Apollo: Collaborative and Scalable Manual Genome Annotation

Nathan A. Dunn¹, Monica C. Munoz-Torres¹, Deepak Unni², Eric Rasche⁴, Anthony Bretaudeau⁵, Eric Yao³, Ian Holmes³, Christine G. Elsik² and Suzanna E. Lewis¹

(1) Lawrence Berkeley National Laboratory, Berkeley, CA, (2) Division of Animal Sciences, University of Missouri, Columbia, MO, (3) Department of Bioengineering, Berkeley, CA, (4) Center for Phage Technology, Department of Biochemistry and Biophysics, Texas A & M University, College Station, TX, (5) BIPPA, Rennes, FR

http://genomearchitect.org/

https://github.org/GMOD/Apollo/
Example Genome Analysis Workflow

1. **Sequencing**
2. **Experimental design, sampling**
3. **Create Assembly**
4. **Comparative analyses**
5. **Automated Annotation**
6. **Manual Annotation**
7. **Curated Gene Set**
8. **Synthesis & dissemination**
Example Genome Analysis Workflow
Annotation

**Structural Annotation**
- exons, introns, UTRs
- repeat regions
- transposable elements
- tRNA, snRNA, snoRNA, miRNA, ncRNA, rRNA

**Functional Annotation**
- metabolic pathways / functions
- Gene Ontology
  - molecular function
  - biological process
  - cellular component
- expression
- gene families

Photo Credit: Alex Wild at [http://www.alexanderwild.com/](http://www.alexanderwild.com/)

http://geneontology.org
Analysis Requires High Quality Data

Sequencing

Create Assembly

Automated Annotation

Experimental design, sampling

Manual Annotation

Comparative analyses

Consensus Gene Set

Synthesis & dissemination
Automated Identification is not Perfect

Generation of Gene Models
Find ORFs, multiple rounds of gene prediction

Annotation of Gene Models
Predicting function, expression patterns, metabolic network memberships

• Assembly errors can cause fragmented annotations
• Limited coverage makes precise identification difficult
Manual Annotation Refines Genome

- Additional data
- Biological knowledge
- Curator best represents underlying evidence

Experimental Evidence

cDNAs, HMM domain searches, RNAseq, genes from other species
Analysis Requires High Quality Data

Sequencing

Create Assembly

Automated Annotation

Manual Annotation

Experimental design, sampling

Synthesis & dissemination

Analysis

Consensus Gene Set

Note: Data Science depends on good data
Note: assemblies often come in bad condition

Comparative analyses

Sequence. Transposable Elements. Gene models and much more.
Analysis Requires High Quality Data

- Understand Error
- Reduce Error

Sequencing

Create Assembly

Automated Annotation

Manual Annotation

Experimental design, sampling

Comparative analyses

Consensus Gene Set

Synthesis & dissemination
Apollo is Used to Produce High Quality Annotations

- Over 100 organizations use Apollo
- Multiple genomes and labs per server
Apollo is a Tool for Collaborative Annotation

- Web-based Editor
- Visual feedback
- Real-time collaborative
- Genomic browser

Photo Credits: i5K; Alex Wild at [http://www.alexanderwild.com/](http://www.alexanderwild.com/): leaf cutter ant, ensign wasp; Leo Bukeboom: *Nasonia vitripennis* jewel wasp; Wikimedia Commons: *Apis mellifera* honey bee; Mike MacNeil USDA/ARS Fort Keogh LARRL: *Bos taurus* cow.
1 - Evidence Viewer (Genome Browser)
1 - Evidence Viewer (Genome Browser)

Configure Multiple Tracks
1 - Evidence Viewer (Genome Browser)

Configure Multiple Tracks

Dynamically Open

Append via URL

Statically Configure

```
addStores=[{"url":{"type":"JBrowse/Store/SeqFeature/GFF3","urlTemplate":"http://host/genes.gff"}],
addTracks=[{"label":"genes","type":"JBrowse/View/Track/CanvasFeatures","store":"url"]
```
1 - Evidence Viewer (Genome Browser)

Configure Multiple Tracks

Dynamically Open

Append via URL

Statically Configure

- BAM
- BigWig
- GFF
- GTF
- GBK
- VCF
- FASTA
- FASTAi
- SPARQL
- custom types (e.g., REST endpoint)
1 - Evidence Viewer (Genome Browser)

Configure Multiple Tracks

Dynamically Open

Append via URL

Statically Configure

Customizable Views

- BAM
- BigWig
- GFF
- GTF
- GBK
- VCF
- FASTA
- FASTAi
- SPARQL
- custom types (e.g., REST end-point)

The GMOD Plugin Directory

https://gmod.github.io/jbrowse-registry/
### JBrowse Registry

<table>
<thead>
<tr>
<th>Plugin</th>
<th>Description</th>
<th>Author</th>
<th>License</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>agp_parser</strong></td>
<td>A JBrowse plugin to parse AGP files and display them</td>
<td>Colin Diesh</td>
<td>BSD 3-clause &quot;New&quot; or &quot;Revised&quot; License</td>
<td>cmdcolin/agp_parser</td>
</tr>
<tr>
<td><strong>Apollo</strong></td>
<td>Genome annotation editor with a Java Server backend and a Javascript client that runs in a web browser as a JBrowse plugin.</td>
<td>Nathan Dunn</td>
<td>BSD 3-clause &quot;New&quot; or &quot;Revised&quot; License</td>
<td>GMOD/Apollo</td>
</tr>
<tr>
<td><strong>BlastView</strong></td>
<td>JBrowse plugin to render gff3 formatted Blast results nicely and provide useful help menus</td>
<td>Eric Rasche</td>
<td>GNU Affero General Public License v3.0</td>
<td>TAMUCPT/blastview</td>
</tr>
</tbody>
</table>

[https://gmod.github.io/jbrowse-registry/](https://gmod.github.io/jbrowse-registry/)
https://gmod.github.io/jbrowse-registry/

- Demo Link
- GitHub Repo
- Configure Apollo to install easily
JBlas$t

Select an existing feature

Highlight Arbitrary Region

• Support any analysis server
• Support local execution

• JBlas$t “Example” Galaxy BLAST Implementation
• Apollo supports BLAT / BLAST, but this will be better
20

2 - Genome Annotation Editor

- **Exported Refined Genomic Elements**
- **Transcripts (GFF3, GBK)**
- **BAM Reads**
- **Automated Annotation**
  - **Transcripts (GFF3, GBK)**
  - **BigWig XY**
  - **BigWig HeatMap**

---

**Genome Annotation Editor**

- User-created Annotations
- Genome Annotation Editor
- Transcripts (GFF3, GBK)
- BAM Reads
- Transcripts (GFF3, GBK)
- BigWig XY
- BigWig HeatMap

---

**Manual Annotation**

- Reference Sequences
- Organism
- Other

---

**Honeybee**

- Annotations
- Tracks
- Ref Sequence
- Organism
- Other
Create Annotation

Add Annotation by Dragging a Genomic Element

Alignments shown in red

Annotate other genomic types with drop-down

- View details
- Highlight this mRNA

Create new annotation
- gene
- pseudogene
- tRNA
- snRNA
- snoRNA
- miRNA
- repeat_region
- transposable_element
Edit Annotation Structure

Adjust exon by dragging
Editing Annotations

- Change Annotation Type
- Edit Additional Structural Data (right-click popup)
- History of Structural Edits

Edit Associations
- PubMed / dbxref
- Gene Ontology
- Metadata
- key/value
- status
- comments

Change annotation type
- gene
- pseudogene
- rRNA
- snRNA
- snoRNA
- tRNA
- ncRNA
- miRNA
- repeat_region
- transposable_element

History of Structural Edits

<table>
<thead>
<tr>
<th>Operation</th>
<th>Editor</th>
<th>Date</th>
<th>Revert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add transcript</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>6/14/16 12:53 PM</td>
<td></td>
</tr>
<tr>
<td>Set translation end</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>6/14/16 12:53 PM</td>
<td></td>
</tr>
<tr>
<td>Set readthrough stop codon</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>6/14/16 12:53 PM</td>
<td></td>
</tr>
<tr>
<td>Change annotation type</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>6/14/16 12:53 PM</td>
<td></td>
</tr>
</tbody>
</table>
Edit Annotation Structure

**Revertible History of Structural Operations**

<table>
<thead>
<tr>
<th>Operation</th>
<th>Editor</th>
<th>Date</th>
<th>Revert</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merge transcripts</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>10/20/16 9:17 PM</td>
<td></td>
</tr>
<tr>
<td>Delete exon</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>10/20/16 9:17 PM</td>
<td></td>
</tr>
<tr>
<td>Set exon boundaries</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>10/20/16 9:18 PM</td>
<td></td>
</tr>
<tr>
<td>Set exon boundaries</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>10/20/16 9:18 PM</td>
<td></td>
</tr>
<tr>
<td>Set translation start</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>10/20/16 9:18 PM</td>
<td></td>
</tr>
<tr>
<td>Set exon boundaries</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>10/20/16 9:18 PM</td>
<td></td>
</tr>
</tbody>
</table>

Highlighted row shown

Current position
Annotate Reference Sequence Alterations
3 - Annotator Panel

Collapsible

Link to Location

Navigate Sequence and Organism
Annotation Tab - Alternate View
Reference Sequence - Search and Export
Organism: Configuration

Create JBrowse tracks from FASTA / GFF3 / BAM / BigWig


Import JBrowse directory

Share “Public” organisms
Users and Groups

Add / Search Users

Edit User Permission

User Can “Admin” an Organism

Use Groups to Manage Bulk Permissions
Admin Panel

- Predefine Curation Terms
- Reports

- Annotation Types
- Ref Sequence
- Organism
- Users
- Groups
- Admin

- Canned Comments
- Canned Key
- Canned Values
- Feature Types
- Statuses

- Proxies
- Reports

- System Info
- Performance Metrics
- WebServices
Summary of Features

- **Genome Annotation Editor**
- **Visual Feedback**
- **Evidence Viewer**
- **History**
- **Structural + Functional**
- **Annotator Panel**
- **Customizable**
- **Multiple data types**
- **Easy to navigate**
- **Fast**
- **Alternate view of data**
- **Simplified administration**
Architecture

Web Services Client
Perl, Shell, Groovy, PHP, etc.

File System

Security
SHIRO

Server
Apollo Server - Grails

REST

Apollo

Annotators

WebSocket

GFF

JSON

FASTA

JBrowse

DOJO / jQuery

Google Web Toolkit (GWT) / Bootstrap

WebSocket

JDBC

File System

Apollo Client(s)

Client(s)

Web Services Client
Perl, Shell, Groovy, PHP, etc.
Architecture

Web Services Client
Perl, Shell, Groovy, PHP, etc.

Client(s)

WebSocket

REST

Annotators
Apollo

Apollo Server - Grails

File System

Security

Server

File System

Apollo Server - Grails

JDBC

H2

MySQL

Annotation Services
Group Services
File Services

Annotation Services - methods for setting the annotation engine
Group Services - methods for managing groups
File Services - methods for managing and querying files

{JSON}

{GFF}

{FASTA}

SHIRO

Spring

MyEclipse Edition

Groovy

MyVisual Studio

Jetty

Resin

HQL

Hibernate

Web Services Client

Perul

Shell

Groovy

PHP

etc.

Annotators

Apollo

Google Web Toolkit (GWT) / Bootstrap

JBrowse

DOJO / jQuery

WebSocket

JDBC

File System
Architecture

Web Services Client
Perl, Shell, Groovy, PHP, etc.

Client(s)

WebSocket

Server
Apollo Server - Grails

File System

Security
SHIRO

REST

JDBC

Apollo Annotators

GFF

FASTA

JSON

Apollo Server - Grails

JBrowse

DOJO / jQuery

Google Web Toolkit (GWT) / Bootstrap

WebSocket

JDBC

File System

Security
SHIRO

Client(s)

WebSocket

Server
Apollo Server - Grails

File System

Security
SHIRO

Client(s)
Integration Strategies

- Configure `apollo-config.groovy`
- Grails plugin
- JBrowse plugin
- Database
- Fork
- Ask
- Web services

https://github.org/GMOD/Apollo/
apollo@lists.lbl.gov

Perl, Shell, Groovy, PHP, etc.
Scriptable Web Services

- Examples: Groovy, Perl, shell, Python
- Autogenerated REST API doc in Apollo

BASE URL: http://icebox.lbl.gov/Apollo2

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation Services</td>
<td>Methods for running the annotation engine</td>
</tr>
<tr>
<td>Group Services</td>
<td>Methods for managing groups</td>
</tr>
<tr>
<td>IO Services</td>
<td>Methods for bulk importing and exporting sequence data</td>
</tr>
<tr>
<td>Organism Services</td>
<td>Methods for managing users</td>
</tr>
<tr>
<td>User Services</td>
<td>Methods for managing users</td>
</tr>
</tbody>
</table>

Example curl command:
```
curl -d "{"operation": "get_features", 'track': 'Group1.10', 'username': 'ndunn@me.com', 'password': 'demo'}" http://localhost:8080/apollo/AnnotationEditorService
```
Workflow

Automated Annotations
GFF3 FASTA BAM, etc

JBrowse

Manual Annotation

GFF3 FASTA CHADO

Next Workflow

TRIPAL
Automated Annotations
GFF3 FASTA BAM, etc

Manual Annotation

GFF3 FASTA CHADO

TRIPAL?

Next Workflow

{ REST }
Get Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Statistics
Graph/Display Data

**JBrowse**

JBrowse genome browser

JBrowse – Data Directory to Standalone upgrades the bare data directory to a full JBrowse instance

**Apollo**

Register Account with WebApollo

Create or Update Organism will create the organism if it doesn’t exist, and update otherwise

Retrieve Data from Apollo into Galaxy

Retrieve JBrowse for an organism, from Apollo

Annotate opens an iFrame to Apollo

List Organisms in Apollo

https://github.com/GMOD/docker-compose-galaxy-annotation

View directly in <server>:<port>/apollo/
**Galaxy**

**REST**

View directly in `<server>:`<port>/apollo/

https://github.com/GMOD/docker-compose-galaxy-annotation

Get Data
Lift-Over
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
Extract Features
Fetch Sequences
Fetch Alignments
Statistics
Graph/Display Data

**JBrowse**

JBrowse genome browser

JBrowse – Data Directory to Standalone upgrades the bare data directory to a full JBrowse instance

**Apollo**

Register Account with WebApollo
Create or Update Organism will create the organism if it doesn’t exist, and update otherwise
Retrieve Data from Apollo into Galaxy
Retrieve JBrowse for an organism, from Apollo
Annotate opens an iFrame to Apollo
List Organisms in Apollo
Other Integration Projects

https://github.com/galaxy-genome-annotation/

- Pre-built workflow with Annotation Tools, Dockerized
- Genome Annotation Birds of a Feather Session at 6 pm.

https://github.com/galaxy-genome-annotation/python-apollo

- Arrow from Parsec, python library that wraps web-services

```bash
$ arrow groups create_group university
{
"publicGroup": false,
"class": "org.bbop.apollo.UserGroup",
"name": "university",
"users": null,
"id": 558319
}

$ arrow users get_users | \
  jq '.[] | select(.username | contains("@tamu.edu")) | .username' | \
  xargs -n1 arrow users add_to_group university
```

https://github.com/GMOD/docker-apollo
Summary

Curators refine genome annotations

Real-time collaborative

Visual evidence and feedback

Integrates within workflow
Future Work: Projection

2.3 Phenotype annotation

2.2 Variant annotation and visualization

2.1 Projection

DB backend, Sidebar, Grails, Multi-organism, WS

2.0 Mavenize

Web Apollo
Desktop Apollo

Assembly Composition

Genome Folding

Variant annotation and visualization
Reverse Complement: 2.1.0 (Jul/Aug)
Reverse Complement: 2.1.0 (Jul/Aug)
Genome Folding (2.1.1)

Fold Between Exons

Fold All Introns

View Sequence Boundaries
Genome Folding (2.1.1)

Fold Between Exons

Fold All Introns

View Sequence Boundaries
Fold Individual Features (2.1.2)
Assembly Composition (2.1.3)

Select to Combine

Drag to Rearrange

Used Scaffolds

Flip Entire Orientation

Set Orientation

Partial Scaffold
Cross-Scaffolds Operations

- Merge
- Modify Exon Boundary
Variant Annotation and Visualization (2.2)

- Phenotype annotation
- Variant annotation and visualization
- Projection
- DB backend, Sidebar, Grails, Multi-organism, WS
- Mavenize
- Web Apollo
- Desktop Apollo

Annotate Variants

Create from Evidence (e.g., VCF)

Visual Predictions
• Berkeley Bioinformatics Open-source Projects (BBOP), Berkeley Lab: Apollo and Gene Ontology teams. Suzanna E. Lewis (PI).

• § Christine G. Elsik (PI). University of Missouri.

• * Ian Holmes (PI). University of California Berkeley.

• Apollo is supported by NIH grants 5R01GM080203 from NIGMS, and 5R01HG004483 from NHGRI. Also supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

Alex Wild at http://www.alexanderwild.com/: leaf cutter ant, ensign wasp; Leo Bukeboom: Nasonia vitripennis jewel wasp; Wikimedia Commons: Apis mellifera honey bee; Mike MacNeil USDA/ARS

• Thanks to you and the Apollo / GMOD Communities

• http://genomearchitect.org

• https://github.org/GMOD/Apollo/

NIH National Institutes of Health
Berkeley Bioinformatics Open-source Projects (BBOP), Berkeley Lab: Apollo and Gene Ontology teams. Suzanna E. Lewis (PI).

Christine G. Elsik (PI). University of Missouri.

Ian Holmes (PI). University of California Berkeley.

Apollo is supported by NIH grants 5R01GM080203 from NIGMS, and 5R01HG004483 from NHGRI. Also supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

Alex Wild at http://www.alexanderwild.com/: leaf cutter ant, ensign wasp; Leo Bukeboom: Nasonia vitripennis jewel wasp; Wikimedia Commons: Apis mellifera honey bee; Mike MacNeil USDA/ARS

Thanks to you and the Apollo / GMOD Communities

http://genomearchitect.org
https://github.org/GMOD/Apollo/
Extra Slides