

Apollo: A Genome Annotation Editing Tool

Suzanna Lewis, Nomi Harris, Mark Gibson, Sima Misra and Guanming Wu

Application Area: Bioinformatics, Model Organism Database Software

Platform: Any (MacOSX, Windows, Unix, including Linux)

Dependencies: Java 1.4

Licensing: BSD License

Apollo is an interactive genome browser and editor. It allows researchers to explore genomic annotations at many levels of detail, and to perform expert annotation curation, all in a graphical environment. The Generic Model Organism Database (GMOD) project, also being presented as a demo at this meeting, aims to provide a complete ready-to-use toolkit for analyzing whole genomes. GMOD has adopted Apollo as its annotation workbench.

Apollo can display many different types of computational evidence for genome annotations, such as alignments, similarities based on BLAST searches, and gene predictions. It enables expert biologists to utilize the evidence to create and edit gene transcript models and other genomic features, such as transposable elements, pseudogenes, and tRNA genes, and to resolve differences between cDNA clone sequence alignment data to refine exon-intron structures predicted by gene prediction algorithms. Unlike typical annotation browsers, Apollo permits very detailed annotation editing. Editing is facilitated by edge matching, drag-and-drop annotation creation and modification, a multiple alignment viewer, and several specialized editing tools. The integrated detail editor displays the three-frame translation of the genomic sequence, with all start/stop codons highlighted, and enables drag and drop adjustment of exon endpoints. Starts and stops of translation can be modified, and the sequence of almost any feature can be downloaded as a FASTA file for further analysis (e.g. virtual cDNA or translation, genomic sequence). Text information such as gene and transcript synonyms and free-text comments can be associated with annotations and modified with another integrated editor. Non-consensus splice sites and other irregularities are flagged to aid in quality control. Translational frameshifts, genomic sequence errors, and other exceptions can be annotated as well.

Apollo is a Java application that can be downloaded and run on Windows, Mac OS X, or any Unix-type system (including Linux). Over 5000 users have downloaded Apollo, and an increasing number of groups are customizing this Open Source software to better fit their needs. These modifications range from changing display preferences to writing new data adapters to read from proprietary databases. A range of different data adapters enable loading data in GAME XML, Chado XML, GFF, GenBank, or EMBL formats, as well as via connections to a Chado or Ensembl relational database. User preferences, visual layouts, and many other aspects of Apollo are configurable as well. Extensive user documentation is available with the Apollo installation as well as at the Apollo web site, <http://www.fruitfly.org/annot/apollo/>.

We will demonstrate how to read data into Apollo, view the computational evidence, and create and edit detailed annotations. If time permits, we will explain some of the many ways in which Apollo can be customized.

References:

Apollo: a sequence annotation editor. Lewis SE, Searle SMJ, Harris N, Gibson M, Iyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews B, Prochnik SE, Smith CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME (2002). *Genome Biology*, 3(12):research0082.1-0082.14.