Gibbon Draft Sequencing Proposal

Annotating the Human Genome Working Group
April 28, 2005

Rationale: Two major reasons for the sequencing of the gibbon are:

1) First, as a member of the lesser apes, this species represents the only major branch of non-human primates for which no whole genome shotgun sequence has yet been approved. The lesser apes represent the link between humans/great apes and the Old World monkey species. Its sequence provides a unique view of evolutionary divergence from the human genome over 18-20 my of species separation (Goodman et al., 1999, Muller et al., 2003).
2) Second, gibbons demonstrate one of the most rapid rates of karyotype evolution compared to other primates and most mammals (Muller et al. 1997; Muller and Weinberg 2001). Comparative studies indicate an unusually large number (n>45) of chromosomal rearrangements when compared to hominoid species. Unlike most other mammalian species, the gibbon karyotypes have been subjected to an extraordinary number of rearrangement events. Comparative sequencing would be used to understand the molecular basis for chromosomal rearrangements—i.e. the transition regions and sequences that may have predisposed the gibbon to such events. Information obtained from such studies could provide valuable insight into both germline and somatic chromosomal instability associated with chromosomal rearrangement.

Proposal:

In light of current demands on sequencing capacity, we propose a two-stage approach to sequencing the gibbon. The first stage would consist of end-sequencing the entire gibbon BAC library (180,000 clones) and should begin now. This would place minimal demand on current sequence production (0.1X sequence coverage), but provide extraordinary physical coverage (10-fold). The second stage would begin after completion of the BAC framework, and would consist of a whole genome shotgun assembly at 6X coverage. This initial light draft of end sequence mapped against the human genome would serve two purposes:

1) It would provide an overview physical map of the gibbon genome prior to full plasmid draft sequencing. This is especially important in the case of the gibbon, where many rearrangement breakpoints are anticipated. The unambiguous assignment of gibbon chromosomal groups should proceed well in advance of a genome assembly exercise. A BAC scaffold would provide the substrate for FISH assignment to confirm relationships detected by BAC end-sequence.

2) Paired-end sequence mapping against the human genome would simultaneously clone all regions of large-scale rearrangements within large-insert BAC clones. Preliminary analyses of several great ape genomes has indicated that >50% of all chromosomal rearrangement sites map to regions of segmental duplication greater than 100 kb in size. Unlike fosmids, BACs provide the requisite insert size to allow these breakpoints to be readily traversed. Such regions are unlikely to be resolved by WGS plasmid-based sequencing.

Other Considerations: 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. Material from *Nomascus leucogenys* (female white-cheeked gibbon) has recently been used in the construction of the first gibbon BAC library (CHORI-271). Material for this library was kindly provided by Dr. Alan Mootnick, Director of Gibbon Conservation at the Santa Barbara Zoo, CA. Since the BAC library was generated from a female donor and additional material may be obtained for plasmid and fosmid library construction, initial BAC-end sequence data should be generated from this library.