Proposal: *Saccoglossus kowalevskii* (Hemichordata, Enteropneusta, Harrimanidae) as a candidate for having a BAC library constructed from its genome

From: John Gerhart
Department of Molecular and Cell Biology
LSA 142
University of California
Berkeley CA 94720-3200
e-mail: gerhart@socrates.berkeley.edu
Tel: 510-642-6382

1. The importance of the organism to biomedical or biological research:

Saccoglossus kowalevskii is a member of the phylum Hemichordata. Based on its development, adult characteristics, and phylogenetic position, it promises to yield insights about the origin of chordates, the origin of deuterostomes, and the origin of bilateral animals. Current molecular phylogenies place hemichordates and echinoderms together as the sister group of chordates, sharing a last common ancestor with chordates, the phylum to which vertebrates belong (including humans). Since echinoderms have evolved pentaradial adult organization whereas chordate adults are bilateral, it is difficult to make echinoderm-chordate comparisons of anatomy. Hemichordates, on the other hand, are bilateral worm-like animals as adults, and their comparison to chordates is easier. Since chordates and hemichordates (as well as echinoderm larvae) are bilaterally symmetric, their common ancestor is assumed to have been so.

William Bateson studied *S. kowalevskii* in 1884-6 and concluded that hemichordates are so chordate-like that they should be included in the chordate phylum. The similarities of hemichordates to chordates include:

1) gill slits (similar to those of amphioxus), up to 70 pairs,

2) a short stomacord homologized by Bateson to the notochord. It is formed from the archenteron roof as is the notochord (and prechordal plate) in chordates,

3) a dorsal nerve cord that formed by neurulation (discovered by TH Morgan) as in chordates but only along one part of its length, and

4) a transient post-anal extension somewhat like the chordate post-anal tail.

As time went on, though, the notochord and tail comparisons seemed weak, and the similarity of the nerve cord was hard to evaluate because the hemichordate nervous system turned out to be mostly an intra-epidermal diffuse nerve net. It is not a centralized nervous system. Nerve axons from the diffuse net extend to the dorsal and ventral midlines and form bundles which serve as through-conduction tracts. Hemichordates have two "nerve cords", and it is unclear whether the dorsal or ventral, or neither, resembles the chordate dorsal nerve cord. These comparisons have attracted attention recently in connection with proposals about the inversion of the dorsoventral axis in chordates relative to an ancestor their line shared with arthropods and annelids (the presumed ancestor of bilateral animals). Hemichordates occupy a critical place--are they inverted or not, and was inversion (as opposed to a dorsal-directed nerve condensation) the mode by which the chordate dorsal cord arose? By the 1940's the ambiguities of hemichordate-chordate comparisons convinced knowledgeable biologists (e.g. L. Hyman) that hemichordates should be placed in a separate phylum close to chordates. They were thereafter little studied, although to this day they are mentioned in textbooks as holding keys to chordate origins.

Attempts to address chordate origins will ultimately, we anticipate, have to include hemichordate-chordate comparisons. We (Dr. Marc Kirchner [Harvard Medical School], Dr. Chris Lowe [University of California, Berkeley], and myself) have initiated such comparison starting two years ago, literally bringing the organism out of the mud. From our in situ hybridization studies of 28 expressed genes thusfar, it appears that the hemichordate tripartite body (composed of a prosome [or proboscis], mesosome [or collar], and metasome [or pharynx and gut]) corresponds to the anteroposterior dimension to the chordate body, with:

•prosome-expressed genes (*six3, rtx, dlx, and pitx*) corresponding to those of the chordate ventral forebrain (eye-forming region) and roof of the mouth,

• mesosome- and anterior metasome-expressed genes (*tll, otx*) corresponding to those of the chordate dorsal forebrain and midbrain back to the first branchial arch. Remarkably, *otx* is expressed back to the first gill slit in hemichordates. And

•posterior metasome-expressed genes (*hox1, hox3, hox4, hox9*) corresponding to those of the hindbrain and spinal cord.

In the dorsoventral dimension, the developing embryo has a thin stripe of *bmp2/4* expression at the dorsal midline and a thin stripe of *netrin* expression at the ventral midline. The entire ectoderm is neurogenic, as known from histology of the 1950's and from our recent finding of ubiquitous *sox2/3* expression in the ectoderm. For comparison, it is noteworthy that the chordate neural tube has *bmp2* and *bmp4* expressed in the dorsalmost region, the roofplate, and *netrin* expressed in the ventralmost region, the floorplate, with neurogenic ectoderm in between, expressing *sox2* and *sox3*. The Bmp proteins of the neural tube are thought to specify sensory fates in dorsal neurons, whereas motoneurons arise ventrally. Netrin protein of the neural tube is thought to attract some axons to the ventral midline and repel others dorsally. It appears that the entire hemichordate body is organized like a chordate neural tube. However, the chordate neural tube expresses *shh* in the floorplate, and the Sonic Hedgehog protein plays a role in motoneuron specification, whereas *shh* seems to have play no role in the hemichordate nervous system. Furthermore, hemichordates produce an ectoderm that remians both epidermogenic and neurogenic during development, whereas chordates separate ectoderm into exclusive epidermogenic or neurogenic regions by way of dorsoventral Bmp-Chordin interactions. The question remains whether hemichordates even have these interactions.

Chordate-hemichordate comparisons are likely to extend beyond overall body organization to details of organ physiology and development, such as the origins of the chordate kidney from the hemichordate axial (podocyte-containing) complex, the chordate thyroid from the hemichordate endostyle, the chordate pituitary from the hemichordate proboscis pore, and aspects of chordate skin from the hemichordate epidermis (which has no chitinous exoskeleton, but probably does have cytoplasmic intermediate filaments). Organ and tissue comparisons to chordates are likely to be much more extensive than is afforded by arthropods or nematodes of the more remotely related ecdysozoa or by molluscs or annelids of the lophotrochozoa.

Hemichordate-chordate differences are also important as ones revealing novel aspects of chordate evolution: the hemichordate stomacord may be more like the chordate prechordal plate than a notochord. It expresses *dkk* and *otx* as does the chordate prechordal plate, but not *bra, admp,* or *shh*, the latter group of genes being expressed in the notochord in chordates. Notochord origins remain a mystery.

Looking beyond chordate origins, the superphylum group composed of chordates, hemichordates, and echinoderms constitutes the deuterostomes. Hemichordates promise to give information on the ancestor of deuterostomes

since this is also the ancestor of chordates. Only these three phyla remain as deuterostomes after several venerable members were recently re-classified as ecdysozoa (e.g. chaetognaths) or lophotrochozoa (e.g. phoronids). Finally, hemichordates promise to give information about the ancestor of all bilateral animals. Recent molecular phylogenies have eliminated all intermediate phyla between radial (or bi-radial) animals (ctenophores and coelenterates) and the 25 phyla of bilateral animals. Hemichordates are as deeply split from the bilateral line as any group. The presence of deuterostome traits in phyla of ecdysozoa and lophotrochozoa raises the possibility that this ancestor had deuterostome traits. Therefore, hemichordates occupy a key phylogenetic position appropriate to inform us about the ancestors of chordates, deuterostomes, and all bilateral animals.

2. Uses to which the BAC library would be put, in addition to genomic sequencing:

The BAC library will also be useful for acquisition of information about gene organization and about regulatory sequences. The extended Hox cluster is of course of interest to us, that is, the Hox cluster plus the NKL cluster plus the EMGbox cluster, which P. Holland and others propose to have existed as a huge single cluster in a chordate ancestor. Hemichordates are so basally split from the deuterostome side that they promise to reveal aspects of gene organization and gene regulation present before chordates arose and numerous genes underwent extrensive duplication and divergence reflected, for example, in the expansion of the wnt family, the TGF-beta family, and the Hox genes.

3. The size of the research community that could potentially use the BAC library and the community's interest in and support for having a BAC library:

At present the community is small (but intense). There are three researchers working with *S. kowalevskii*, namely, Dr. Marc Kirschner (Harvard Medical School), Dr. Christopher Lowe (University of California, Berkeley), and myself. After two years of work, we have made substantial progress and will submit several articles in the next few months about hemichordate-chordate comparisons (anteroposterior organization, dorsoventral organization and the diffuse nervous system, and the stomacord similarity to the prechordal plate). A large research community could potentially use a hemichordate BAC library since most chordate sequences, molecules, and molecular interactions are expected to be traceable to a hemichordate correspondence. Hemichordates are the phylum of bilateral animals closest to chordates.

Most venerable hypotheses about the origins of chordates go back to hemichordates but necessarily get very sketchy at this juncture due to lack of information about hemichordates. For example, the pituitary has been proposed to derive from the hemichordate proboscis pore (ES Goodrich, 1918). Related to this, we have recently found that *pitx* is indeed expressed there. *Pitx* is a vertebrate pituitary transcription factor (for POMC and GH). The thyroid has been long proposed to derive from the endostyle (a mucus-secreting specialization of the mouth cavity, to trap food particles for conveyance to the gut), which is large in hemichordates. Relevant to endostyle-thyroid proposals, we find a thyroglobulin sequence in our EST collection, as well as T4 and T3 deiodinases and the relevant thyroxine nuclear receptor sequence. There are many good reasons to have a hemichordate BAC library, and an eventual genome sequence, to facilitate chordate comparisons and insights into chordate origins.

4. Whether the organism will be, or has been, proposed to NHGRI or another publicly funded agency for BAC-based genomic sequencing and the status of that request:

To my knowledge, the only proposal for a BAC library (no sequencing, though) from *Saccoglossus kowalevskii* is from Dr. Linda Holland of UCSD, Scripps Institution of Oceanography, submitted to NSF recently. She listed numerous organisms of which *Saccoglossus* was one. It is possible that other researchers have proposed *Ptychodera flava*, an indirect developing hemichordate.

5. Other genomic resources that are available that will complement this resource:

As other resources we have:

• three high quality cDNA libraries, one from blastula-gastrula stages, one from neurula-gill slit stages, and one from two-week old juveniles (2 gill slits). These were prepared by Dr. Chris Gruber of Invitrogen/Life Technologies.

•5000 arrayed clones for which sequences are known.

•65,000 arrayed clones for which sequences are not yet known, but are planned for systematic analysis after the major known sequences are located and eliminated,

•an EST collection representing about 1200 blastx-identifiable contigs and an equal number of unidentifiable contigs (from 11,000 random sequences). This was kindly made by Dr. Eric Lander's laboratory (Dr. Nicole Stange-Thomann organizing the project). Dr. Lander has expressed an interest in the possibility of future sequencing.

•about 30 coding sequences have been fully sequenced, chosen from developmentally "interesting" genes (including *six3, rtx, dlx, pitx, tll, pax6, otx, pax1/9, hox1, hox3, hox4, hox9, bra, netrin, bmp2/4, bmp7, twg, tbx3 (omb), sox1/2/3, nkx2.2, mnx (HB9), gbx, dan, admp, dkk, sna,* and *thyroglobulin*). Many other sequences have been identified and are in the process of clone isolation and sequencing (including *gsc, tld, wnt1/14, notch, delta, serrate/jagged, hox2, hox6, meis, bf-1, emx, pax2.1, pax2/5 (poxneuro), pax3/7, en, gsh,* and *neurofilamentH*). While this wouldn't represent a large collection for a model organism, it is large for a non-model.

6. The strain of the organism proposed and rationale for its selection:

We propose *Saccoglossus kowalevskii* because it is a direct developing species amenable to intensive study of embryonic, juvenile, and adult stages. The egg is rather large, 0.4 mm, suitable for microsurgery and injection. Development to hatching of a juvenile (adult body organization but not yet sexually mature) takes 7 days from fertilization. Development resembles that of amphioxus with respect to cytoplasmic rearrangements, cleavages, and gastrulation. The anteroposterior axis develops in line with the animal-vegetal axis, not orthogonal to it as chordate embryos eventually do.

S. kowalevskii lives in silty-sandy intertidal zones of the east coast of the US, from Massachusetts to South Carolina. It was studied by William Bateson in 1884-6, by Arthur and Laura Colwin from 1956-62 at Woods Hole MBL, MA, and by us since 1998. It is collected by the MBL collecting staff (Ed Enos and Andy Sexton in charge). We have had success in extending the mating season by two months by cooling gravid animals in the laboratory.

Embryos can be collected in large numbers in season (40,000 collected and fixed last September for use for in situ hybridizations throughout the year). We have avoided indirect developing species because the tornaria larva takes several months to grow in the plankton and the egg is relatively small (0.1mm).

At this time, *S. kowalevskii* is the most fully described hemichordate in terms of its regional gene expression during development. We have performed high-quality in situ hybridization studies for the regional expression of 28 genes (including *six3, rtx, dlx, pitx, tll, pax6, otx, pax1/9, hox1, hox3, hox4, hox9, bra, netrin, bmp2/4, bmp7, twg, tbx3 (omb), sox1/2/3, nkx2.2, mnx (HB9), gbx, dan, admp, dkk, shh,* and *sna*). We have done double in situ hybridizations as well.

7. The size of the genome:
Dr. Kirschner has measured *S. kowalevskii* sperm DNA (cytophotometry, fluorimetry) and estimates that the genome is about 35% that of *Xenopus laevis* sperm (3.1×10^9 bases), namely, about 1.1×10^9 bases.

8. The availability of a source of DNA for construction of the BAC library (evidence of its quality for this purpose):
Sperm and sperm DNA from *S. kowalevskii* can be prepared fresh from males in two seasons: from March to early May and from late August to mid October. A single male contains 10^7-10^8 sperm in bilateral testes 3-5 cm in length. These can be dissected from the animal, free of other tissues. When the testis is mature, it is filled with packets of finished spermatocytes. The contribution from immature sperm, spermatogonia, or support cells is very small. Mitochondrial DNA contamination should be minimal (although each sperm is reported to carry 4 mitochondria). We have dissected testis and prepared sperm suspensions many times and can provide material for the embedment and lysis of sperm in agar blocks, or we can do those steps ourselves. We have made DNA by standard procedures for various uses such as pcr and genomic libraries, but the pieces would be too small for a BAC library.

9. Specifications for the library (e.g., library depth, BAC insert size) and supporting scientific rationale for these specifications. (Note: any request for an unusual vector for a particular application must be thoroughly discussed):
No special requirements.

10. Time frame in which the library is needed:
Any time, but within a year we would like to analyze the extended hox cluster.

11. Other support that is available or has been requested for the construction of the desired library:
No other support available to me. I have a small NASA grant (about $110,000 direct costs) for hemichordate work, ending July 2002. I am applying for a renewal but won't be able to include BAC library construction in the small budget. As noted above, Dr. Linda Holland has included *Saccoglossus* in her list of organisms for BAC library construction in a request submitted to NSF. She asked whether I would provide the *S. kowalevskii* sperm if her request is accepted, and I have said yes.

12. The need for an additional BAC library if one or more already exists:
No library available at present.

13. Any other relevant information:
None. If there are questions, let me know.