Pristionchus pacificus var. California Sequencing Plan

Washington University Genome Sequencing Center

Pristionchus pacificus is a free-living nematode belonging to the family Diplogastridae. The working wild-type strain of *P. pacificus* was isolated in Pasadena, California and is designated as PS312. This is the strain that will be sequenced. Please note that planning for this genome project involved consultation with Ralf Sommer (Max-Planck Institut fur Entwicklungsbiologie, Tuebingen Germany).

In sequencing the genome of *Caenorhabditis briggsae* we learned that even with 8X whole genome shotgun (WGS) coverage, local repeats in that nematode genome confounded even the best assembly algorithms. However, for the nematode *P. pacificus*, we are fortunate to have additional genomic resources that should significantly facilitate sequence assembly. Most notably, two BAC libraries exist; one constructed using *Hind*III and the other using *Eco*RI. A total of 34,187 BAC clones already have been end sequenced (14,812 from the *Eco*RI library and 19.375 from the *Hind*III library). This yields approximately 0.16X sequence coverage in BAC end sequences (BES). All of these sequences are available from the dbGSS division of GenBank. A *P. pacificus* fosmid library also is available and 46,130 clones have been sequenced at one end (~0.25x sequence coverage). These data also are available from the dbGSS division of GenBank. Gridded filters for both the BAC and the fosmid clones have been prepared and are available upon request from the RZPD German Resource Center for Genomic Research, Berlin. As part of our nematode cDNA sequencing program (http://nematode.net/), the GSC has 11,047 ESTs from mixed stage *Pristionchus* animals and submitted these data to the dbEST division of GenBank. Funds for this work were provided by the Max-Plank Institute. We also have prepared RNA from larval stages L1, L3, and L4 for which we are constructing cDNA libraries for further EST sequencing. In addition to these resources, the research community has provided a genetic map (six linkage groups), a physical map (now at 396 contigs) and 547 STS markers.

After consideration of the available resources, we propose to: 1) sequence the genome of *P. pacificus* to 6.5-fold coverage in plasmids, 2) sequence the missing ends of the available fosmid clones, and 3) perform one round of directed sequence improvement ("pre-finishing"). Genomic DNA will be prepared from synchronized L1 stages in the laboratory of Ralf Sommer and sent to the WUGSC. At our current average read length (~690 Q20 bp), we expect to attempt just over 1 million reads to achieve the desired coverage level. The cost for WGS library construction ($1,200.00) and sequence production (~$0.60 per read) will be approximately $620K. Fosmid sequencing will cost approximately $35K.

Once libraries are available, sequence production should require no more than one month. After we have assembled all of the WGS data, we will utilize our pre-finishing pipeline to improve the quality and contiguity of the genome sequence. This is an automated production-style pipeline that generates directed sequence reads to an initial whole genome shotgun assembly. Here, each assembly is analyzed using the program *autofinish*, and oligonucleotide primers are selected to extend reads through low quality regions or off the ends of contigs. The program also generates lists of templates that are robotically re-arrayed to match 96-well plates of primers. All primer-directed sequencing reactions are subsequently processed through our

main production pipeline. We propose to perform one round of pre-finishing on the *P. pacificus* genome, at an estimated cost of $1M (~$0.01/base). At this point, an option that could be considered to close many of the remaining gaps would be to perform low-coverage shotgun sequencing of BAC clones that span those gaps (i.e., as determined by BAC end sequence placement). The cost of this option would be $200-470K, depending on the type of DNA purification method available.

Following final genome assembly, the resulting sequence contigs and supercontigs will be made available through GenBank and AppaDB (an AcedB based database housed in Tuebingen, Germany [http://appadb.eb.tuebingen.mpg.de]). We will work with AppaDB and members of the Pristionchus community to determine the best approach to subsequent analysis and annotation of the sequences. Analysis and annotation of the *Pristionchus* genome will be accomplished in a manner similar to our work on the *C. elegans* genome, with Dr. John Spieth, the leader of our WormBase efforts, playing a lead role. We will work with *Pristionchus* researchers such as Dr. Rolf Sommer and others to ensure that the resulting annotation is consistent with the requirements of their research community. If sufficient resources are available within the GSC, we would be interested in contributing substantial effort and expertise to the initial annotation and the long-term curation of the *Pristionchus* genome sequence.

Table 1. Estimated cost: *Pristiochus pacificus* genomes

WGS library production & sequencing	$ 620,247
Fosmid end sequencing (one end only)	$ 34,560
Sequence improvement	$1,000,000
TOTAL	**$1,654,807**
Optional:	
BAC skims	$200 to 470K