Additional BAC Coverage of the Male Chimpanzee represented by CHORI-251: A Request for BAC Library Construction Through the BAC Resource Network

Submitted to the NHGRI, summer 2003

Submitted by David C. Page
page_admin@wi.mit.edu
Voice: 617 258 5203
Fax: 617 258 5578

Overview. This is a request for an additional BAC library providing 10X autosomal coverage for the male chimpanzee represented by the CHORI-251 library (*Pan troglodytes*, Yerkes animal #C0471). Combined with the existing RPCI-43 (also representing Yerkes #C0471) and CHORI-251 libraries, the requested library will provide sufficient coverage of the chimpanzee Y chromosome to allow its essentially complete sequencing, which would not be possible otherwise.

Preliminary data indicate that, in contrast to the autosomes and the X chromosome, the chimpanzee and human Y chromosomes are substantially diverged. This divergence is not primarily in the form of dispersed single nucleotide differences. Rather it is in the form of abundant rearrangements at small and large scales, including substantial sequence in the chimpanzee Y chromosome that is lacking in humans. Therefore, the sequence content of gaps in the chimpanzee BAC map cannot be inferred from the human Y sequence, and the order and orientation of chimpanzee BAC contigs cannot be based on the human Y sequence. Instead, the BAC map of the chimpanzee Y chromosome must be constructed and confirmed based on overlaps between chimpanzee BACs. Experience with the human Y chromosome has shown that in some regions, these overlaps must be > 30kb, in order to distinguish BACs representing different copies of large (> 200 kb) and nearly identical repeats. This allows differences in apparent overlaps to be recognized as differences between repeat copies rather than differences between individuals. Therefore, it is important that the chimpanzee represented in the requested library be Yerkes #C0471. In other, non-duplicated regions, higher coverage will allow selection of BACs for sequencing that avoid unnecessarily large overlaps. Such unnecessarily large overlaps would be costly to sequence because full depth of subclone shotgun reads would be generated for both overlapping BACs.

In the remainder of this request we address the issues enumerated in the NHGRI's "Instructions for Proposing Organisms From Which to Make New BAC Libraries."

1. The importance of the organism to biomedical or biological research. The chimpanzee is listed as a high priority sequencing project by the NHGRI. As discussed above, there are many rearrangements between the human and chimpanzee Y chromosomes. These necessitate that the chimpanzee Y chromosome be sequenced largely independently, using the iterative mapping sequencing strategy we and our colleagues used to finish the human Y chromosome.

2. Uses to which the BAC library would be put, in addition to genomic sequencing. Genomic sequencing would be the primary use for this library.

3. 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. The additional male chimpanzee will enable completion of NHGRI R0-HG00257-10. The aims of this grant include providing a sequence ready map of the chimpanzee Y chromosome, and with clones sequenced at the Washington University Genome sequencing Center in St. Louis ("WashU"), the finished sequence of the chimpanzee Y chromosome. By enabling completion of the sequence of the chimpanzee Y chromosome, this library will ultimately be of great value to those studying genes on the chimpanzee Y chromosome, some of which are likely to play critical roles in spermatogenesis. The sequence will also be valuable to those investigating the processes of sex-chromosome evolution.

4. 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 sequence of the chimpanzee is listed as a "high priority" sequencing project by the NHGRI, and the NHGRI has approved and funded the finished sequencing of the chimpanzee Y chromosome. The requested library will enable attainment of this objective and cost-effective clone selection and gap closure.

5. Other genomic resources that are available that will complement this resource. The value of the requested library will be increased by the existing, partial sequence of the Y chromosome of chimpanzee Yerkes #C0471 (BACs from the RPCI-43 and CHORI-251 libraries).

6. The strain of the organism proposed and rationale for its selection. We request that the library be constructed from the same individual (Yerkes #C0471) that is represented in the RPCI-43 and CHORI-251 libraries.

7. The size of the genome. ~3 Gb.

8. The availability of a source of DNA for construction of the BAC library. DNA extracted from Yerkes #C0471 for construction of the CHORI-251 library is still available.

9. Specifications for the library (e.g., library depth, BAC insert size) and supporting scientific rationale for these specifications. 10X autosomal coverage in the requested library would provide an additional 5X BAC coverage of the chimpanzee Y chromosome. Taken together with the existing RPCI-43 and CHORI-251 libraries, the requested library would provide a total of 12X coverage of the chimpanzee Y chromosome. Insert size of ~170 kb, as for the CHORI-251 library, would be suitable. Availability of high-density hybridization filters for BAC identification would be essential. To provide better probability of filling clone gaps, an enzyme other than *Eco*RI (the one used in constructing the RPCI-43 and CHORI-251 libraries) should be used.

10. The time frame in which the library is needed. As soon as possible, to allow timely completion of the project (R01-H00257) goals of mapping, sequencing, and analyzing the chimpanzee Y chromosome.

11. Other support that is available or has been requested for the construction of the desired library. No other support is available or has been requested for construction of an additional BAC library representing the genome of a male chimpanzee.

12. The need for an additional BAC library if one or more already exists. The existing publicly available male chimpanzee libraries, RPCI-43 and CHORI-251, represent the Y chromosome at approximately 7X coverage. Our experience mapping and (with WashU) sequencing the human Y chromosome indicates that at this low level of coverage there will be an unacceptably high number of gaps (~20) in the BAC map of the chimpanzee Y chromosome. As discussed above, it is not feasible to use the human Y chromosome sequence to infer the contents of these gaps or to order and orient isolated BAC contigs.. We anticipate that many of these gaps will be spanned by BACs in the requested library.

In addition, at only 7X coverage we would be forced to select for sequencing many clones with needlessly large overlaps, simply because there would be few BAC choices in many areas. This would increase the cost of sequencing the chimpanzee Y chromosome. Even though such redundant overlaps will not have to be finished in both overlapping BACs, in many cases both BACs will have to be fully shotgun-sequenced at substantial expense.

13. Other relevant information. None.