Snail Genome Meeting (September 1-2, 2005)

Four years ago (2001), ideas for a snail (*Biomphalaria glabrata*) genome project were discussed at the American Society of Parasitologists (ASP) meeting in New Mexico by various members of the snail community and subsequently, a snail genome consortium was established (see http://biology.unm.edu/biomphalaria-genome/). In 2002, a joint proposal put forward by the University of New Mexico (UNM), Albuquerque, NM and the Biomedical Research Institute (BRI), Rockville, MD to produce a snail BAC library was prioritized by the National Human Research Institute (NHGRI). This BAC library was made from the BB02 strain of *B. glabrata*, a recent field isolate from Brazil, that is susceptible to *Schistosoma mansoni*. The BAC library (136 kb average insert size, ~9.1x genome coverage) constructed at the Arizona Genomics Institute became available at cost starting in August 2004. Following the success of the BAC library proposal, a white paper for sequencing the snail genome was submitted to NHGRI in 2003 by investigators from BRI, UNM and The Institute for Genomic Research (TIGR, Hervé Tettlin). In August 2004, *B. glabrata* was prioritized as genomic sequencing target by NHGRI. In summer 2005, the Genome Sequencing Center (GSC) at Washington University (WashU), St. Louis, MO was assigned the sequencing of this snail genome. The details of this project can be obtained from the URL: http://genome.wustl.edu/genome.cgi?GENOME=Biomphalaria%20glabrata

Compared to the genome size of *S. mansoni* (270 Mb) the genome of *B. glabrata* is considerably larger (estimated to be around 931 Mb). The preliminary, investigatory sequencing (whole genome shotgun) of *B. glabrata* has been completed at WashU and 26,398 trace reads from this project have already been deposited in GenBank (see http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=6526) With these preliminary contributions, the time had come to convocate members of the snail genome consortium* to discuss the plan of action for sequencing the snail genome.

*The current snail genome consortium:
Coen Adema* - USA
Gennady Ataev- Russia
Thierry Backeljau - Belgium
Chris Bayne* - USA
David Blair - Aus
Mark Blaxter- UK
Daniel Boana - Ghana
Joanna Bridger* - UK
Paul Brindley* - USA
Omar Carvalho - Brazil
Christine Coustau* - France
Jason Curtis* - USA
Lawrence A. Curtis - USA
Reinhard Dallinger - AUSTRIA
Angus Davison - UK
Randy DeJong - USA
Marijke deJong-Brink - Holland
Colette Dissous - France
Georges Dussart* - UK
Aiden Emery* - UK
Najib El-Sayed* -USA
Sharon File-Emperador – USA
Petr Horak - Czech
Judith Humphries - USA
Renzo Nino Incani* - Venezuela
David Johnston - UK

Catherine Jones* - Scotland
José Jourdane* - France
Narcis Kabatereine - Uganda
Bernd Kalinna - Germany
Matty Knight* - USA
Thomas K. Kristensen*- Denmark
Hammou Laamrani* - Morocco
David Lambert - USA
Fred Lewis* - USA
Bernard Lieb* - Germany
Anne Lockyer – UK
Eric Loker* - USA
Nicholas Lwambo* -Tanzania
John Malone - USA
Juergen Markl - Germany
Andre Miller - USA
Dennis Minchella* - USA
Gerald Mkoji* - Kenya
Helene Mone - France
Les Noble* - UK
Guri Roesijadi* - USA
Nithya Raghavan*-USA
David Rollinson* - UK
Russell Stothard* -UK
John Sullivan* - USA
Hervé Tettelin* - USA
Andre Théron - France
Jackie Trigwell* - UK
Mingyi Xia* - China
Tim Yoshino* - USA
Don McManus*- Aus
Si-Ming Zhang* - USA
Ulrike Zelck*-Germany

*Indicated willingness to contribute experimental effort.

Thus, the first snail genome meeting, jointly sponsored under the auspices of the National Institutes of Allergy and Infectious Diseases (NIAID) and supplemented by the World Health Organization (WHO) was held on September 1-2, 2005 at Bethesda, MD in conjunction with the already established schistosome and filarial genome networks. The efforts of NIAID, WHO, and other organizers, especially Philip LoVerde the chair of this meeting who also made available funding for travel of the overseas participants and for the stay of all participants, are greatly acknowledged for making the snail genome meeting a success. Various representatives of international consortium "the *Biomphalaria glabrata* genome initiative" were present at this meeting. A separate session for the snail genome chaired by Matty Knight (BRI, Rockville, MD) was held on September 2, 2005.

Matty Knight initiated the session with a brief history of the snail genome project, restating the need for such a project. Genome sequences of the three organisms that are pertinent to transmission of schistosomiasis -the parasite, the intermediate snail host, and the human definitive host will be useful. She expressed the hope that such a sequencing project will help determine whether or not molecular co-evolution has allowed these two organisms (parasite and snail) to adapt to each other and to the human host. To date (September 19, 2005) there are only 9433 *B. glabrata* nucleotide sequences which includes

8307 ESTs. She reiterated the fact that the snail genome was very large (estimated at 931 Mb) and that it will require the co-operative efforts of the entire consortium to see the genome project to its fruition. She also indicated the availability of the SR3 (Schistosome Related Reagent Repository) at BRI, as a centralized facility for depositing, curating and distributing clones, libraries and other reagents to both the schistosome and snail genome community. The URL of this repository is http://www.afbr-bri.com/SR3/sr3home.htm

The list of speakers at the *B. glabrata* genome meeting:

1) E. Sam Loker (Univ. of New Mexico) - What is *Biomphalaria glabrata*?
2) Sandra Clifton (Washington Univ. St. Louis) - A preliminary sequencing analysis of the *B. glabrata* genome.
3) Nithya Raghavan (Biomedical Research Institute) - Preliminary genome sequencing and analysis of *B. glabrata* transposable elements.
4) Joanna Bridger (Brunel University, UK) - Chromosome and Nuclear Organization in the snail *B. glabrata*.
5) Coen Adema (Univ. of New Mexico) - *Biomphalaria glabrata*; initial insights from the mitochondrial genome and BAC-door access to the nuclear genome.
6) David Johnston (Natural History Museum, UK) - Transcriptome analysis in *B. glabrata*: a combined EST and microarray approach.
7) Les Noble (Univ. Aberdeen, UK) - 'Preliminary mapping of *B. glabrata*: a top down bottom up approach.
8) Judith Humphries (Univ. of Wisconsin-Madison) - Bge cells as a model for snail hemocyte gene discovery.
9) Chris Bayne (Oregon State Univ.) – Underlying mechanisms of schistosome-snail compatibility.

Also present were:
Adriana Costera (NIAID/NIH), Carolyn Cousin (University of the District of Columbia), Jason Curtis (Purdue North-Central), Randy DeJong (Malaria and Vector Research NIAID/NIH), Najib El-Sayed (TIGR), Adam Felsenfeld (NHGRI), Hirohisa Hirai (Kyoto University), Fred Lewis (BRI), Andre Miller (BRI), Mitzi Sereno (BRI), Liz Thiele (Purdue).

Sam Loker (UNM) gave an overview of the taxonomy, biology and evolutionary aspects of the snail *B. glabrata*. Of the 34 species of *Biomphalaria* identified, at least 18 were known or likely hosts of schistosomiasis. *Biomphalaria glabrata* is a complex organism with tolerance to adverse environmental conditions, capable of mounting diverse immune reactions to invading parasites. The species is worthy of study from multiple scientific viewpoints.

Sandra Clifton, Assistant Director of the Washington University Genome Sequencing Center, proposed plans of action for the snail genome project. She started of by indicating the desirability for this project of at least 12,500 ESTs by the end of year 2005. She also indicated that the current resources from the consortium included linkage maps being prepared using 25 microsatellite markers, ~75 RAPDs/AFLPs in addition to anticipating having the BACs placed for these markers. The BAC library generated from BB02 *B. glabrata* (sponsored by NHGRI) is was procured by the GSC. BACs identified as harboring transposons were being sequenced and finished and used to develop snail FISH. Within the consortium a first generation cDNA microarray and end sequences of some 180 BAC clones have been developed. Analysis performed at WashU using 2 animals indicated one heterozygous position every 4000-5000 bases (common repeats masked) in the genome of BB02 snails. From investigatory WGS sequencing, the genome was found to 64% AT-rich, and this certainly introduces the possibility of a cloning bias. In conclusion, she put forth a possible road map for sequencing the genome, constructing a genetic landscape and for curating and annotating the snail genome via a web interface. There was to be an initial organizational meeting at the GSC (October 6, 2005) with a few invited individuals from the consortium who contributed the white paper, following which, to design a final plan for approval by NHGRI. The outcome of this process will be communicated to the consortium.

Nithya Raghavan (BRI) presented an overview of the past research in snail genomics at BRI, and their identification of mobile genetic elements (MGEs). Noteworthy among the identified elements were the LINE-1-like retro-elements and reverse transcriptases. In addition, functional activity of this enzyme had been demonstrated in both resistant and susceptible snails and also in the snail cell line Bge. Nucleotide and protein analysis of the trace reads deposited by WashU in GenBank also revealed the presence of an abundance of MGEs. In summary, analysis of the WashU data indicated the presence of Class I non-LTR (3-4 %), Class II transposons (0.2%) in addition to LTR elements and SINEs. One of the LINE elements has been located on to a BAC that is currently being assembled and finished at TIGR. The same BAC has been used in FISH mapping of the snail chromosome for the first time, in collaboration with Joanna Bridger at Brunel University, UK. She concluded that some of the elements may still be active and the presence of a significant number of MGEs and repeat sequences may pose a potential problem in the genome assembly process.

Joanna Bridger from (Brunel Univ. UK) has experience in FISH mapping of human, pig, mouse, chicken, turkey, snail and trypanosomes. She also works on the genome organisation/chromosome behaviour in development, differentiation and disease. She is currently developing chromosome paints for the snail *B. glabrata* and is using the Bge cells as a platform for this process. She has successfully produced a chromosomal paint for one pair of snail chromosomes by micro-dissection and this will be expanded to remaining 17 chromosomes. She has also standardized the FISH technique for snail (Bge cell) chromosomes and has positioned one BAC (for the LINE element) on the chromosomes.

Coen Adema (UNM) presented on the characterization of the entire mitochondrial genome of the snail and on the generation of snail BACs in conjunction with the Arizona Genome Institute (AGI). He also maintains the *B. glabrata* genome initiative web site (http://biology.unm.edu/biomphalaria-genome/). The snail's mitochondrial genome is 13670 nt with13 protein-encoding genes, 22 tRNA genes, 2 rRNA genes (16S and 12S) with 74.6% A/T content. The identification of snail mitochondrial genes now benefits gene annotation efforts. The BAC initiative, while jointly proposed by UNM and BRI, was spearheaded by UNM. The *B. glabrata* snails used in this endeavor were collected by Omar dos Santos Carvalho from the field in the south east of Brazil (S 19°59'/W44°02'), Belo Horizonte, district of Barreiro, an endemic area for schistosomiasis. Molecular means confirmed species identity as *B. glabrata*. The F1 generation was established of the BB02 (Barreiro, Brasil, 2002) strain and this isolate will be used in the snail genome project also. The BAC library generated comprises of 61,824 BAC clones in 161 (384 well) plates, with an average insert size of 136 Kb thus providing a 9.1x genome coverage. The different targets recovered from this BAC library include FREPs 4 and 13, myoglobin, ferritin, actin, hsp 70 etc. and approximately 200k nucleotides of BAC ends have been sequenced.

David Johnston (Natural History Museum, UK) presented part of the work jointly done by the Natural History Museum (Anne Lockyer and David Rollinson), Aberdeen University, UK (Cathy Jones and Les Nobel) and Cambridge University, UK (Karl Hoffman and Jenny Fitzpatrick) on the development of *B. glabrata* microarrays. This group has generated a large number of snail ESTs using various techniques such as conventional methods, differential display, Open Reading frame ESTs (ORESTES) and Suppressive Subtractive Hybridization (SSH). The first generation microarray was printed in June 2004 using 1062 gene fragments from the ORESTES and 980 transcripts from the SSH libraries. In total, 4896 genes were printed in duplicate in a 18 x 18 genes/ subarray, 4 x 4 subarrays grid design with the controls consisting of vector (pGEM), spotting buffer and blanks. The next generation microarray will include many more sequences as contributed by several different international research groups.

Les Nobel (Aberdeen University, UK) continued on the work done by his group at Aberdeen, UK in the search for resistance markers in *B. glabrata* using expression analysis, linkage analysis and interval mapping. Analysis of the first generation Bg microarray using hemocyte RNA from resistant and susceptible snails identified 25 candidates with significant differential expression in 3 or > replicates (between 2-6 fold differences in expression). The resistant group identified 18 candidates (92% unknown); while the susceptible RNA identified 7 genes. The identified candidates from the resistant group included Titin, a signal transducer and ornithine decarboxylase 1, a catalytic enzyme with lyase activity. From their linkage analysis they have also established 16 preliminary linkage groups, and 3

microsatellites have been linked to resistance trait(s). In addition preliminary quantitative data to interval map resistance was also available.

Judith Humphries (Univ. of Wisconsin-Madison) presented the usefulness of Bge cells as a discovery tool for *B. glabrata* hemocytes. The two ongoing projects in the lab of Tim Yoshino and Judith Humphries include random EST library of Bge cells and a subtracted library before and after exposure to excretory-secretory proteins of *S. mansoni*. From their random library they have thus far generated 1625 sequences that fall into 235 contigs with 661 singletons. Some of the identified homologs from the random and subtracted libraries include the Y-box protein, cytochrome b, perforin, peroxiredoxin, serine protease inhibitor etc.

Chris Bayne (Oregon State University) talked on the functional genomic aspects of the project. His interest was in identifying snail genes encoding enzymes involved in the respiratory burst, and their expression Superoxide dimustase 1 (SOD1) is a Cu/Zn dependent enzyme that is expressed cytoplasmically. It has 5 exons and 4 introns ranging in sizes from 75-118 nt and 119-2500 nt respectively. He felt that identifying specific genes and defining their intron-exon boundaries and figuring out a functional pathway was very necessary for the success of the snail genome project.

At the end of the session there was an hour-long discussion on the various aspects of the snail genome project. One of the primary concerns was the setting up of snail genome database and its long-term maintenance. Increasing the number of ESTs in the database was also defined as an important prerequisite for the success of this project. The lack of an adequate number of full-length snail cDNAs was discussed as a potential stumbling block for training "gene-finder" programs. It was decided that while the snail genome sequence is being characterized, individual groups could fish out, analyze and publish on their favorite genes. The snail genome in it entirety will only be published with the involvement of all the participating groups and the sequencing center.

For database and or webinterface facilities needed to manage, annotate and make available sequence data, various members would explore possibilities within and outside of the consortium. This topic will be decided on later.

Earlier during the meeting, Matthew Berriman (Welcome Trust Sanger Institute) had indicated that the Sanger Center was interested in contributing large scale (~100,000) EST sequencing to the *B. glabrata* genome effort (pending a proposal from consortium members). With this added consideration, the consortium members present were in favor of initial 4x coverage by WGS sequencing of the genome of *B. glabrata*. Then an assembly effort will be made to decide how to proceed with the project. Other resources will be considered for the sequencing project also. It was understood that inherent properties may make it impossible to arrive at a final assembly of 18 contiguous clusters. More likely, the final genome sequence will contain gaps.

The topics discussed will be developed in a smaller setting (WUGSC and invited members of the consortium; 6 October 2005). A resulting final proposal will be drafted for approval by NHGRI.

Matty Knight summarized the discussion as shown below.

Summary of *B. glabrata* genome project (Sept. 1-2, 2005)
- Whole genome shotgun - at least 4x coverage
- Make fosmids
- Finger-print BACs - do pilot run
- Request Sanger to do 100,000 ESTs - whole snail
- Full-length cDNA project
- Database concern - discussions
- Publication from the consortium
- Group meeting at Wash U with at least members of the white paper proposal