Gibbon (*Nomascus leucogenys)* Sequencing Proposal

Rationale: The two major reasons for complete sequencing of the gibbon genome are 1) its unique phylogenetic position on the primate tree and 2) the unusual frequency of chromosome rearrangement events since its divergence from the last common ancestor. Similar to other non-human primate index species (Figure 1), high quality genome sequence provides a unique opportunity to reconstruct the ancestral primate genome athe basepair level. We expand on these points below.

1) **Reconstruction of the Ancestral Primate Genome**: As a representative of the lesser-apes, this species is the only phylogenetic branch point of non-human primates for which a whole genome shotgun sequencing project has yet to be approved. The lesser apes are the link between humans/great apes and the Old World monkey phylogeny. This position provides a unique view of evolutionary divergence from the human genome over 18-20 million years of species separation (Goodman 1999; Muller et al. 2003). Sequencing of the gibbon is critical to achieve comprehensive annotation of the human genome by reconstruction of the ancestral genome at the basepair level. Complete sequencing of this genome will allow the evolutionary history of mutational change to be resolved since the divergence of common ancestor humans and great apes (14 million years ago) from the Old World monkey lineage (23-25 million years ago). Genome-wide sequence representation of this phylogenetic node is important to interpreting the history of functional genetic differences that have emerged. Similar to other non-human primate genome projects, the focus on sequence differences requires high-quality whole-genome shotgun sequence (including resolution of regions of segmental duplication (SD), and chromosomal rearrangement). This is the primary motivation for sequencing the gibbon genome.

2) **Mechanism of Chromosomal Rearrangement:** Unlike most other mammalian species, the gibbon karyotypes have been subjected to an extraordinary number of rearrangement events. (Muller et al. 1997; Muller and Wienberg 2001). Comparative studies indicate an unusually large number (n>45) of chromosomal fissions and translocations when compared to other primates. It is quite likely that an exponential increase in the number of rearrangements will be revealed as smaller structural changes between the genomes of humans and gibbon are refined. BAC-end sequencing efforts have already been approved and the framework for some of these differences is becoming apparent. Whole-genome shotgun sequencing of the entire genome would be used to further elucidate the molecular basis for chromosomal rearrangements—particularly the frequency of smaller inversions and intrachromosomal events. Sequencing of BACs will provide the necessary precision and quality in regions of genomic complexity. Information obtained from such studies could provide valuable insight into the mechanism underlying both germline and somatic chromosomal instability associated with human disease and evolution.

Species Selection: There are at least five different species or subspecies belonging to the family Hylobatidae (two main genera Syndactylus and Hylobates). All show rapid evolutionary rearrangement with respect to human and other Old world monkey lineages. As an index species with respect to human genome annotation, all species are comparable. Material from *Nomascus leucogenys* (female white-cheeked gibbon) has recently been used in the construction of the first gibbon BAC library (CHORI-271)(material was kindly provided by Dr. Alan Mootnick at the Santa Barbara Zoo, CA, Director of Gibbon Conservation). The 0.1X fold sequence coverage in BAC-end sequences has been generated from this source. Since the donor is female and additional material may be obtained for plasmid and fosmid library construction, this represents the most pragmatic choice.

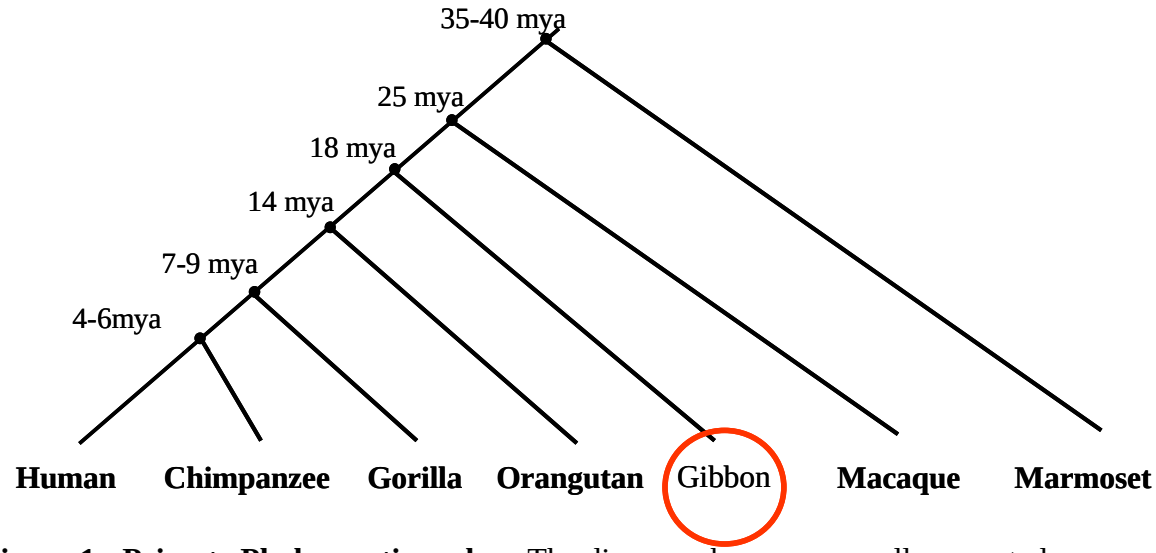

Figure 1: Primate Phylogenetic nodes: The diagram shows a generally accepted phylogenetic tree with divergence times with respect to the human lineage (Goodman et al., 1999). Index species for which full-genome sequencing has been approved are shown in bold. There are six phylogenetic nodes indicated. Each provides a unique opportunity to reconstruct the primate ancestral genome at the basepair level at these points during lineage. The gibbon lineage is the only branch for which whole-genome shotgun sequencing has not yet been approved.

Proposal:

Two years ago, we proposed (and council approved) BAC end-sequencing (BES) of the entire gibbon BAC library (180,000 clones). This limited survey was requested (0.1X sequence coverage) to minimize demand on sequence production at that time, but would still provide an overview of sites of large-scale chromosomal rearrangement in anticipation of a more extensive draft sequencing project (6X). With the BES now complete (150,000 clones corresponding to ~300,000 BES are now available through the trace repository as of May 1), we now propose a complete whole genome shotgun sequencing project with complementary high-quality BAC sequencing within regions of complex rearrangement and segmental duplication.

•**WGS component**: We propose a 6-fold WGS sequencing phase. Long, high-quality sequence reads are required to accurately assemble this repeat-rich genome. In theory, six-fold sequence coverage will provide 99.3% coverage of the lesser-ape euchromatic sequence for genome assembly. Comparisons between the finished genome and WGS assembly for human have shown that strict WGS of 5.3X paired-end human sequence will allow 92.7% of the sequence to be accurately assembled and mapped. An additional 3.6% could be readily recovered by primer-end walking. The remainder is expected to map to segmental duplication (if these are equivalent between the species). These data suggest that the bulk of the euchromatin will be reliably assembled at this depth. Furthermore, this depth is sufficient to detect (but not assemble) 99% of segmental duplication regions (>95% identity and >20 kb) (Bailey 2002, Bailey 2004, Cheng, 2005).

•**BAC-based component.** To complement the WGS assembly, we propose shotgun sequencing of targeted BAC clones. As proposed for other index non-human primate projects, we will select BACs from sites of rearrangement, segmental duplication, centromere and telomere transition regions as identified by BES data. In particular we will look for BACs with BES anchored at one end within unique portions of the WGSA but with the other end in duplicated regions. Alleles and paralogues from such regions will be distinguished based on an initial, light draft data of representative clones. From this set ~1000-1500 BACs will be selected for higher quality large-insert sequencing. These finished or near-finished inserts will be preferentially inserted into the WGS assembly to produce a high quality genome sequence.

Concluding Remarks: The seven distinct phylogenetic branchpoints between human and non-human primate species provide considerable power for annotation of the human genome, with each branch point revealing significant changes. Multiple species comparisons are expected to be significantly more informative than individual pairwise comparisons that have been performed to date (CSAC 2005). Complete (~6 X) genome sequencing projects have been approved for five of these seven nodes, represented by chimpanzee, orangutan, macaque and marmoset and most recently gorilla (approved by the Wellcome trust, October 2005; see Table 1 (see below), Figure 1). Low coverage (~2X) has been approved for the prosimians, mouse lemur and galago, representing the sixth branch point. The hylobatids, the seventh branch point, are now the only index point for which significant genome-sequence data does not exist. Sequencing of the gibbon genome will provide the necessary information to reconstruct the ancestral primate genome at this critical junction of human evolution. With the other primate sequences we will have a comprehensive view of the evolution of our genome.

Specific Application:

- **Gene annotation**— Less than 50% of human genes can be mapped as 1:1 orthologs between man and mouse (Clark et al. 2003). High quality sequence from non-human primate genomes such as gibbon will further improve 1:1 annotation(Olson and Varki 2003). High quality sequence from gibbon will provide temporal information during this 10 million year window and help delineate factors (gene conversion, duplication, rearrangement) that confound 1:1

mapping of genes and other conserved sequences. In addition, genome sequence from this species will provide an outgroup to humans and great apes and identify non-neutral patterns of selection (adaptive and balancing) that emerged after separation from the Old-World monkeys.

- **Segmental Duplication and Structural Variation** . ~5% of the genome exists as blocks of duplication (>95% >20 kb in length). In particular, many of the intrachromosomal duplications that contribute to disease and disease susceptibility appear to have expanded recently in the African great-ape and human lineage. Segmental duplications are preferential sites of breakpoints of synteny (Armengol et al. 2003; Bailey et al. 2004; Murphy et al. 2005) between man and mouse and man and apes—between 30-50% of breakpoints in conserved synteny between mouse and human and ~60% of the breakpoints in conserved synteny between human and African apes map to regions of segmental duplication. Whole-genome sequencing of the gibbon and delineation of the breakpoints will provide a unique opportunity to determine whether SDs have been directly responsible for the accelerated rate of karyotype evolution in this species.
- **Mechanism of Chromosomal Evolution**: The tempo of karyotype evolution (defined here as the number of syntenic rearrangements) is 10X faster among hylobatids. What is the underlying genomic mechanism? The relatedness of hylobatid genome sequence (~5% divergence) and the high quality of the human genome reference provide the ability to interrogate the junctions precisely. This precise knowledge should provide insights into the basic molecular mechanisms driving these changes.

Table 1: Primate Whole-genome Sequencing Status

Species	Target	Center	Progress	Expected completion	WGS Reads (11/15/05)*	WGS Reads (1/02/06)*	WGS Reads (4/15/06)
Pan troglodytes (Common chimpanzee)	high quality draft assembly and refinement	WIBR, WUGSC	6X WGS assembly available; in finishing	WGA and analysis done; BAC finishing mid- 2006	31,337,683	31,338,467	31,339,206
Macaca mulatta (Rhesus macaque)	High quality draft assembly (7X)	BCM, WUGSC	5-6X WGS assembly available	WGA completed; analysis and targeted phase	22,853,623	22,853,623	22,853,623
Pongo pygmaeus (orangutan)	High quality draft assembly (6X); prefinishing; finish ENCODE regions; 100k ESTs	WUGSC, BCM	WGS collection nearing completion	WGA mid 2006	24,503,869	25,513,505	26,606,602
Callithrix jacchus (common marmoset)	High quality draft assembly (6X); prefinishing; finish ENCODE regions; 400k ESTs	BCM, WUGSC	WGS collection 1/2 complete	WGA late 2006	6,497,052	16,213,747	25,901,303
Otolemur garnetti (galago or bushbaby)	2X coverage		underway	2X WGS February 2006	242,592	5,029,194	7,632,316
Nomascus leucogenys (white-cheeked gibbon*)*	BAC-end sequencing (180k BACs)	WUGSC	WUGSC awaiting BAC library from DeJong	?	-	-	231,565
Gorilla gorilla (gorilla)	High quality draft assembly; likely at least some finishing	Sanger	**Sanger Institute**	early 2007?	14,812	14,812	136,607

References.

Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X (2003) Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum Mol Genet 12:2201-8

Bailey JA, Baertsch R, Kent WJ, Haussler D, Eichler EE (2004) Hotspots of mammalian chromosomal evolution. Genome Biol 5:R23

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M (2003) Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960-3

CSAC (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87

Goodman M (1999) The genomic record of Humankind's evolutionary roots. Am J Hum Genet 64:31-9

Muller S, Hollatz M, Wienberg J (2003) Chromosomal phylogeny and evolution of gibbons (Hylobatidae). Hum Genet 113:493-501

Muller S, O'Brien PC, Ferguson-Smith MA, Wienberg J (1997) A novel source of highly specific chromosome painting probes for human karyotype analysis derived from primate homologues. Hum Genet 101:149-53

Muller S, Wienberg J (2001) "Bar-coding" primate chromosomes: molecular cytogenetic screening for the ancestral hominoid karyotype. Hum Genet 109:85-94.

Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Auvil L, Beever JE, Chowdhary BP, Galibert F, Gatzke L, Hitte C, Meyers SN, Milan D, Ostrander EA, Pape G, Parker HG, Raudsepp T, Rogatcheva MB, Schook LB, Skow LC, Welge M, Womack JE, O'Brien S J, Pevzner PA, Lewin HA (2005)

Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science 309:613-7
Olson MV, Varki A (2003) Sequencing the chimpanzee genome: insights into human evolution and disease. Nat Rev Genet 4:20-8