Request for approval for sequencing of the sooty mangabey (*Cercocebus atys*) genome

Despite more than 20 years of effort and the investment of substantial resources, the biomedical research community has not yet been able to develop an effective strategy to prevent infection with HIV. The unusual properties of this virus have confounded all efforts develop a vaccine, and as a result more than 2.5 million new infections occur worldwide each year. Recent high-profile failures of vaccine trials have caused the HIV-AIDS research community to re-think strategies and initiate new avenues of research. One of those avenues is expanded investigation of nonhuman primate species that have evolved the ability to successfully cope with persistent infection with HIV-related lentiviruses, without progressing to AIDS-like disease. These natural host species, all of which are native to Africa, include the sooty mangabey (*Cercocebus atys*), the African green monkey and the mandrill. These species exhibit high rates of natural infection in the wild with various simian immunodeficiency viruses (SIV). When Asian rhesus or cynomolgus macaques are infected with SIV obtained from sooty mangabeys, these macaques develop a variety of symptoms that are remarkably similar to humans infected with HIV, and generally succumb to an AIDS-like disease in a few months. While researchers have begun to explore the mechanisms by which sooty mangabeys and other natural primate hosts cope with high SIV viral loads and remain free of disease, the details of this resistance are still poorly understood. Access to whole genome DNA sequences for natural hosts, and the resulting ability to compare genes and genetic pathways among natural hosts and non-natural hosts (i.e. susceptible species such as macaques) would accelerate research in this area. For these reasons, sooty mangabeys were one of the species recently proposed to NHGRI for whole genome sequencing. The Co-ordinating Committee was supportive of the general goal of sequencing this and other nonhuman primate genomes, but declined to approve sooty mangabeys at this time. Instead the CC requested additional information and recommended developing a plan and justification for larger number of primate species, more than the 15 species proposed via White Paper in early 2010.

The Baylor Human Genome Sequencing Center has been approached by external investigators who wish to pursue the sequencing of the mangabey genome now, and have substantial resources to support the project. Dr. Guido Silvestri (currently at University of Pennsylvania but soon moving to Emory University) is the leader of this group. Dr. Silvestri is willing to support the cost of generating approximately 80X sequence coverage of the sooty mangabey genome, using Illumina fragment paired-end and large insert mate-pair libraries. Dr. Silvestri is also interested in sequencing RNA from a series of mangabey tissues related to immune function and SIV. HGSC requests approval to match Dr. Silvestri's investment and cover the cost of sequence analysis and genome assembly. We propose to charge Dr. Silvestri for all costs associated with producing 80X coverage of the mangabey genome, and deep RNA sequencing on four tissues. HGSC would use available resources to pay one full-time bioinformatics analyst to work 100% effort for one year to manage the data, conduct quality control tests, produce the whole genome assembly of the mangabey genome, and assemble full-length transcripts to generate well-validated gene models (see attached budget).

Drs. Kim Worley and Jeff Rogers would supervise these activities, under current HGSC salary support.

We believe that sooty mangabeys are an outstanding candidate for immediate sequencing, given their relevance to critical questions in HIV-AIDS research and the opportunity to leverage HGSC resources with significant support from the external primate research community. First, the virus that occurs naturally in sooty mangabeys is the origin of both SIVmac (the virus most commonly used in studies of the macaque model of AIDS) and HIV-2 (responsible for millions of infections in humans). Second, the only colony of naturally SIV-infected nonhuman primates available for research in the U.S. is the large colony of SIVsmm-infected sooty mangabeys housed at the Yerkes Primate Research Center. Lastly, the CD4 cell biology of sooty mangabeys is more similar to that of humans or macaques than is the CD4 cell biology of the other widely available natural SIV host, African green monkeys. Typically, sooty mangabeys are similar to humans and macaques in their circulating levels of CD4+ T cells, comprising 30-50% of all lymphocytes. In contrast, African green monkeys show lower levels of circulating and mucosal CD4+ T cells, and higher fraction of double positive CD4+CD8+ T cells. For this reason it can be argued that mangabeys represent a unique model of the effects of SIV infection on CD4+ T cells. Overall, natural infection of sooty mangabeys with SIV represents a highly relevant and intensively studied model of lentivirus infection. The possibility of integrating the available virological and immunological information on this species with a complete knowledge of the mangabey genome represents an invaluable resource and opportunity for the national (and indeed global) AIDS research effort.

Budget

Sequencing costs to be paid by Dr. Silvestri (Emory Univ.)

40x genome coverage, Illumina fragment sequencing (2 PE libraries and 10 HiSeq lanes)	$ 34,570.
40x genome coverage, Illumina 3kb mate-pairs (5 mate-pair 3kb libraries and 11 HiSeq lanes)	43,693.
1x sequence coverage, 454 20kb mate-pairs (1 454 20kb PE library and 1 454 run)	7,967.
RNA sequencing, per tissue (1 454 fragment library and ½ 454 run per tissue)	3,670. (per sample)
Total cost of sequencing (assumes 4 tissues for RNA-seq)	$ 100,910.

Bioinformatic analysis and genome assembly (to be covered by BCM-HGSC)

Bioinformatics support (2000 hours)	$ 110,000.

Total Est. Cost of Sooty Mangabey Genome Project $ 210,910.