WebApollo 2.0: A Scalable, Distributed Sequence Annotation Editor

Nathan Dunn; Mónica C Muñoz-Torres; Colin Diesh; Deepak Unni; Ian Holmes; Christine G Elsik; Suzanna E Lewis

Berkeley Bioinformatics Open-source Projects, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA; (CD, CE) University of Missouri, Columbia, MO, 65211, USA; (IH) Department of Bioengineering, University of California at Berkeley, Berkeley, CA, 94720, USA

WebApollo is a genomic annotation editor built on top of the JBrowse framework that supports a multitude of genome annotation projects. To support the diverse needs of a growing community and their data, we have undertaken several major efforts:

1) More flexible and informative layout via a side-dock allows a deeper alternate view of the data
2) Improved search functionality, validation checks, and editing capability streamlines quality control checks and improve adherence to best practices for submission to genome repositories
3) A simpler, more extensible, and transparent architecture that will allow users to more easily query their data and build extensions
4) Support for multiple organisms per server to minimize server resources
5) Provide user and group level permissions for editing and viewing annotations, tracks, and projects to support larger teams and provide public access
6) Allow more generic set of sequence annotations based on the Sequence Ontology

Additional information is available at http://genomearchitecture.org and https://github.com/modiapollo

1) More Flexible and Informative Layout
A) Removable side-dock provides an alternate view of annotation information.
B) Tabs have search functionality, additional details, and editing capability.
C) Allows detailed view of data.
D) Allows annotator to quickly switch between organisms and sequences.

3) Simpler Extensible, Transparent Architecture
Architecture moved to a single queryable data store and a simplified installation and configuration. Allows embedded or remote datastore for local or large production setup.

Grails controllers (a J2EE servlet) routes requests to the appropriate JBrowse data directory for a given organism.

GWT / Bootstrap used on front-end to provide rich application behavior. Communication with server is via REST and WebSockets for flexibility, speed, and service re-use.

4) Multiple Organisms
Multiple organisms may be browsed at once and hosted on the same server.
Each organism will have its own JBrowse data directory in addition to a database.

5) User and Group Permissions
Permissions can be set at a user and/or group level including public access.

Web Apollo is an established collaborative genome annotation tool. Due to the success of the Web Apollo platform and the growth of multi-organism collaborative annotation and education, it was necessary to re-architect the software stack and update the user interface. These changes will allow significant improvements in scalability, performance, and features.

Looking forward, we will work to allow annotation of variant data, display of multigenomic tracks, house JBrowse tracks directly in our datastore (allow persistence of analysis and remotely loaded tracks), and support further integration with other annotation pipelines and platforms.

This project was funded by the NIH (GM080203, HG004483), National Agricultural Library at the United States Department of Agriculture (NAL-USDA) (60-8260-4-005) and made possible by a very active group of contributors and collaborators from the GMOD, JBrowse, and Apollo communities. We encourage you to get involved.