Re: Proposal for Construction of a BAC library from the tsetse fly *Glossina morsitans morsitans*

Date: June 10, 2003

From: Serap Aksoy, Yale University, New Haven, CT, USA
Terry Pearson, University of Victoria, Canada
Michael J. Lehane, University of Wales, Bangor, U.K.
Wendy Gibson, University of Bristol, Bristol, UK

To: BAC Library Resource Network, National Human Genome Research Institute

Importance of the organism to biomedical or biological research. Tsetse flies are the sole vector of trypanosomiasis. They transmit the African trypanosomes, which cause the fatal disease Sleeping Sickness in humans and various wasting and fatal diseases in domestic animals. Currently, human epidemics still rage throughout Central Africa and the animal diseases are still the major hindrance for the development of agricultural practices in Africa. In addition to their biomedical significance, tsetse flies are of interest as a member of higher Diptera. Both male and female adults are strict blood feeders, depend on obligate symbiotic microbes for many vital functions and display viviparous reproductive biology, all characteristics that make this insect a great model to address unique physiological questions.

Uses to which the BAC library would be put in addition to genomic sequencing. The clones of this library (accompanied with high-density hybridization filters) would provide a vital resource for the characterization of genetic loci important for the vectorial competence of tsetse. Promoter regions of genes with inducible expression profiles and especially those genes important in immune competence would be among candidates to isolate for the development of systems towards engineering parasite refractory flies. These genes can be subsequently used in transgenic-based approaches both in tsetse and in the related insect Drosophila for further functional analysis.

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. Work on tsetse flies as an important disease vector is expanding. Because of the laborious colony rearing requirements, the research circle that works solely on tsetse has been relatively small. However, as DNA technologies and genomics based research has expanded and made research less dependant on live colony material, investigators who work on this disease system are increasing. Currently there are about fifteen research groups worldwide who would immediately benefit from the availability of the BAC library. The availability of this BAC library resource will increase the involvement of investigators with interests in comparative genetics, genomics and evolutionary biology.

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. The

goal of the tsetse fly community is to proceed to genome-sequencing project, which will be required for post-genome functional genomics-proteomics approaches to tsetse and parasite control. The initial proposal towards a genome sequence will focus on the BAC end sequences. The recently completed genome sequences of the parasite *Trypanosoma brucei* and the human host and to the sequencing projects being planned for *Trypanosoma vivax* and *Trypanosoma congolense* makes the sequencing of the tsetse genome timely. The sequencing of the tsetse genome will also have broad support from the fly community as *Glossinidiae* occupies a key phylogenetic position in Diptera.

Other genomic resources that are available that will complement this resource.
A midgut tissue specific EST sequencing project from *G. m. morsitans* has now been completed by the Wellcome Trust Sanger Institute, UK (http://www.sanger.ac.uk/Projects/G_morsitans). In this analysis, 21, 427 expressed sequence tags were produced from *G. m. morsitans* adult midgut and grouped into 8,876 clusters or singletons potentially representing unique genes. Putative functions are ascribed to 4,035 of these by homology. A similarly immune challenged fat body specific EST library sequencing project is *in progress* at The Institute for Genomic Research (TIGR) and funds have been secured through the Wellcome Trust for the sequencing of a salivary gland EST library (MJ Lehane, University of Wales, Bangor). These projects are done with normalized libraries prepared at the B. Soares laboratory in Iowa and showed high quality cDNAs. In addition, a cDNA library from the cardia tissue is currently under analysis. A grant application for micro array construction of all the sequenced ESTs is in progress (MJ Lehane) and will allow for the development of functional studies.

The strain of the organism proposed and rationale for its selection. We are proposing to construct the BAC library *Glossina m. morsitans* which is the species from which all of our ESTs have been prepared. This species is an important vector in Africa for both human and animal diseases.

The size of the genome. Currently, two different approaches are being used to determine the size of the tsetse genome; Cot analysis and DNA estimation based on Epics Elite flow cytometer propidium iodide stained nuclei extracted from sample brain. Sequences of several genomic loci show lack of extensive intervening sequences and furthermore cDNA sequences analyzed have short 3'-non-translated regions indicating that the genome size may be similar to the other drosophilatids.

The availability of a source of DNA for construction of the BAC library. We maintain breeding colonies of *Glossina m. morsitans*. DNA can be easily prepared for library construction.

Specifications for the library and supporting scientific rationale for these specifications.

We prefer to have two libraries with 6-8x coverage but constructed using different restriction digests. A large insert size, minimum of 100 is desirable for eventual complete genome sequencing project. This library can also be used for the preparation of high-density hybridization filters for screening with genes of interest, mapping and gap closure.

The time frame in which the library is needed. This library will provide the impetus for our functional genetic studies and provide a drive for our genome sequencing funding efforts. The library is needed now.

Other support. There is no other support for construction of this library.

Need for an additional library. Irrelevant at this time point.