THE GENOME CHANNEL: A FOUNDATION FOR GENOMES TO LIFE AND COMPARATIVE GENOMICS

Miriam Land, Frank Larimer, Jay Snoddy, Denise Schmoyer, Doug Hyatt, Manesh Shah, Inna Vokler, Phil LoCascio, Gwo-Liang Chen, Loren Hauser, Ed Uberbacher. Computational Biology Program, Life Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee. http://compbio.ornl.gov/

Representing DOE's unique perspective on eukaryotic and microbial genomes, the Genome Channel has made major improvements to its usability, organism coverage and supported types of analysis. Channel and related web sites receive an average of 25,000 web hits per day with usage comparing favorably with other major genome web resources. Genome Channel contains over 20 unique microbial genomes not contained on other sites and presents more types of human and mouse gene modeling than any other site. Integration of these many organisms in a single framework is making it possible to develop new tools and views that will help GTL researchers use comparative genomics to identify regulatory motifs and other conserved non-coding genome elements.

The Genome Channel has improved navigation tools which provide faster and easier access to the data. A new text-based home page has links for quick access to human and mouse data and an ever increasing list of both finished and draft microbials. Estimates of eukaryotic gene content in the Channel include five types: gene models based on GrailEXP, Genscan, Genbank annotation, U.Penn DOTS EST assemblies, and modeling based on Genbank Refseq mRNAs. Three of the methods include prediction of gene variants and these can be compared along with the evidence supporting each model. tRNA gene predictions have been added to both the text and graphical display. Microbial gene modeling is based on three methods; Generation, Glimmer, Critica, and an additional combined ORNL gene call based on these three methods. These suites of methods and organisms make the views provided for both eukaryotic and microbial genomes richer, more accurate, and more comprehensive than in other available resources. Many of the web pages now include links to 'lists' of data which can be downloaded and used with other tools. For example, lists of contigs, genes, or repeats on a chromosome or entire organism.

Annotation for JGI-sequenced draft microbials is accessible with links to several different summaries of the data which can be downloaded and put into other tools. These include Btab output for the predicted genes, Pfam summaries, and COGS analysis. Blast, Pfam, and COGs analysis can all be refreshed to get the latest interpretation of the data.

For GTL to be successful with Goal 2 (Gene Regulatory Networks), foundational data must be available that describes gene regulatory regions and potential transcription factor binding sites in genomes of interest. The most powerful paradigm for obtaining such information is through comparative genomics, since regulatory signals are often conserved among genomes, and also repeated within each genome proximal to genes of related function. Such inter- and intra-genome comparisons provide the basis for estimating regulatory signals and the role of genes in cell processes. The Channel and its underlying data warehousing structure and update processes are a natural foundation for facilitating comparison of genomes, including detailed comparison and analysis of regulatory units and structures in both microbes and eukaryotes. Steps are being taken to develop the Channel as a comprehensive resource for (1) facilitating and viewing large-scale genome alignment, (2) clustering and

comparing gene/protein orthologs and paralogs, and identifying conserved blocks and potential regulatory motifs in their upstream regions, and (3) recognizing operon structure and comparing regulatory signals in operons of related species, and (4) identifying shared regulatory elements in regulons in genomes of interest to the GTL program.

(Research was supported by the Office of Biological and Environmental Research, USDOE, under contract number DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed by UT-Battelle, LLC and through an ORNL Exploratory Studies Project.)