

Lawrence Berkeley National Laboratory

Recent Work

Title

Genome Annotation: Going Beyond Human Genome

Permalink

<https://escholarship.org/uc/item/55t4r8mq>

Author

Grigoriev, Igor V.

Publication Date

2005-06-12

Genome Annotation: Going Beyond Human Genome

Igor V GRIGORIEV

US DOE Joint Genome Institute, Walnut Creek, CA 94598, ivgrigoriev@lbl.gov

Three human chromosomes [1-3] and a variety of eukaryotic genomes [4-9], dozens of microbial genomes and environmental communities [10-12] are being sequenced, assembled and annotated at the DOE Joint Genome Institute (<http://genome.jgi-psf.org>). Differences in genome organization and gene structure as well as variation in the data available for training and validation of gene predictions require custom approaches for each genome and create additional challenges for annotation. To address these challenges we have developed JGI annotation pipeline that allows us to annotate these genomes in high-throughput manner but also take into account specifics of individual genomes. The pipeline integrates a suite gene prediction and annotation methods, custom methods for validation and consolidation of various gene models, and visualization tools for genome-scale analysis. Analysis of genomes annotated using the Annotation Pipeline compared to the earlier genome releases, predictions made by other groups, and tests on data not used for gene prediction training indicate overall improvement in quality of predicted gene models.

1. J. Martin et al. (2004) The sequence and analysis of duplication-rich human chromosome 16. *Nature*. **432**: 988-994.
2. J. Shmutz et al. (2004) The DNA sequence and comparative analysis of human chromosome 5. *Nature*. **431**: 268-274.
3. J. Grimwood et al. (2004) The DNA sequence and biology of human chromosome 19. *Nature* **428**: 529-535.

4. E.V. Armbrust et al. (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science*. **306**: 79-86.
5. D. Martinez et al. (2004) Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nat Biotechnol*. **22**: 695-700.
6. J.J. Gibson-Brown et al (2003) A proposal to sequence the amphioxus genome submitted to the Joint Genome Institute of the US Department of Energy. *J Exp Zoolog B Mol Dev Evol*. **300**: 5-22.
7. A.R. Grossman et al. (2003) *Chlamydomonas reinhardtii* at the crossroads of genomics. *Eukaryot Cell*. **2**:1137-1150.
8. P. Dehal et al. (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*. **298**: 2157-2167.
9. S. Aparicio et al. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*. **297**: 1301-1310.
10. S. J. Hallam et al. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science*. **305**: 1457-1462.
11. G.W. Tyson et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. **428**: 37-43.
12. S.G. Tringe et al. (2005) Comparative metagenomics of microbial communities. *Science*. **308**: 554-557