

## Apollo: Scalable & collaborative curation for improved comparative genomics

Monica C Munoz-Torres<sup>1</sup>, Nathan A Dunn<sup>1</sup>, Deepak Unni<sup>2</sup>, Seth Carbon<sup>1</sup>, Colin Diesh<sup>2</sup>, Heiko Dietze<sup>1</sup>, Christopher Mungall<sup>1</sup>, Nicole Washington<sup>1</sup>, Christine E Elisk<sup>2</sup>, Ian Holmes<sup>3</sup>, and Suzanna E Lewis<sup>1</sup>

<sup>1</sup> Lawrence Berkeley National Laboratory, Genomics Division, Berkeley, CA. Email: [MCMunozT@lbl.gov](mailto:MCMunozT@lbl.gov)

<sup>2</sup> University of Missouri, Divisions of Plant and Animal Sciences, Columbia, MO.

<sup>3</sup> University of California Berkeley, Bioengineering, Berkeley, CA.

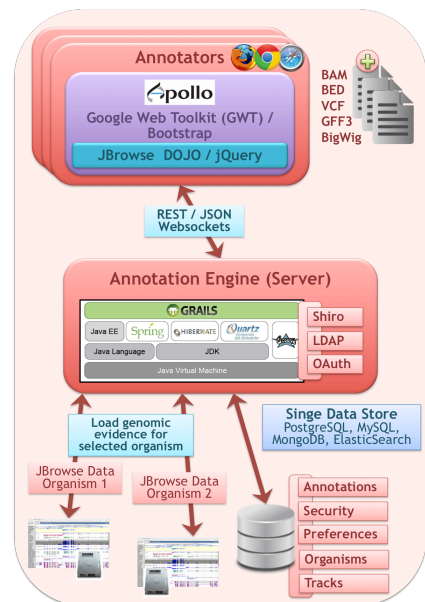
**Project Website:** <http://genomearchitect.org/>

**Source Code:** <https://github.com/GMOD/Apollo>

**License:** Berkeley Software Distribution (BSD) License. See <https://github.com/GMOD/Apollo/blob/master/LICENSE.md>

Comparative genome analysis requires high quality annotations of all genomic elements. Today's sequencing projects face numerous challenges including lower coverage, more frequent assembly errors, and the lack of closely related species with well-annotated genomes. Apollo is a web-based application that supports and enables collaborative genome curation in real time, analogous to Google Docs, allowing teams of curators to improve on existing automated gene models through an intuitive interface.

Apollo's architecture is built on top of the JBrowse framework and is composed of a web-based client, an annotation-editing engine, and a server-side data service. It allows users to visualize automated gene models, protein alignments, expression and variant data, and with these, conduct structural and/or functional annotations. To support the diverse needs of a growing community, we have recently completed two major efforts to improve functionality and performance: 1) Significant architectural changes to adopt the Grails JVM framework, and 2) Adoption of a queryable datastore to house annotations. The improved architecture allows users to more easily query the data and build extensions, supports multiple organisms per server, and also allows a larger set of sequence annotations based on the Sequence Ontology. A more flexible user interface via a removable side-dock provides improved search functionality, validation checks, and editing capability, and offers fine-grained user and group level permission.



Researchers from nearly one hundred institutions worldwide are currently using Apollo for distributed curation efforts in over sixty genome projects across the tree of life: from plants to arthropods, to fungi, to species of fish and other vertebrates including human, cattle (bovine), and dog. We are training the next generation of researchers by reaching out to educators to make these tools available as part of curricula, offering workshops and webinars to the scientific community, and through widely applied systems such as iPlant and DNA Subway. We are currently integrating Apollo into an annotation environment combining gene structural and functional annotation, transcriptomic, proteomic, and phenotypic annotation. In this presentation we will describe in detail its utility to users, introduce the new architecture, and offer details of our future plans.