Title
Assembly and Annotation of the Xenopus tropicalis Genome

Permalink
https://escholarship.org/uc/item/8p7944h8

Authors
Richardson, Paul
Detter, Chris
Terry, Astrid
et al.

Publication Date
2005-10-07
Assembly and Annotation of the *Xenopus tropicalis* Genome

Paul Richardson, Chris Detter, Astrid Terry, Nicholas Putnam, Andrea Aerts, Harris Shapiro, Inna Dubchak, Erika Lindquist, Darren Platt, Igor Grigoriev, Susan Lucas, Dan Rokhsar, Jim Bristow and Eddy Rubin

Paul M. Richardson  
[pmrichardson@lbl.gov](mailto:pmrichardson@lbl.gov)  
US DoE Joint Genome Institute  
2800 Mitchell Drive  
Walnut Creek, CA 94598

The US Department of Energy Joint Genome institute is sequencing the genome of the pipid frog *Xenopus tropicalis*. The genome is estimated to be approximately 1.7 billion base pairs contained in 10 pairs of chromosomes, and is sequenced to approximately 8X depth. Assembly Version 4.0 is the fourth in a series of preliminary assembly releases that are part of the ongoing *X. tropicalis* genome project. The assembly was constructed with the JGI assembler, Jazz, using 22.5 Million paired end sequencing reads in various sized insert libraries. This data set representing a coverage of 7.65X, assembled into 19,501 scaffolds totaling 1.51 Gbp. Roughly half of the genome is contained in 272 scaffolds all at least 1.56 Mb in length. A final draft assembly based on this sequence incorporating other data-including BAC-end sequence and physical map information is expected within the year.

The genome assembly was annotated using the JGI Annotation Pipeline, which runs analysis tools on a distributed UNIX system. Approximately 36k genes are predicted and annotated with the models built by *ab initio* and homology-based gene prediction methods, validated using available EST and cDNA data for *X. tropicalis* and *X. laevis* and genomic conservation between *X. tropicalis* and the human genome, and then annotated according to GO, KEGG and KOG functional classification. Average gene length is 16.5kb and average transcript length is 1.3kb while proteins average 409 amino acids in length. The genome sequence and functional analysis of predicted genes and their classification are available from JGI Genome Portal ([http://www.jgi.doe.gov/xenopus](http://www.jgi.doe.gov/xenopus)).

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.