Biomphalaria glabrata GENOME INITIATIVE

Why focus on ***Biomphalaria glabrata?***

The biology of ***Biomphalaria glabrata*** comprises many aspects that make this organism a logical choice for a molluscan genome project. Below, such aspects are highlighted from the standpoints of basic science and of infectious disease (~2001).

Obtaining the genome sequence of the mollusc ***Biomphalaria glabrata***: **A rationale from the standpoint of basic science**

Introduction

The beginning of the 21st century will long be remembered by historians as the age of genome biology. One of the challenges for biologists during this exciting period is to determine a logical order with which to procure the abundance of fascinating information lying within the genomes of Earth's various organisms. This challenge becomes particularly difficult now that a first tier of genomes that encompass the most prominent model organisms is in hand.

One of the major considerations in determining which eukaryotic organisms make their way onto a planned list of genomes to be completed is to begin to fill in some of the major phylogenetic gaps that exist in our current sample. Thus far, only a small sample of genomes representing the Kingdom Animalia is available: a nematode, an insect, and two mammals (humans and mice). Other projects already underway will provide additional genomes for a variety of vertebrates and medically significant parasitic animals, including additional nematodes and the flatworm, *Schistosoma mansoni*. The available sequences represent well the deuterostome lineage of animals, and the ecdysozoan lineages, but leave relatively under-represented one of the prominent groups of invertebrates, namely the more complex members of the Lophotrochozoa. Below, the case is made for obtaining the genome sequence for the freshwater snail *Biomphalaria glabrata*, a representative of a prominent lophotrochozoan lineage, the Phylum Mollusca.

The Case for Obtaining a Molluscan Genome

One of the most prominent lophotrochozoan phyla, and indeed one of the largest of all animal phyla, is the Mollusca. There are an estimated 50,000 extant molluscan species, making it one of the largest and most successful of all phyla. Furthermore, their success is not transitory – molluscs have one of the best documented fossil records of any animal group and have played a prominent role in animal life since their origins in the Cambrian period over 550 million years ago. Of all the animals on earth, perhaps none have adopted body plans and life styles more distinctive than molluscs. Molluscs are remarkable for their possession of a soft, mucus-covered body frequently protected by a shell and a file-like radula for obtaining and processing food. They are often unusual in their coloration and body organization, making them unique among animal life. Molluscs occupy habitats ranging from the oceanic depths to the tops of trees in tropical rain forests. As invertebrates go, molluscs like *Architeuthis* are the largest and the octopus the most intelligent. Some molluscs like squids have adapted to pelagic existence and are the equals of fish with respect to their swimming speed and ability. Some of the longest lived of all animals are molluscs: freshwater unionid bivalve species routinely live for more than 100 years. The unionids also emphasize the point that many molluscs are endangered and face extirpation. Molluscs also play a prominent role in the lives of humans. No other invertebrate group is as

frequently exploited as a source of food or for products of commercial significance. Molluscs also play an essential role in the life cycles of many parasites, including some, like the digenetic trematodes, that are widespread and significant pathogens of human beings and domestic animals.
Finally, there is an argument to be made purely from an aesthetic point of view. Molluscs are often arrestingly beautiful – if you have ever watched a cuttlefish or squid hover in the water, watching you watch them, you know you are viewing one of the pinnacles of evolution's accomplishments. To our knowledge, presently there is no initiative underway anywhere in the biological community to obtain the genomic sequence of a mollusc. The prospects that a molluscan genome holds for the discovery of novel genes and as yet unglimpsed biochemical or physiological capabilities are extraordinarily exciting.

Molluscs as Model Organisms

Although there is a wealth of information pertaining to the paleontology and ecology of molluscs, the group as a whole is remarkably understudied at the molecular level. This is true with respect to the application of molecular methods to reveal phylogenetic relationships among molluscs, and the unique developmental and physiological pathways undertaken to produce their distinctive body plans. When molecular techniques have been applied to molluscs, the results are often surprising, and provide general enlightenment that extends well beyond the limited confines of molluscan specialists.
One example of how the study of molluscs has provided broad enlightenment is in the field of neurobiology. Molluscs like the squid were used early on to define the basic physiology of axonal conductance and the strong tradition established in molluscan neurobiology is carried on today in groundbreaking studies of molluscs like *Aplysia* and *Lymnaea.* Studies of gastropods are providing fundamental insights into how neuronal plasticity develops, and the underlying molecular and cellular basis of learning and memory. Cephalopods, like the octopus, are prominent models for the study of vision and capacity for problem solving in invertebrates. Another molluscan group that has provided novel biological insights is the cone snails (*Conus*). A single species of cone snail can produce over 100 different toxic peptides and studies of such conotoxins have provided new insights into a diversity of mechanisms that can be used for post-translational modification of peptides. The mechanisms used by molluscs to protect their moist body surface from pathogens is a topic that has barely been approached but should yield a wealth of new information regarding peptides and other anti-microbial factors. Gastropods play an essential role in the transmission of most species of digenetic trematodes, and studies of gastropod-digenean interactions have done much to enhance our overall understanding of host-parasite interactions. The evolution of sexuality in response to parasitism has been effectively studied using a gastropod-digenean system. Because digeneans typically castrate their molluscan hosts, they provide an excellent model system to explore the physiological mechanisms exploited by parasites to re-direct the energy metabolism of their hosts.

Biomphalaria glabrata as a Model Organism

One of the most commonly investigated of all molluscs is the freshwater gastropod *Biomphalaria glabrata*. This distinction is well-deserved because this snail serves as one of the most important intermediate hosts for a widespread pathogen of humans, the digenetic trematode *Schistosoma mansoni*. To a large extent, the geographic distribution of this snail defines the distribution of S. mansoni in the Western Hemisphere. The snail is widely distributed on several Caribbean islands, and in extensive areas of South America, especially in Brazil. *Biomphalaria glabrata* also hosts a variety of other digenetic trematodes and has been adopted as the most commonly used model host to study the basic biology of digenean-snail interactions. One of the advantages of working with *B. glabrata* is that it is easily maintained in the laboratory.

Several studies underway with *B. glabrata* highlight its contribution to fields like evolutionary biology, parasitology or comparative immunobiology. As one example, *B. glabrata* has been found to produce after exposure to digeneans a unique family of hemolymph molecules termed FREPs (fibrinogen-related proteins). FREPs consist of a unique juxtaposition of fibrinogen and immunoglobulin superfamily domains, and have proven to be remarkably diverse in their composition. *Biomphalaria glabrata* thus serves as a new model system to examine the nature and diversity of non-self recognition molecules produced by invertebrates. Other studies with *B. glabrata* have begun to reveal the phenomena underlying adherence of invertebrate defense cells to foreign objects, and the mechanisms used by such cells to kill helminth parasites. A cell line derived from *B. glabrata* embryos has been used to support, for the first time, the complete in vitro development of a digenetic trematode. The genes responsible to resistance to digenean infection are actively being sought using

B. glabrata. The possible role of transposons in altering susceptibility of snails to infection is also being actively pursued using *B. glabrata*.
Examination of GenBank reveals that more sequence data by far are available for *B. glabrata* (approximately 1400 nucleotide sequences) than for any other mollusc. Collectively, molluscs have only about 8400 entries. The more this database grows, the more *B. glabrata* will be used as a model organism. The genome of *B. glabrata* is distributed on 18 pairs of small chromosomes and is estimated to be 9.31 x 10^8 base pairs in size (TR Gregory, University of Guelph, Ontario, Canada), with a CG content of 46%.

The Rationale for Obtaining the Genome Sequence for *Biomphalaria glabrata*

A concerted attempt to obtain the genomic sequence for *B. glabrata* would be valuable for several reasons:

1. Genome sequencing efforts would thus include a member of a large, as yet unrepresented lophotrochozoan phylum, the Mollusca. The study of molluscan biology will be seriously impaired if there is not an attempt to provide a genome sequence for a representative mollusc.
2. *B. glabrata* represents the most abundant class within the Mollusca, the Gastropoda, a lineage that has been successful over geological time and that remains successful today.
3. *B. glabrata* has already proven its worth as a valuable model for basic biological studies. By providing a genome sequence, its value as a model would be further increased, and the provision of the sequence would greatly assist ongoing studies of *B. glabrata* biology. Furthermore, the application of novel techniques to study proteomes and transcriptomes rely increasingly on the availability of sequence data.
4. The genome itself is likely to be interesting for several reasons. *B. glabrata* is an exclusively tropical, aquatic, hermaphroditic organism, a combination of attributes not associated with any animal thus far sequenced. The molluscan genome will yield interesting insights regarding how a shell is formed, how the moist molluscan body surface is protected from pathogens, and how an asymmetrical body plan is produced.
5. There is an active international community of scholars working with *B. glabrata* to both assist in procurement and analysis of the sequence data, and to use it in the future.

Coen M. Adema and Eric S. Loker

Obtaining the genome sequence of the mollusc ***Biomphalaria glabrata***: **A rationale from the standpoint of infectious disease**

Freshwater snails of the genus *Biomphalaria* are the intermediate hosts for *Schistosoma mansoni*, the most widespread of the three species that cause schistosomiasis in humans. This progressive and debilitating disease is one of the most intractable public health problems in many parts of the developing world. By most estimates, up to 10% of the world's human population is infected with one of the three major schistosome species. The decline in public health measures and sanitation, along with construction of dams and new irrigation schemes, serve to spread the disease into previously unaffected regions.

The most thoroughly studied snail host for schistosomes is *Biomphalaria glabrata*. It is the one most closely associated with schistosomiasis in the Western hemisphere, but has also been reported in Egypt, having been introduced there accidentally sometime during the last few years. Of the schistosome snail hosts, it is the easiest one to maintain in the laboratory, and the volume of scientific literature on this species exceeds that of all others combined.

Since this snail species is linked to such an important human health problem, studies on its genetics have largely been focused on how this may affect its relationship with *S. mansoni*. These studies began in the mid-1950's, when it was found that susceptibility to infection by *S. mansoni* is a heritable trait. Since then, we have found out a great deal about the genetics of parasite/snail compatibility. It is hoped that by identifying genes and their products that interfere with parasite survival in the snail, we may develop better methods of controlling transmission of schistosomiasis to humans. Information on the nature of those genes involved in the host parasite relationship however is still rudimentary. Efforts to identify genetic loci using a marker driven

positional mapping approach is now under way, but this method is technically challenging because little in the way of either genetic/physical or RFLP maps exist for the *B. glabrata* genome.

Much of the problem centers around the genome size of *B. glabrata*, which is estimated to be 9.31 x 10^8 base pairs in size (TR Gregory, University of Guelph, Ontario, Canada). The chromosomes (haploid number = 18) are small, relatively monomorphic, and have been organized into groups according to size and shape. To better understand the molecular make-up of *B. glabrata*, various gene libraries (cDNA, genomic, cosmid, BAC) have been constructed, and several laboratories are actively engaged in gene identification efforts.

It is clear that the entire phylum Mollusca is under-represented in proportion to its numbers and importance. A recent search of the public databases revealed the following information (table) on the nucleotide sequences of molluscs that have been deposited.For *B. glabrata* several genes have been sequenced and characterized, and there currently (November 2001) are approximately 1400 ESTs deposited in GenBank.

organism	seq#
Mollusca (all species)	11229
Biomphalaria sp	1673
B. glabrata	1525
Aplysia sp.	224
Oncomelania sp.	158
Lymnaea sp.	111
Bulinus sp.	49

Compared to the genomic studies of invertebrate vectors of other parasitic diseases (most notably several mosquito species), molecular biology studies of these molluscs are considerably less advanced. The genomes of several invertebrates have now been sequenced and the molecular make-up of organisms such as *Drosophila* and *Caenorrhabditis elegans* are forthcoming. In the field of tropical medicine sequencing of the malaria parasite and its vector host *Anopheles gambiae* are in progress, and several other genome projects are underway, including that of *S. mansoni*. We realize that mapping the genome of *B. glabrata* would be one of the larger sequencing efforts to be undertaken in the infectious disease arena. The molecular information already gathered on this species however places it in the forefront of molecular studies for any member of the phylum Mollusca. In the animal kingdom, members of this phylum are second only to those of the Arthropoda (insects, spiders, crustaceans) in their numbers and diversity. Information on the genome of *B. glabrata* would also have relevance for molluscan species that serve as hosts for a number of other trematode, and some nematode, infectious agents. Besides schistosomiasis, diseases such as fascioliasis, clonorchiasis, and paragonimiasis represent only a few of the snail transmitted diseases with worldwide medical and economic impact.

Since schistosomes alternate between a vertebrate and invertebrate host, we believe that ongoing sequencing efforts for *S. mansoni* will produce some significant gaps in our knowledge without having comparable sequence information for its snail host. The degree of parasite differentiation in the snail is much greater than it is in the mammalian host, and gene expression by the parasite in snail tissue is likely more varied.

As evidenced by the success of the human genome project, and the sequencing efforts of several other complex genomes, the technology certainly exists for sequencing this species as well. Such an effort will likely be the most efficient way (in labor and material cost) to advance our molecular knowledge in such a complex system.

We propose that, for a meaningful start to a genome project for this organism, our collective efforts could be organized into 2 major phases. The approach proposed is flexible and may be adjusted on consultation with other investigators.

Phase 1 – Presequencing approaches

Using a multigroup effort, the labor for this could be partitioned in the following way

- Initiate "gene discovery" (EST) projects by individual investigator labs
- Construct BAC libraries with inserts at least 120Kb with 5- to 6-fold coverage
- Construct cDNA libraries from specific regions or tissues of interest for EST-based sequencing
- Map new or already known genes to the BACs so that contigs can be identified
- Map BACs to specific chromosomes

Phase 2 - Sequencing approaches

- Employ whole genome shotgun sequencing, or shotgun sequencing of relevant chromosomes or BACs

In summary, schistosomiasis research has been supported by NIH-NIAID for roughly 50 years. As a consequence great strides have been made in understanding the biology of the parasite and, most notably, deciphering immune components of the disease process in the mammalian host. We feel the time is appropriate now to sequence the genome of *B. glabrata*, thus giving more relevance to the information coming from the sequencing efforts for *S. mansoni*, and spearheading the NIH-supported sequencing movement into an entirely new phylum of medically and economically important organisms.

Fred Lewis and Matty Knight