Community Annotation with Apollo: Engaging the Collective in a Genome Annotation Workflow

Nathan A. Dunn¹, Deepak Unni¹, Eric Yao², Robert Buels², Ian Holmes², Christine G. Elsik³ and Suzanna E. Lewis¹

(1) Lawrence Berkeley National Laboratory, Berkeley, CA, (2) Department of Bioengineering, Berkeley, CA, (3) Division of Animal Sciences, University of Missouri, Columbia, MO,

http://genomearchitect.org/

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

Sequencing

Experimental design, sampling

Create Assembly

Comparative analyses

FGENESH

Automated Annotation

Manual Annotation

Curated Gene Set

Synthesis & dissemination
Analysis Requires High Quality Data

- Sequencing
- Create Assembly
- Automated Annotation
- Manual Annotation
- Experimental design, sampling

Comparative analyses

Consensus Gene Set

Synthesis & dissemination

Analysis Requires High Quality Data

- Experimental design, sampling
- Automated Annotation
- Manual Annotation
- Comparative analyses

- Consensus Gene Set

Analysis Requires High Quality Data

- Experimental design, sampling
- Automated Annotation
- Manual Annotation
- Comparative analyses

- Consensus Gene Set

Analysis Requires High Quality Data

- Experimental design, sampling
- Automated Annotation
- Manual Annotation
- Comparative analyses

- Consensus Gene Set
New Genome Annotation Tool Requirements

• Can assist when assemblies are poor
• Intuitive user interface
• Easy to install and configure to suit any lab’s sequencing workflow
• Intuitive programmer interface
• Support collaborators who are geographically dispersed
Automated Identification is not Perfect

- Assembly errors can cause fragmented annotations
- Limited coverage makes precise identification difficult

**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

- Dynamically include additional analysis
- Make use of the researcher’s expertise
- Integrate all underlying evidence

Experimental Evidence

- cDNAs, HMM domain searches, RNAseq, genes from other species.
New Genome Annotation Tool Requirements

- Can assist when assemblies are poor
- Intuitive user interface
- Easy to install and configure to suit any lab’s sequencing workflow
- Intuitive programmer interface
- Support collaborators who are geographically dispersed
Apollo is a Tool for Collaborative Annotation

Biology is a team sport

Manual tasks require more hands

Photo Credits: i5K; 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 Fort Keogh LARRL: Bos taurus cow.
Apollo Overview

- Genome Annotation Editor
- Evidence Viewer
- Annotator Panel
1 - Evidence Viewer / Genome Browser

- Transcripts (GFF3, GBK)
- BAM Reads
- Automated Annotation
- Transcripts (GFF3, GBK)
- BigWig XY
- BigWig HeatMap

Themes (dark/light)
Color CDS Frame
Configure Multiple Tracks

Dynamically Open

Append via URL

Statically Configure

Customizable Views

1 - Evidence Viewer (Genome Browser)

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

The GMOD Plugin Directory

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

https://gmod.github.io/jbrowse-registry/

44 registered plugins

Sashimi Plot – Converts RNA-seq BAM files into intron support coverage features.

Multi Variant Viewer – Multi-VCF viewer 1000 Genomes

@cmdcolin

Lollipop, SVG Tracks with MyVariants.info plugin
Getting JBrowse

[Diagram showing Docker, npm, Galaxy Tool Shed, and Desktop Electron App]

- Docker Image: https://hub.docker.com/r/jbrowse/gmod-jbrowse/
- npm Package: https://www.npmjs.com/package/@gmod/jbrowse
- Desktop Electron App: @cmdcolin
- Galaxy Tool Shed: @erasche
- Website: http://jbrowse.org
JBConnect

JBrowse Server

@ennugetry

Queueing (e.g., JBlas)

Connect to Remote (e.g., Galaxy) and Local Servers
2 - Genome Annotation Editor

- Transcripts (GFF3, GBK)
- BAM Reads
- Automated Annotation
- Transcripts (GFF3, GBK)
- BigWig XY
- BigWig HeatMap

Exported Refined Genomic Elements
Create Annotation

Add Annotation by Dragging a Genomic Element

Alignments shown in red

Annotate other genomic types with drop-down
Edit Annotation Structure

Adjust exon by dragging
Editing Annotations

Change Annotation Type

Edit Additional Structural Data (right-click popup)

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

History of Structural Edits
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
Create Variant Annotation (2.1.0)

Add Variant Annotation by Dragging a Genomic Element

Or by a Right-click

Copy / Edit Properties

Export VCF

@deepakunni3
Annotate Reference Sequence Alterations

Alteration Reflected
3 - Annotator Panel

- Link to Location
- Collapsible
- Navigate Sequence and Organism
Annotations
Reference Sequence - Search and Export
Organism (Admin)

Create JBrowse tracks from FASTA / GFF3 / BAM / BigWig

Import JBrowse directory

Share “Public” organisms

Users and Groups (Admin)

http://gonramp.wustl.edu/  @Yating-L

Added Instructor Role to Manage Organisms
- Edit user permissions
- Create / edit organisms

Use Groups to Manage Bulk Permissions

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
<th>Global Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nathan Dunn</td>
<td><a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a></td>
<td>admin</td>
</tr>
<tr>
<td>Suzie Lewis</td>
<td></td>
<td>admin</td>
</tr>
<tr>
<td>Deepak Unni</td>
<td></td>
<td>admin</td>
</tr>
<tr>
<td>Moni Munoz-Torres</td>
<td></td>
<td>admin</td>
</tr>
<tr>
<td>Test User</td>
<td></td>
<td>user</td>
</tr>
<tr>
<td>Colin Diesh</td>
<td></td>
<td>admin</td>
</tr>
<tr>
<td>New Test</td>
<td></td>
<td>user</td>
</tr>
<tr>
<td>Human Admin</td>
<td><a href="mailto:ndunn@me.com">ndunn@me.com</a></td>
<td>user</td>
</tr>
<tr>
<td>Demo User</td>
<td><a href="mailto:demo@demo.com">demo@demo.com</a></td>
<td>user</td>
</tr>
<tr>
<td>New User</td>
<td><a href="mailto:newuser@newuser.com">newuser@newuser.com</a></td>
<td>user</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details</th>
<th>Groups</th>
<th>Organisms</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nathan Dunn [<a href="mailto:nathandunn@lbl.gov">nathandunn@lbl.gov</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suzie Lewis [<a href="mailto:telewis@lbl.gov">telewis@lbl.gov</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deepak Unni [<a href="mailto:deepak.unni3@gmail.com">deepak.unni3@gmail.com</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moni Munoz-Torres [<a href="mailto:McMunozT@lbl.gov">McMunozT@lbl.gov</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test User [<a href="mailto:test@user.com">test@user.com</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colin Diesh [<a href="mailto:colin.cliesth@gmail.com">colin.cliesth@gmail.com</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Test [<a href="mailto:newtest@newtest.com">newtest@newtest.com</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Admin [<a href="mailto:ndunn@me.com">ndunn@me.com</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demo User [<a href="mailto:demo@demo.com">demo@demo.com</a>]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Added Instructors Role to Manage Organisms

Edit user permissions
Create / edit organisms
Use Groups to Manage Bulk Permissions
Admin

**Predefine Curation Terms**

**Reports**
Summary of Features

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

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

BASE URL: http://demo.genomearchitect.org/Apollo2

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

Create Organisms and Tracks on the Fly
Track and Variant Services

https://github.com/GMOD/GenomeFeatureComponent

ALLIANCE of GENOME RESOURCES
Python API

https://pypi.org/project/apollo/  @erasche  @abretaud

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

@luke-c-sargent  @Yating-L  http://gonramp.wustl.edu/

Automated Annotations
GFF3 FASTA BAM, etc

Manual Annotation
GFF3 FASTA CHADO VCF
Integration into Workflow and Tools

- Over 100 organizations refine annotation
- Multiple genomes per organization

Refined Annotations Distributed to Public

- NCBI
- Ensemble
- OrthoDB
Summary

Curators refine genome annotations

Real-time collaborative

Visual evidence and feedback

Integrates within workflow
Thank You

**BBOP**

- **Apollo**
  - Nathan Dunn
  - Deepak Unni

- **JBrowse***
  - Eric Yao
  - Robert Buels

- **“Galaxy”**
  - Eric Rasche
  - Anthony Bretauadeau
  - Yating Lui
  - Luke Sargent

- **Gene Ontology**
  - Monica Munoz-Torres
  - Colin Diesh*
  - Chris Mungall
  - Seth Carbon

- **“Galaxy”**

- **IntelliJ IDEA**

- **YourKit**

- **GitHub**

- **NIH** National Institutes of Health

- **U.S. DEPARTMENT OF ENERGY**

---

- **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 was supported by NIH grants 5R01GM080203 from NIGMS, and 5R01HG004483 from NHGRI.

- **Thanks to you and the Apollo and GMOD Communities**

---

Apollo: [http://genomearchitect.org](http://genomearchitect.org)
https://github.org/GMOD/Apollo/

---

Poster: B28
Getting Apollo
http://genomearchitect.org
GitHub GMOD/Apollo

Future?
• Improved widget integration
• Tapering support, Collaborators?
  • Further integrations
  • Micro-attributions
  • Phenotypes
  • Variant Effect
  • ???

Activities
• apollo@lbl.gov
  Poster: B28 (tomorrow)
• There *was* a JB birds of a feather today
• JBrowse Collaboration Fest Friday-Sunday
• https://gitter.im/GMOD/jbrowse

Questions?
Extra Slides
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/)
Integration Strategies

• Configure `apollo-config.groovy`

• Grails plugin

• JBrowse plugin

• Database

• Fork

• Ask

• Web services

Perl, Shell, Groovy, PHP, etc.

https://github.org/GMOD/Apollo/
apollo@lists.lbl.gov
Automated Annotations
GFF3 FASTA BAM, etc

Manual Annotation
GFF3 FASTA CHADO VCF

Public Data

Next Workflow

Example

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

Automated Annotations
GFF3 FASTA BAM, etc

JBrowse

Manual Annotation

GFF3 FASTA CHADO

Next Workflow

TRIPAL
Architecture

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

Client(s)

REST

Server

File System

Security

Apollo Server - Grails

WebSocket

Annotation Services
Group Services
ID Services

GFF

FASTA

JSON

Apollo

File System

JDBC

H2

MySQL
Architecture

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

Client(s)

WebSockets

REST

Server
Apollo Server - Grails

Security
SHIRO

File System

Annotators
Apollo

Apollo

File System

JDBC

GFF

FASTA

JSON

Annotation Services
Group Services
Job Services

Perl, Shell, Groovy, PHP, etc.
Architecture

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

Client(s)

REST

Server
Apollo Server - Grails

File System

Security

Annotators

WebSocket

File System

Apollo

Annotators

JBrowse

JUnit

Google Web Toolkit (GWT) / Bootstrap

DOJO / jQuery

WebSocket

JDBC

JDBC

File System

H2

MySQL
Tracks

Annotaions | Tracks | Ref Sequence | Organism | Users | Groups | Admin
---|---|---|---|---|---|---

Search

454 Contigs  0/0

BigWig Density  0/1

- Forager RNA-Seq HeatMap

BigWig XYPlot  1/1

- Forager RNA-Seq XY Plot

EST  0/0

Gene Predictions  3/12

- Official Gene Set v3.2
- Fgenes
- Fgenes++ without RNASeq training data
- Fgenes++ with RNASeq training data
- Official Gene Set v1.0
- NCBI RefSeq Protein Coding Genes
- NCBI RefSeq Low Quality Protein Coding Genes
- NCBI RefSeq miRNA
- NCBI RefSeq Pseudogene
- NCBI RefSeq Noncoding RNA
- OGSv3.2 genes that merge RefSeq genes
- OGSv3.2 genes that split RefSeq genes

Search