Request for making a BAC library from *Biomphalaria glabrata* (gastropod mollusc), the prominent snail species contributing to transmission of human schistosomiasis.

by **Coen Adema[1*], Matty Knight[2], Fred Lewis[2] and Eric (Sam) Loker[1]**.
Members of the *Biomphalaria glabrata* Genome Initiative.
[1]Department of Biology, University of New Mexico, Albuquerque, NM 87131 USA.
[2] Biomedical Research Institute, Rockville, MA 20852 USA.
*Contact information, phone 505 277 3134, fax 505 277 0304, e-mail coenadem@unm.edu

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

Biomedical research.

Freshwater snails of the genus *Biomphalaria* are the intermediate hosts for *Schistosoma* , the most widespread of the three parasite species that cause
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 suffer from schistosomiasis. 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,* most closely associated with schistosomiasis in the Western hemisphere. In addition, the recent accidental introduction in Egypt of this snail host may further exacerbate an already existing public health problem in this part of the world. Because of this snail species medical significance, studies on its genetics have largely been focused on how these may affect its relationship with *S. mansoni*. Susceptibility to infection by *S. mansoni* proved to be a heritable trait, and a great deal has been learned about the genetics of parasite/snail compatibility. The identification of genes and their products that interfere with parasite survival in the snail may benefit the control of transmission of schistosomiasis to humans. Despite a marked increase in the application of molecular methods, information on the nature of genes involved in the host parasite relationship remains rudimentary.

In the field of tropical medicine, several parasite genome projects are underway, including *S. mansoni*. Compared to the genomic studies of invertebrate vectors of other parasitic diseases (such as the malaria mosquito *Anopheles gambiae*), molecular study of *B. glabrata* as intermediate host for *S. mansoni* is considerably less advanced. Since schistosomes alternate between a vertebrate (human) and an invertebrate host, the available human genome sequence and the ongoing sequencing efforts for *S. mansoni* will likely leave some significant gaps in our understanding of schistosomiasis if comparable sequence information for the snail host remains unavailable. The degree of parasite differentiation in the snail is greater than it is in the mammalian host, and gene expression by the parasite in snail tissue is likely more varied. 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 nematode infectious agents. Besides schistosomiasis, diseases such as paragonimiasis, clonorchiasis, and fascioliasis represent only a few of the snail transmitted diseases with worldwide medical and economic impact.

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. The availability of a BAC library representing DNA from *B. glabrata* will provide a fulcrum for further study of the biology of the schistosome lifecycle.

Biological Research.

At this time, molecular data are becoming available on a vast scale, enabling sweeping comparisons between diverse organisms towards a more profound understanding of general biological phenomena. The selection of organisms for molecular study critically determines the extent of the possible comparisons. Most information has been obtained from a selected group of vertebrates and invertebrates. Within the protostome invertebrates, most attention has been directed towards insects and nematodes, both belonging to the Ecdysozoa, or molting animals. Little information has been gathered for organisms of the other clade of protostome invertebrates, the Lophotrochozoa. This clade includes the Mollusca, a phylum second in size only to the arthropods. The Mollusca are an ancient phylum that likely predates the Cambrian

radiation. It includes the largest (giant squid), most intelligent (cephalopods), and longest living (Unioid bivalves live for several decades) extant invertebrates. Molluscs are an important source of food but also impact humans as pests and by transmission of disease. *Biomphalaria glabrata* represents the most abundant class within the Mollusca, the Gastropoda. From a biological perspective, *B. glabrata* combines unique properties in that it is an exclusively tropical, aquatic, hermaphroditic organism. *Biomphalaria glabrata* has already proven itself as a model system in fields of basic biological study as diverse as comparative immunology, evolutionary systematics, biogeography, and environmental toxicology. A relatively modest number of 1530 nucleotide entries in GenBank has been derived from *B. glabrata*. While this provides an indication that molluscs are underrepresented in molecular studies, this number also makes *B. glabrata* the molluscan species for which most molecular data is available. With continued study, *B. glabrata* can become a model organism that yields 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.

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

The genome of *B. glabrata* has a GC content of 46% and comprises 18 (haploid number) chromosomes that are small and relatively monomorphic. About 15 genes of *B. glabrata* have been characterized full-length (at cDNA or DNA level). Little is known about the genome regarding nature of repetitiveness, mobile genetic elements, or arrangement of genes. Previous research has identified questions relating to e.g. diversity and distribution of large gene families of putative innate defense factors and of transposable elements in the genome of *B. glabrata*. Further investigations would directly benefit from having a BAC library for obtaining contiguous full-length sequences (also in light of characterization of EST sequences), determining numbers of genes and investigation of clustering of genes and identifying regulatory sequences. Additionally, several known genes would be mapped to BACs, and BAC clones will be developed as markers for individual chromosomes towards initial mapping of the genome.

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.

A BAC library for *B. glabrata* will provide relevant information for several diverse fields of research that are already involved with snails and other molluscs, such as malacology, parasitology, comparative immunology, (neuro-)physiology, endocrinology, pest control, and evolutionary systematics. At present, no high quality BAC library is available for any member of the phylum Mollusca. If this void can be filled, researchers from additional fields may also be attracted to the novel possibilities.
Specifically, this proposal is presented in light of the *Biomphalaria glabrata* Genome Initiative. This is an international affiliation of some 40 researchers from all five continents of the world, that have expressed interest in genome-type analysis of *B. glabrata* (letters available on request, * indicates willingness to actively contribute to gene discovery and other genome-type research). As pointed out by one of the members, Dr D. Johnston UK, Secretary to the WHO/UNDP/World Bank Schistosoma Genome Network <http://www.nhm.ac.uk/ hosted_sites/ schisto/index.html>, information derived from a *B. glabrata* BAC library would greatly benefit schistosome research. Some parasite stages can not be obtained without snails, and without appropriate sequence information it is difficult to distinguish parasite and snail products.

Supporters of the *Biomphalaria glabrata* genome initiative

Coen Adema* - USA
Gennady Ataev- Russia
Chris Bayne* - USA
David Blair - Australia
Paul Brindley* - USA
Omar Carvalho - Brazil
Christine Coustau* - France
Jason Curtis* - USA
Lawrence A. Curtis - USA
Marijke deJong-Brink - Holland
Colette Dissous - France
Georges Dussart* - UK
Sharon File-Emperador - USA
Petr Horak - Czech
Renzo Nino Incani* - Venezuela
David Johnston - UK
Catherine Jones* - Scotland
José Jourdane* - France
Bernd Kalinna - Germany
Matty Knight* - USA
Thomas K. Kristensen*- Denmark
Hammou Laamrani* - Morocco
Fred Lewis* - USA
Eric Loker* - USA
Nicholas Lwambo* - Tanzania
John Malone - USA
Don McManus* - Australia
Dennis Minchella* - USA
Gerald Mkoji* - Kenya
Helene Mone - France
Les Noble* - UK
Guri Roesijadi* - USA
David Rollinson* - UK
John Sullivan* - USA
Herve Tettelin TIGR - USA
Andre Théron - France
Jackie Trigwell* - UK
Mingyi Xia* - China
Tim Yoshino* - USA
Ulrike Zelck* - Germany

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.

Not to our knowledge.

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

Several laboratories across the world are actively studying molecular biology of *B. glabrata*. Based on interactive exchange of information and methods within this research community, it is valid to conclude that know-how and expertise are available to study *Biomphalaria* with standard molecular techniques. Currently, several cDNA libraries are available, representing different *B. glabrata* strains, parasite-infected and uninfected snails, and specific tissues (cerebral ganglia, albumen gland, hemocytes or circulating defense cells). Genomic libraries have been produced for several *B. glabrata* strains. With 1530 (1426 ESTs) nucleotide entries in GenBank, the amount of molecular data from *B. glabrata* exceeds that of all other molluscs. Initial efforts indicate the feasibility of obtaining intact high molecular weight genomic DNA from *B. glabrata* and subsequent production of BAC libraries.

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

Of several *B. glabrata* strains that are maintained in different laboratories, the BS90 strain is proposed for production of a BAC library. Since this strain is available from several groups, the culture of BS90 snails is less likely to crash catastrophically. The strain derives from an original field isolate from Brazil, and it has remained inherently resistant to schistosome infection. The BS90 is preferred over other laboratory strains of *B. glabrata* that result from crosses between different isolates and thus may not represent a wild-type genome. **NOTE**. This proposal received a positive recommendation from the Genome Resources and Sequencing Prioritization Panel (Aug 2002). Since then, NHGRI and the *Biomphalaria glabrata* Genome Initiative have interactively agreed to employ a *B. glabrata* strain that is **susceptible** for *S. mansoni* to generate the BAC library. Thus, molecular data collected from the *B. glabrata* BAC library will provide a relevant context for study of the intramolluscan biology of schistosomes. A new *B. glabrata* field isolate from Brazil, susceptible to *S. mansoni*, collected in Brazil (by Omar dos Santos Carvalho, Centro de Pesquisas René Rachou/Fiocruz, Brazil) will be maintained as a strain by several laboratories.

7 The size of the genome.

The haploid genome size or C-value of *Biomphalaria glabrata* is estimated at 0.95pg (± 0.01 SE). This converts (C-value x 0.98 $x10^9$) to **931Mbases** (T. Ryan Gregory, University of Guelph, Canada, see http://www.genomesize.com/). This value is based on analysis of hemocyte (blood cells) samples from three different strains of *Biomphalaria glabrata* (prepared by Coen Adema, University of New Mexico). The inbred M-line and 13-16R1 laboratory strains were included as these are used extensively in research, the BS90 strain represents an original field isolate of *B. glabrata* from Brasil. Integrated optical density (IOD) measurements were taken from 25 Feulgen-stained nuclei of hemocytes of individual snails (n = 14, total of 350 nuclei), and compared against >100 chicken erythrocyte nuclei (1C = 1.25pg), using an image analysis system. The variation in IODs was minimal between samples from individuals within and among strains.

8 The availability of a source of DNA for construction of the BAC library (evidence of its quality for this purpose).

It is feasible to isolate high quality genomic DNA from *B. glabrata* for construction of BAC libraries. The methods described here were part of initial efforts from Matty Knight to generate a BAC library in-house at the Biomedical Research Institute (MD).

After overnight incubation in ampicillin (100µg/ml), head-foot tissues were dissected from BS90 adult snails (10mm shell diameter) and ground to powder under liquid nitrogen on dry ice with pre-chilled (-70^{o}C) pestle and mortar. Powdered tissues were resuspended in 3 volumes of 50mM EDTA (pH 8.0) on ice and moved to 65^{o}C water bath. An equal volume of molten 1.0% agarose (FMC) in 0.125mM EDTA pH 8.0 (65^{o}C) was added. Aliquots (500µl) were transferred into moulds using a large pipette tip (cut and flamed for a smooth edge and a wide bore to prevent shearing of DNA), and allowed to set for 20 minutes at 4^{o}C. Resulting agarose plugs were transferred into pre-warmed CTAB DNA lysis buffer, 2mg/ml proteinase K and incubated 2days at 55^{o}C-60^{o}C with one change of buffer. The plugs were then rinsed (8x,

50mM EDTA pH 8.0, 55°C; 3x, TE pH 8.0, 37°C, 5-10 min intervals), incubated overnight in 200µM PMSF, TE buffer pH 8.0 at 37°C, and rinsed again (3x, TE buffer, 37°C; 4x TE at room temp). Plugs were then digested with *HindIII* overnight at 37°C in 1.0ml of incubation buffer.
The yield of genomic DNA per plug (10 plugs were derived from the head foot tissues of 6 adult snails) was estimated at 2-10µg by ethidium bromide staining intensity of undigested and *HindIII* digested samples on transverse agarose gel electrophoresis (TAFE) gels (1.0% agarose low melting agarose). The size of the undigested genomic DNA exceeded 1500 kb (see **Figure 1**), *HindIII* restriction yielded high molecular weight DNA fragments in the range of 100 kb (see **Figure 2**).
Genomic DNA recovered with these methods was used directly for BAC library construction The resulting clones contained inserts with an average size of 40.09 kb, ranging from 23 up to 89 kb. The insert sizes recovered do not seem to correctly reflect the size of DNA fragments that where used for the ligation. Problems during the freezing of the clones are a likely cause for loss of clones and of large inserts from this BAC library.

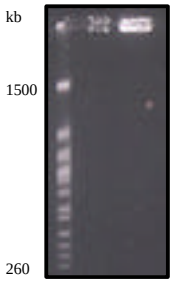

Figure 1. High quality genomic DNA from BS90 *B. glabrata*. Two different samples were run on a 1% TAFE gel. The size of genomic DNA (bands just migrated out of the wells) is in excess of 1500 kb, the upper band of the marker. Note that the absence of smearing indicates minimal degradation of DNA.

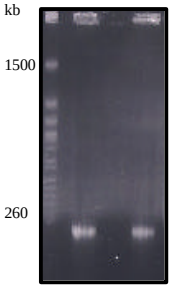

Figure 2. *HindIII* digestion of high quality genomic DNA from BS90 *B. glabrata*. Comparison with various molecular weight standards (intact λ phage DNA not shown) indicated that *HindIII* digestion of two samples yielded DNA fragments in the range of 100 kb.

9 Specifications for the library (e.g., library depth, BAC insert size) and supporting scientific rationale for these specifications.

General considerations. A standard, usual BAC vector system such as pBeloBAC11 is preferred. The likelihood that relevant know-how of a more standard vector system is generally widespread will increase the accessibility and relative ease of experimental manipulation of a BAC library by diverse members of the research community.

Specifications for the library. Initial applications for a BAC library include recovery of complete gene sequences and investigation of gene arrangements. To also allow more complex analyses such as initial mapping and genome walking, the specifications of the BAC library must ensure a reasonable representation of the whole genome sequence, as well as connectivity between different BAC inserts.

Insert size. General recommendations from the literature suggest that 100 kb is an appropriate average insert size in BAC libraries. As described above (**section 8**), high quality DNA can be obtained from *B. glabrata*, with an undigested fragment size upward of 1500 kb. Treatment with *HindIII* yielded fragments in the range of about 100 kb. Genes of *B. glabrata* that have been sequenced full-length at genomic level range from about 4 kb (e.g. *myoglobin*, deWilde *et al*., 1998 Journal of Biological Chemistry, 273:13583-13592) to 15 kb (e.g. *BgMFREP7*, Zhang *et al*., 2001, Immunogenetics 53:684-694) in size, from start codon to polyadenylation signal. Although the intergenic distance in *B. glabrata* remains unknown, it is likely that an average 100 kb BAC insert could accommodate at least several genes of *B. glabrata.* It is therefore not unrealistic to propose an average insert size of 100 kb.

Coverage. Again, literature provides a rule of thumb that a BAC library should provide at least 5x coverage of the genome size to allow genome mapping, by ensuring gene representation and connectivity of BAC clones. A 5-fold coverage of the 931Mb genome of *B. glabrata* requires 4655Mb of insert DNA, or 46,550 BAC clones with an average insert size of 100 kb. The following formula can be used to estimate the number of clones required of a certain insert size to give a particular probability that any given single copy sequence will be present in the BAC library.

N=ln(1-P)/ln(1-I/GS) where N= number of clones, P= probability (99% tested)
I= 0.1 Mb insert size, GS= 931 Mb genome size (for *B. glabrata*)

This calculation indicates that a BAC library, containing an average insert size of 100 kb, requires 42,872 clones to yield a 99% chance of finding any given unique sequence. Therefore, a 5x coverage of the genome, providing close to 99% probability that any given single copy sequence is present in the BAC library, requires at least 46,550 clones with an average insert size of 100 kb.

10 The time frame in which the library is needed.

One year. Several research groups (within and outside the USA) either are in the process of, or have already identified genes of *B. glabrata*, that are likely relevant for parasite/host interactions. The initial characterization of some of these genes has already prompted additional questions that could be addressed directly if a high quality BAC library were available. For example, such questions relate to identification of regulatory sequences, number of related sequences in the genome, and whether some genes are clustered or randomly distributed in the genome. Moreover, some members of the international *Biomphalaria glabrata* genome initiative have expressed interest in genome mapping. One group in particular (Biological Research Institute, MD) has initiated work in this area by generating a first (if not optimal) BAC library (see **section 12**, below). All these efforts would immediately benefit greatly from a high quality BAC library representing the genome of *B. glabrata*.

11 Other support that is available or has been requested for the construction of the desired library.

None.

12 The need for an additional BAC library if one or more already exists.

A first effort at generating a BAC library representing genomic DNA of *B. glabrata* inserts in the pBeloBAC11 vector was performed by Matty Knight (Biological Research Institute, MD). While the feasibility of producing a BAC library from *B. glabrata* DNA was demonstrated, unfortunately, problems during the freezing of the BAC clones caused the characteristics of the resulting first-effort library to be less than optimal. With an average insert size of 40.09 kb (range 23-88 kb) and 29376 clones available, the BAC library provides about a 1.3x coverage of the 931Mb genome of *B. glabrata*. As such, the pre-existing library may not provide adequate coverage of all chromosomes and lack connectivity, also it is sub-optimal for initial mapping studies. The pre-existing library was generated with genomic DNA from BS90 *B. glabrata*.

Clearly, a high quality BAC library representing *B. glabrata* genomic DNA is needed as optimal starting point to maximize chances of successfully encountering and characterizing single genes, and for initial mapping studies, and other genome-type projects in the future. The use of a *B. glabrata* strain that is **susceptible** to *S. mansoni* to generate the BAC library (mutually agreed upon by NHGRI and the *Biomphalaria glabrata* Genome Initiative, Aug 2002) will complement the pre-existing BS90 *B. glabrata* BAC library. Thus, molecular data collected from a BAC library derived from a susceptible *B. glabrata* will provide a relevant context for study of the intramolluscan biology of schistosomes.

13 Any other relevant information.

An informal consortium of international investigators has formed the *Biomphalaria glabrata* genome initiative. The members of this initiative plan to interact extensively to share and develop standard methodology, and to prevent duplication of investigate efforts. A uniform, standard nomenclature will be proposed for convenient and unequivocal annotation of *Biomphalaria* sequences. Sequence information obtained will be released to public databases in a timely fashion. A web page is under development to present progress and development of this initiative <http://biology.unm.edu/biomphalaria-genome/index.html>.

In light of space limitation imposed on this proposal, literature references have been kept to an absolute minimum. Upon request, the authors of this proposal can provide additional and more detailed documentation, including relevant literature references, in further support of the statements presented.