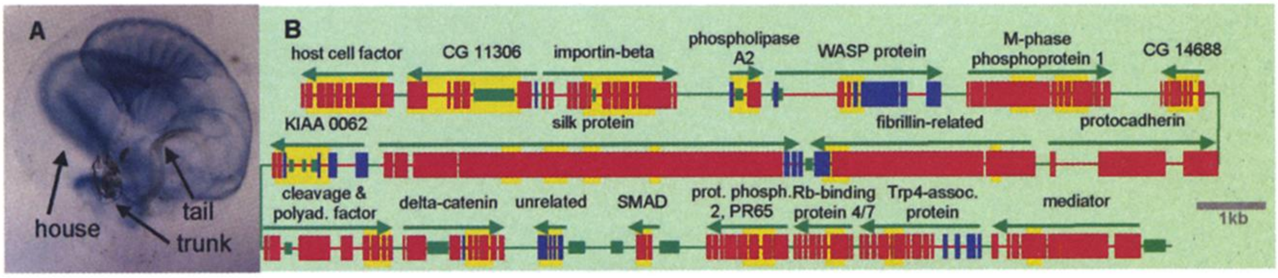

**Fig. 1.** (**A**) *Oikopleura dioica* in its "house," which allows seawater filtration. As an adult, the whole animal measures about 5 mm. (**B**) Organization of *Oikopleura dioica* genes in one BAC insert [see (*6*)], with exons detectable with BLAST (red), exons predicted by other tools (blue), regions of coding sequences confirmed by cDNA cloning (yellow), and repetitive sequences (green).

**References and Notes**

1. S. Bassham, J. Postlethwait, *Dev. Biol.* **220**, 322 (2000).
2. A. Nishino, N. Satoh, *Genesis* **29**, 36 (2001).
3. C. M. Fraser, R. D. Fleischmann, *Electrophoresis* **18**, 1207 (1997).
4. M. W. Simmen, S. Leitgeb, V. H. Clark, S. J. Jones, A. Bird, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4437 (1998).
5. P. Monaghan, N. B. Metcalfe, *Trends Genet.* **16**, 331 (2000).
6. Complete BAC insert, acknowledgements, and GenBank accession numbers are available at www.sciencemag.org/cgi/content/full/294/5551/2506/DC1.

---

[1]Sars Centre for Marine Molecular Biology, Thormøhlensgt. 55, 5020 Bergen, Norway. [2]Max-Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany. [3]CNRS-INSU/UPMC/LOV, Observatoire Océanologique, 06234 Villefranche, France.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: Daniel.Chourrout@sars.uib.no