Genomics of Basal Metazoans

ROBERT E. STEELE
Department of Biological Chemistry and the Developmental Biology Center, University of California, Irvine, California 92697-1700

SYNOPSIS. An in-depth understanding of the biology of animals will require the generation of genomics resources from organisms from all phyla in the metazoan phylogenetic tree. Such resources will ideally include complete genome sequences and comprehensive EST (expressed sequence tag) datasets for each species of interest. Of particular interest in this regard are animals in the early diverging non-bilaterian phyla Porifera, Placozoa, Cnidaria, and Ctenophora. Publications describing the results from the use of genomics approaches in these phyla have only recently begun to appear (Kortschak et al., 2003; Yang et al., 2003; Steele et al., 2004). Issues to be considered here include choosing the basal metazoan species to examine with genomics approaches, the relative advantages and disadvantages of genome sequencing versus EST projects, and the resources and infrastructure required to carry out such projects successfully.

WHAT CAN ONE EXPECT TO LEARN BY DOING GENOMICS ON BASAL METAZOANS?

Before discussing the issues associated with pursuing genomic studies on basal metazoans, it is important to consider what one can expect to learn from such studies. If the whole genome of an organism is sequenced, one obviously expects ultimately to end up with a catalog of all of the genes for that organism. The value of such a catalog is substantial. It allows one to evaluate critically a number of interesting hypotheses related to evolutionary, physiological, and developmental processes. For example the absence of a gene from a cnidarian genome but its presence in bilaterian genomes could indicate that the gene evolved after the divergence of cnidarians. Alternatively, the gene could have been secondarily lost from cnidarians. To determine which hypothesis is correct one would then need to have information on this gene from other taxa, particularly ones which diverged prior to cnidarians. If, for example, sponges were found to have the gene, then one would conclude that secondary loss explains its absence from cnidarians.

Secondary loss is not as rare as might be expected. Kortschak et al. (2003) have provided a preliminary sampling of the genes in the coral Acropora millepora using an expressed sequence tag dataset of 1,376 clusters and found that 53 of the clusters represented genes that are present in vertebrates but absent from Drosophila and Caenorhabditis elegans. A number of additional cases have been found in which a gene is present in a cnidarian and in deuterostomes, but absent from the protostomes Drosophila and/or C. elegans (Chan et al., 1994; Steele et al., 1999; Fedders et al., 2004). Thus these genes were apparently lost secondarily from these protostomes. In essence, then, evolution has carried out gene knockout experiments in worms and flies that indicate that these genes are not required for at least some types of metazoans to survive in their natural environments. This gets around the problem with a number of gene knockout studies in mice, in which absence of an observable phenotype may be due either to compensation by paralogous genes or the failure of the laboratory to duplicate the animal’s natural environment.

Of particular interest would be genes that are absent from all basal metazoans but present in bilaterians. These would presumably represent genes that were selected for their role in a bilaterian-specific biological process. Proof of such genes would require complete genome sequences of representatives from each of the basal metazoan phyla. Of equal interest to biologists would be genes found only in one or more basal metazoan taxa. These genes would be candidates for involvement in taxon-specific biological processes. Examples of such genes would include those encoding proteins involved in taxon-specific signaling processes (e.g., those involved in reproduction), genes involved in taxon-specific defense against pathogens, genes involved in production of taxon-specific structures (e.g., the nematocysts of cnidarians), and genes involved in taxon-specific symbioses (e.g., coral/algal symbiosis). Having genome sequences from sponges and cnidarians in particular will potentially have far-reaching impacts on fields as diverse as ecology and medicine. For example, our understanding of the phenomenon of coral bleaching (Hughes et al., 2003; Pandolfi et al., 2003) might be enhanced by knowledge of a cnidarian gene set. Our understanding of the synthesis and biological roles of the large number of natural products, some of biomedical relevance, identified in marine cnidarians and sponges (Fingerman and Nagabhushanam, 2001) might also be enhanced by having in hand the sequences of all of the proteins from such organisms, presumably including the enzymes responsible for synthesis of such products.

GENOME SEQUENCING VERSUS ESTS

If one wants to make unequivocal statements about the gene content of an organism, a complete, fully annotated genome sequence is the only acceptable data-
set. An alternative approach to obtain information about the gene set of an organism is to generate an EST dataset. ESTs are produced by single pass sequencing from one or both ends of randomly selected clones from a cDNA library (Adams et al., 1992). The presence of a gene in an organism is confirmed if an EST for it is identified even in the absence of a complete genome sequence. However, the absence of a gene can be demonstrated only if the whole genome has been sequenced. Given this requirement, a complete genome sequence from each of the basal metazoan phyla is an obvious goal. In addition, complete genome sequences allow one to ask a variety of important questions related to evolution of genome organization and gene regulation. Some of the advantages and disadvantages of genome projects and EST projects are listed in Table 1.

Despite their incompleteness, EST datasets are powerful resources that allow one to pursue a variety of interesting questions. In basal metazoans, with their relatively simple cell compositions and life cycles, one has a chance of achieving close to saturation coverage of the gene set with an EST project of sufficiently large size and thus a reasonable approximation of a genome sequence with regard to gene content, a so-called “poor man’s genome” (Rudd, 2003). Automated computational methods for gene prediction have their value (Mathe et al., 2002), but comparison between the genome sequence and a large EST dataset is the only sure way to confirm that one has correctly identified protein-coding genes.

Table 1. Genome Sequencing Projects.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequences of all genes are obtained if a finished sequence is produced.</td>
<td>Expensive, particularly if done to finished quality.</td>
</tr>
<tr>
<td>Sequences of potential transcriptional regulatory regions are obtained.</td>
<td>Draft quality sequence will lack genes.</td>
</tr>
<tr>
<td>Genome organization information is obtained.</td>
<td>A large genome size can make sequencing prohibitively expensive.</td>
</tr>
<tr>
<td></td>
<td>Difficulty in demonstrating biomedical relevance may affect priority.</td>
</tr>
</tbody>
</table>

EST Projects

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potentially less expensive than a genome sequence.</td>
<td>Higher sequencing error rate than for genome sequencing.</td>
</tr>
<tr>
<td>Gene identification is easier due to absence of introns.</td>
<td>Multiple cDNA libraries and/or various technical procedures (e.g., normalization) needed to ensure that one achieves good coverage of the gene set.</td>
</tr>
<tr>
<td>Small-scale EST projects don’t require the involvement of a large sequencing center.</td>
<td>Yields only partial sequences for cDNAs less than about 1,000 base pairs (assuming both ends of a cDNA clone are sequenced).</td>
</tr>
<tr>
<td></td>
<td>Don’t get genome organization information.</td>
</tr>
<tr>
<td></td>
<td>Don’t get transcriptional regulatory sequences.</td>
</tr>
<tr>
<td></td>
<td>Don’t get all of the genes in the organism.</td>
</tr>
</tbody>
</table>

Both genome and EST projects generate valuable resources in addition to the sequences themselves

While the sequences are the goal of genome and EST projects, valuable research resources are generated as a product of such projects. These include libraries containing large pieces of genomic DNA—bacterial artificial chromosome (BAC) (Shizuya et al., 1992) and fosmid libraries (Kim et al., 1992). Such libraries provide material for doing transgenic experiments (promoters for driving expression of cloned genes introduced into embryos and adult tissues). cDNA clones identified by EST projects can provide reagents for examining gene expression via high throughput in situ hybridization studies (Pollet and Niehrs, 2001), protein interactions via yeast two-hybrid assays (Li et al., 2004; Formstecher et al., 2005), antibody generation by immunization with proteins expressed in bacteria (Baneyx, 1999), and production of proteins for structural studies (Finley et al., 2004).

Where do we stand currently with regard to genomics for the various basal metazoan taxa?

Genomics studies of basal metazoans are underway, but are still in the very early stages compared to such studies in model bilaterian systems. Below I review the current status of genomics efforts with basal metazoans.
Porifera

Genomic studies in sponges are of critical importance if we hope to understand the evolution of metazoans. Molecular biological studies with sponges have been carried out by only a small number of groups, and the number of sequences for protein-encoding genes from sponges in the public databases is relatively small (569 at the time of writing). This compares with 2,716 such sequences for cnidarians, plus about 150,000 more from the Hydra EST project (www.hydredbase.org). An NSF-funded project to produce a BAC library from the sponge *Callyspongia diffusa* is underway (http://www.nsf.gov/bio/pubs/awards/bachome.htm). This will be the first reported BAC library from a sponge. Although there are no entries for sponges in the EST database at GenBank (dbEST), an EST project is being carried out on the demosponge *Suberites domuncula* (Cetkovic et al., 2004; Chevreux et al., 2004). At present these sequences can only be searched using a database that requires registration (http://spongebase.genoserv.de/).

A recent exciting development in the area of sponge genomics is the initiation of a sponge genome project by the U.S. Department of Energy’s Joint Genome Institute (JGI) (http://www.jgi.doe.gov/sequencing/why/CSP2005/reniera.html). The goal of this project is to sequence the genome of the demosponge *Reniera*. *Reniera* larvae will be used as the source of DNA for this project (Leys et al., 2005) to avoid the presence of DNA from contaminating organisms which have plagued the characterization of sponge DNA (Costantini, 2004). Ultimately one sponge genome will not suffice. The phylum Porifera is believed to be paraphyletic (Borchiellini et al., 2001). Thus a true understanding of this phylum from a genomic perspective will require additional genome sequences.

Placozoa

The phylum Placozoa, with its single described species *Trichoplax adhaerens*, is of great interest to students of metazoan evolution because of its simple composition (Miller and Ball, 2005) and its resistance to secure placement within the metazoan tree (Wainright et al., 1993; Cavalier-Smith, 1998; Collins, 1998; Brooke and Holland, 2003; Cavalier-Smith and Chao, 2003; Ender and Schierwater, 2003; Martinelli and Spring, 2003). Placozoa has been placed in virtually all possible positions among the basal metazoans. It has been suggested to be a distinct phylum or to represent a secondarily reduced cnidarian. Attempts to resolve its relationship to other basal metazoans have so far focused on comparisons based on rRNA and a handful of protein coding genes (Schierwater and Kuhn, 1998; Martinelli and Spring, 2003).

Until recently, it was assumed that *Trichoplax adhaerens* was the only known species in the phylum Placozoa. A recent study of nuclear and mitochondrial rDNA sequences from placozoans collected from diverse locations indicates that the genus *Trichoplax* contains multiple species (Voigt et al., 2004). The evolutionary history of placozoans is clearly more complicated than would have been guessed.

A proposal to sequence the *Trichoplax* genome was submitted to the National Human Genome Research Institute (NHGRI) in 2002 and received a moderate priority ranking, but sequencing was not initiated. Subsequently a proposal was submitted through the Community Sequencing Program at JGI and was approved (http://www.jgi.doe.gov/sequencing/why/CSP2005/trichoplax.html). The small size of the *Trichoplax* genome, about 50 megabases (Grell and Ruthmann, 1991), means that a complete sequence of this genome will be very easy to achieve. It will be very interesting to see what the genome sequence has to tell us about the evolution of this strange little creature.

Cnidaria

Cnidarians are unquestionably currently at the forefront with regard to genomics approaches using basal metazoans. A large-scale Hydra EST project is underway in the United States and a smaller scale one has recently been completed at the National Institute of Genetics in Japan. Data from both projects have been deposited in the public databases. As discussed above, an analysis of a small-scale EST dataset from the coral *Acropora millepora* has recently been published (Kortschak et al., 2003), and this study has received considerable attention (Dennis, 2003; Campbell, 2004; Raible and Arendt, 2004) because of its unexpected findings regarding gene loss in *Drosophila* and *C. elegans*. An arrayed BAC library from the starlet sea anemone *Nematostella vectensis* has been completed (http://www.nsf.gov/bio/pubs/awards/bachome.htm) and is available to interested researchers from Children’s Hospital Oakland Research Institute (CHORI, http://baepac.chori.org/library.php?id=219). A large-scale *Nematostella* EST project is in progress (U. Technau and T. Holstein, personal communication). *Nematostella* was chosen for construction of the first cnidarian BAC library for several reasons, one of which was concern about insert instability with the more A+T-rich genomes of hydrozoans (e.g., 71% A+T in *Hydra*). This concern arose as a result of the severe insert instability problem encountered when attempts were made to produce BAC libraries from *Dickostyelium discoideum* (78% A+T) and *Plasmodium falciparum* (82% A+T). Despite this concern, the Bosch lab (University of Kiel, Germany) attempted to construct a *Hydra* BAC library. They have succeeded in producing a library with 2× coverage and an average insert size of 150 kb. A BAC library has also been constructed from the colonial hydrozoan *Hydractinia symbiolongicarpus* with 13× coverage and an average insert size of 132 kb (Luis Cadavid, U. of New Mexico, personal communication). Thus it now appears that the point at which base composition creates a significant problem for BAC insert stability lies somewhere in the narrow window between 71% and 78% A+T. Undoubtedly there will be segments of hy-
drozzoan genomes which will be unstable in BACs due to high local A+T content, but the discovery that it is possible to make BAC libraries from this class of cnidarians is an important step forward.

Without a doubt, the most exciting development in regard to basal metazoan genomics was the decision by JGI to sequence the *Nematostella* genome. At the time of this writing, the sequencing part of the project is nearly completed and assembly and annotation will begin soon. *Nematostella* will be the first basal metazoan to have its genome sequenced. The *Nematostella* genome project has been followed unexpectedly quickly by a project to sequence the genome of *Hydra magnipapillata*, the species being used for the *Hydra* EST Project. The *Hydra* Genome Project is being funded by NHGRI and is being carried out by the J. Craig Venter Institute. This project is targeted for completion by the end of 2005. Coverage has already reached about 1X, with a goal of 8X coverage. As they are generated, the sequences are made available on the Trace Archive at NCBI (http://www.ncbi.nlm.nih.gov/Traces/tracere.cgi)?.

**Ctenophora**

The ctenophores have so far seen the least effort with regard to genomics. Only 38 sequences from protein-encoding genes have been deposited in GenBank (27 from *Beroe*, 6 from *Pleurobrachia*, and 5 from *Mnemiopsis*). Only 75 ctenophore references are found in PubMed, covering a period of 51 years. This lack of attention relative to other basal metazoans is unfortunate since the relationship of ctenophores to other phyla is still unclear (Medina et al., 2001; Martindale et al., 2002; Rokas et al., 2003) and its resolution is critical to understanding metazoan evolution. Genomics efforts with this phylum would undoubtedly contribute importantly to this. Concerted efforts to clone homeobox genes of the Hox/Parahox class from ctenophores have been unsuccessful (Mark Martindale, personal communication), suggesting the intriguing possibility that ctenophores lack Hox/Parahox genes. A ctenophore genome sequence would be the definitive test of this hypothesis. Ongoing developmental studies with ctenophores (Martindale, 1986; Martindale and Henry, 1997; Martindale and Henry, 1999; Henry and Martindale, 2000; Henry and Martindale, 2001; Yamada and Martindale, 2002) will provide an important foundation for molecular studies arising out of genomics efforts with this phylum.

An important need with regard to ctenophore genomics initiatives is an organized scientific constituency which will lobby for them. A positive sign for such efforts is the fact that the ctenophore *Mnemiopsis leidyi* was one of the animals selected for BAC library construction by the National Science Foundation (http://www.nsf.gov/bio/pubs/awards/bachome.htm). The haploid genome size for *Mnemiopsis leidyi* has been reported to be \(3 \times 10^8\) base pairs (http://www.genomesize.com/ctenophora.htm). This is about twice the size of the *Drosophila* genome and thus well within the reasonable size range for sequencing.

**Choanoflagellates—not metazoans, but close**

Genomic investigations of the basal metazoan phyla will tell us a lot about the evolution of animals. However, to complete the story it will be important to have information on non-metazoan phyla whose divergence immediately antedates the origin of metazoans. Studies of the morphology of choanoflagellates in the 1800s noted the similarity of the their cells to the choanocytes of sponges (Saville-Kent, 1880–82), particularly with regard to the collar which surrounds the flagellum. This shared morphological trait suggested that sponges may have derived from a choanoflagellate ancestor and that modern choanoflagellates constitute a sister taxon to Metazoa. Molecular phylogenetic studies have provided strong support for such a relationship (reviewed in King, 2004). EST projects using two choanoflagellates (*Monosiga brevicollis* and *Proterospongia sp.*) have led to the surprising finding that choanoflagellates contain genes for a variety of signaling molecules that were thought to be restricted to metazoans (King and Carroll, 2001; King et al., 2003). These results make it clear that choanoflagellate genome sequences will be essential in understanding the evolution of basal metazoans.

Both the JGI and the NHGRI have committed to sequencing of choanoflagellate genomes. JGI will sequence the *Monosiga brevicollis* genome and NHGRI is funding the sequencing of the *Monosiga ovata* genome. Comparison of these two genomes to each other and to those of metazoans is certain to provide important insights into the origins of animals.

**Getting a Basal Metazoan into the Genomics Queue**

In choosing an organism with which to pursue genomic studies several criteria must be applied. EST and genome projects are by definition expensive operations. For genome sequencing and generation of BAC libraries, there are now well-defined routes to follow. For EST projects, the path is less well-defined. The NHGRI accepts proposals (called White Papers) for genome sequencing of organisms for its comparative genomics program. Proposals are reviewed three times a year. Instructions for submission can be found at http://www.genome.gov/10002189. Previously submitted proposals which have received either a high or moderate priority ranking can be examined at http://www.genome.gov/10002154 and provide excellent models for preparation of new proposals. Figure 1 shows the points to be addressed in a genome sequencing White Paper submitted to the NHGRI.

As Figure 1 shows, it is important that an organism being proposed for sequencing satisfy a number of criteria. The size and other basic features of the genome must be known. It is obviously much easier to make a case for sequencing if resources such as BAC libraries and/or EST datasets are available or are in the
A. Specific biological/biomedical rationales for the utility of new sequence data

1. Improving human health. How will the genomic sequence of an organism inform our understanding of human disease or hold the potential for improving human health? What, if any, is the relevance to the development of innovative and improved methods of diagnosis, treatment or prevention?

2. Informing human biology. How will the genomic sequence of a particular organism lead to a better understanding of biological function in the human?

3. Expanding our understanding of basic biological processes relevant to human health, e.g. cell or developmental biology, neurobiology.

4. Providing additional surrogate systems for human experimentation, e.g. new disease models, improved opportunities for drug testing, or other medical procedures, such as transplantation.

5. Facilitating the ability to do experiments, e.g. "direct" genetics or positional mapping, in additional organisms.

B. Strategic issues in acquiring new sequence data

1. The demand for the new sequence data. What is the size of the research community that will use it? What is the community’s enthusiasm for having the sequence? Will the new sequence data stimulate the expansion of the research community?

2. The suitability of the organism for experimentation. What are the basic properties of the organism that affect its ability to be studied in the laboratory (e.g. availability, ability to be cultured and propagated in the laboratory, generation time)? Are mutants available with defined phenotypes? How will the new sequence data enhance the experimental use of the organism? What other genomic resources and technologies (e.g. gene transfer, ability to go from molecule to mutation) are available that will allow the new sequence information to be effectively used?

3. The rationale for the complete sequence of the organism. Why would the complete sequence be more useful than the sequences of specific regions, or only the coding sequences, or only ESTs? Are there alternative ways to get the necessary information?

4. The cost of sequencing the genome and the state of readiness of the organism’s DNA for sequencing. What is the size of the genome? Will the repeat structure or other biological features pose challenges for genome sequencing? What quality of sequence product is needed (finished sequence? draft? full shotgun?)? What sequencing strategy will be used? Is suitable DNA readily available?

5. Are there other (partial) sources of funding available or being sought for this sequencing project? NHGRI is willing to consider requests to sequence disease vectors but priority will be given to requests that relate to the programmatic interests of NHGRI and where co-funding from other entities has either been identified or discussions about such funding is underway.

Fig. 1. Points to be addressed in a genome sequencing proposal (White Paper) submitted to NHGRI (from http://www.genome.gov/11509736).

process of being generated. And it is critical that the proposal demonstrate that there is a research community of sufficient size and vigor to make good use of the genome sequence when it becomes available. Whole-genome shotgun sequencing requires that the individual sequence reads be assembled into contigs and scaffolds. Assembly can be complicated by polymorphism in the DNA used for sequencing, even if the DNA that is used comes from a single individual (Dehal et al., 2002). Thus having an inbred line of the target organism is an obvious plus.

The JGI has recently established a Community Sequencing Program which accepts proposals for sequencing projects from the research community. Information on this program can be found at http://www.jgi.doe.gov/CSP/index.html. As with the NHGRI program, the JGI program has a well-defined proposal submission and review process.

For genome sequencing projects, one provides high quality DNA to the sequencing center and they construct the clones that are used for sequencing. In the whole-genome shotgun approach (Adams et al., 2000; Aparicio et al., 2002; Dehal et al., 2002), the DNA is mechanically sheared and then cloned into a plasmid vector. The recombinant plasmids are then end-sequenced. End-sequencing of BAC clones is done to link the shotgun sequences together.

For EST projects, there is no standard process for
project proposals as there is for genome sequencing projects. cDNA libraries for EST projects can be constructed either by individual investigators or RNA can be provided to commercial entities for library construction. The bacterial colonies from the resulting library must then be picked and arrayed in microtiter plates. The clones then have to be sequenced and the data analyzed. Both non-profit and for-profit entities can be engaged to do the sequencing.

**CONSTRUCTION OF BAC LIBRARIES**

BAC libraries are critical for genomics studies (Hoskins et al., 2000; Osoegawa et al., 2000, 2001; McPherson et al., 2001). Given the complexities of making high quality BAC libraries, most laboratories will not want to tackle this task on their own. It is best to have such libraries made by one of the four labs that constitute the BAC Resource Network (http://www.genome.gov/10001844). They have extensive experience in making BAC libraries from a wide variety of organisms. The NHGRI has a program for funding the production of BAC libraries. http://www.genome.gov/10001845. The NSF recently funded the production of a number of BAC libraries including ones from three basal metazoans—the sponge *Callyspongia diffusa*, the ctenophore *Mnemiopsis leidyi*, and the cnidarian *Nematostella vectensis* (http://www.nsf.gov/bio/pubs/awards/bachome.htm). We should soon have in hand high quality BAC libraries from representative species of three of the four basal metazoan phyla. These libraries will all be available to interested researchers through the BACPAC Resources Center at CHORI (http://bacpac.chori.org/). Specific information on how to obtain copies of these libraries when they are completed is available at http://bacpac.chori.org/sharingpolicy.htm.

**DATA PROCESSING**

Securing the money and generating the sequences are only part of the job associated with a genome or EST project. Analysis of the data and presentation of it to the community are next. The centers that carry out animal genome projects have well-established data processing pipelines and sequence release policies. They thus take care of most of the decision-making in this regard.

The data processing and sequence release aspects of EST projects are not as standardized as for genome projects. Much depends on where the sequencing is done. Some sequencing centers may process the data immediately and may require that data be released immediately to the public databases (e.g., the Genome Sequencing Center at Washington University). However, if the EST sequencing is done under a contract with a commercial entity or at an investigator’s home institution, it will likely be up to the investigator to determine when the data will be made publicly available. To facilitate deposition of the large amounts of data typically associated with an EST project, GenBank has developed a streamlined batch submission process (http://www.ncbi.nlm.nih.gov/dbEST/how_to.submit.html) as well as a batch process for updating the records from an EST project once they have appeared.

There is also the question of what to deposit in the public databases. Obviously the sequences themselves will be deposited. It is important that the investigators do a careful job of removing vector, linker/adaptor, and poor quality sequences from their datasets before submission. An additional helpful feature is to include the highest scoring BLASTX hit for each sequence (Gish and States, 1993). Additional information that is of value includes a brief description of the methods used to construct the library. An example of an EST entry for *Hydra magnipapillata* is shown in Figure 2. If a database is created to hold the EST data, the investigators should, as much as possible, use open source software (http://www.gmod.org/) that has been used successfully with other gene databases.

The data from EST and genome projects have their maximum impact when high quality gene predictions are done and the genes are expertly annotated. In some ways this is the most difficult part of the project. Automated annotation can be of some help, but such annotations have a significant error rate. The gold standard is manual annotation by expert annotators. This is obviously both expensive and slow. Alternatives include annotation jamborees and distributed, community annotation. An important aspect of annotation is the assignment of Gene Ontology (GO) terms (Ashburner et al., 2000), which provide a powerful tool for organizing and comparing gene data.

**REAGENT DISTRIBUTION**

As mentioned above, EST and genome projects create valuable reagents which the research community will undoubtedly want access to. In the case of EST projects, these would be arrayed cDNA clones. In the case of genome sequencing projects, they would be arrayed clones from BAC and fosmid libraries. Thus an important consideration for such projects is how the DNA clones will be archived and distributed. Archiving and distribution are expensive (freezers) and labor-intensive (someone has to maintain the freezers, do the record-keeping, and ship the clones). Thus most labs would not want to do this themselves. Fortunately there are centers that will archive and distribute arrayed cDNA libraries (e.g., The I.M.A.G.E. Consortium at Lawrence Livermore National Laboratory; http://image.llnl.gov/). BAC and fosmid libraries may be archived and distributed by the centers where they were constructed (e.g., the BACPAC Resources Center at CHORI). The investigator must anticipate up-front costs of deposition of clones at a distribution center.

**THE NEXT STEP—COMPARATIVE GENOMICS WITHIN BASAL METAZOAN PHYLA**

Based on efforts underway and planned, it seems all but certain that genome sequences will eventually be obtained for representative species from each of the
Genomics of Basal Metazoans

Fig. 2. An example of a GenBank entry for a sequence from the Hydra EST Project.
basal metazoan phyla. Recently, it has become clear that having genome sequences from several organisms within a single taxonomic group provides a very powerful means for addressing questions of evolution (Kellis et al., 2003; Stein et al., 2003). While we are currently at the stage where getting a single genome and/or large scale EST project done for each basal metazoan phylum is the goal, the basal metazoan research community should be thinking ahead and anticipating its needs and interests with regard to genomics initiatives and resources. We should begin lining up the next wave of organisms, paying particular attention to those which will be of the greatest immediate value for comparative studies. For example one might want to try to get a representative done from each of the four (and now possibly five) classes of cnidarians (Marques and Collins, 2004) before trying to do more members of a single cnidarian class. Alternatively one might want to concentrate on a single class so that one can do comparisons between taxa that are closely related but differ in life cycles. Such a comparison could be done between *Hydra*, which lacks medusa stage of most hydrozoans, and *Podocoryne*, which has both medusa and polyp stages (Galliot and Schmid, 2002).

The time is also ripe for investigators to do genome size surveys on additional basal metazoan species and to make attempts to bring new basal metazoan species into laboratory culture. Such material will provide the foundation for future genomic studies with basal metazoans.

**CONCLUSION**

In a period of only 15 years we have gone from the first publications describing the cloning of protein-encoding genes from basal metazoans (Bosch et al., 1989; Fisher and Bode, 1989) to having draft sequences of several basal metazoan genomes within reach. The impact of genomics efforts on our understanding of basal metazoan biology will undoubtedly be great. It will be particularly important for genomics researchers and organismal biologists to interact frequently and extensively so that the wealth of information waiting in the sequences is used as thoroughly and creatively as possible. Functional tests are beginning to be used successfully with genes from cnidarians (Lohmann et al., 1999; Lohmann and Bosch, 2000; Smith et al., 2000; Böttger et al., 2002; Miljkovic et al., 2002; Cardenas and Salgado, 2003; Jakob et al., 2004) and the first paper describing a functional test of a gene in *Trichoplax* has recently appeared (Wikramanayake et al., 2003). The maximum benefit will come as functional tests are applied to the many interesting basal metazoan genes that will become accessible to experimental manipulation from current and future genomics projects.

**ACKNOWLEDGMENTS**

This paper is dedicated to all the biologists who have kept the faith and worked so hard to bring basal metazoans into the genomics era, and to the genomics researchers who have welcomed us so enthusiastically into their community. I am grateful for the thoughtful comments on the manuscript from the associate editor and the two anonymous reviewers. Work in the author’s lab has been generously supported by the National Science Foundation.

**REFERENCES**


