

WEBAPOLLO: A WEB-BASED SEQUENCE ANNOTATION EDITOR FOR DISTRIBUTED COMMUNITY ANNOTATION

Ed Lee¹, Gregg Helt¹, Nomi Harris¹, Mitch Skinner¹, Christopher Childers², Justin Reese², Monica C. Munoz-Torres², Christine G. Elisk², Ian Holmes³, Suzanna E. Lewis¹

¹Berkeley Bioinformatics Open-source Projects, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

²Georgetown University, Washington DC, 20057, USA

³Department of Bioengineering, University of California at Berkeley, Berkeley, CA, 94720, USA

Contact: Ed Lee (elee@berkeleybop.org)

As technical advances make sequencing faster and cheaper, genomic annotation efforts must adapt to keep pace. The upward trend in the number of genome sequencing projects means there will be a larger reliance on contributions from domain specialists. Thus the curation environment is shifting from a traditional centralized model, in which all curators for a given genome project share the same physical location, to a geographically dispersed community annotation model—which requires new tools to support community annotation efforts. WebApollo was designed to provide an easy to use, web-based environment that allows multiple distributed users to edit and share sequence annotations.

WebApollo is comprised of three components: a web-based client, a server-side annotation editing engine, and a server-side service that provides the client with data from different sources, including databases at the University of California at Santa Cruz, Ensembl, and Chado.

The web-based client is designed as an extension to JBrowse, a JavaScript-based genome browser that provides a fast, highly interactive interface for the visualization of genomic data. This JBrowse extension provides the gestures needed for editing annotations, such as dragging and dropping features to create new annotations of genes, transcripts and other genomic elements, dragging to change exon boundaries of existing annotations, and using context-specific menus to modify features. The extension also connects to the annotation-editing service and the data-providing services.

The server-side annotation-editing engine is written in Java. It handles all the necessary logic for editing and deals with the complexities of modifications in a biological context, where a single change can have multiple cascading effects (e.g., when splitting or merging transcripts). Edits are stored persistently in the server, allowing users to quickly recover their data in the event of unexpected browser or server crashes. The server provides synchronized updates over multiple browser instances, so that every edit is immediately visible to all users who are viewing or editing the same region. It offers multiple levels of user accessibility, allowing project owners to decide with whom to share their work, and whether to allow read-only or both read and write access.

The server-side service that provides data to the client is built on top of Trellis, a Distributed Annotation System (DAS) server framework. It sends JBrowse-supported JavaScript Object Notation (JSON) data, rather than the more verbose DAS XML. We also developed a Trellis plugin to access data from the UCSC MySQL genome database, which provides easy access to that popular data source. All three components are open source and provided under the BSD License.

An extensive group of curators and investigators from the bee genome research community, spread over academic institutions worldwide, are currently beta-testing WebApollo. Their curation efforts, findings and interactions will dramatically upgrade the quality of the annotation data for the genomes of honey bee (*Apis mellifera*), and two bumble bees (*Bombus impatiens* and *Bombus terrestris*), which will lead to a better understanding of the biology of these social insects. WebApollo will be publicly released in fall 2012.

Public demo: <http://icebox.lbl.gov:8080/ApolloWebDemo>

Project web page: <http://gmod.org/wiki/WebApollo>