Apollo: Collaborative Manual Annotation for Genome Sequencing Projects

Nathan A Dunn\textsuperscript{1}, Monica C. Munoz-Torres\textsuperscript{1}, Colin Diesh\textsuperscript{2}, Deepak Unni\textsuperscript{2}, Eric Rasche\textsuperscript{3}, Eric Yao\textsuperscript{3}, Ian Holmes\textsuperscript{3}, Christine G. Elsik\textsuperscript{2} and Suzanna Lewis\textsuperscript{1}

(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

http://genomearchitect.org/

https://github.org/GMOD/Apollo/
Genome Annotation

Structural Annotation
- exons, introns, UTRs
- repeat regions
- transposable elements

Functional Annotation
- Gene Ontology
  - molecular function
  - biological process
  - cellular component
- expression
- metabolic pathways / functions
- gene families
- molecular evolution
- sequence conservation
Example Genome Analysis Workflow

1. **Sequencing**
   - Experimental design, sampling
   - Synthesis & dissemination

2. **Create Assembly**
   - HiSAT
   - StringTie
   - FGENESH

3. **Automated Annotation**
   - Stacks
   - Manual Annotation

4. **Comparative analyses**
   - Consensus Gene Set

5. **Synthesis & dissemination**
   - Experimental design, sampling
   - Sequencing

- Tools: HiSAT, StringTie, FGENESH, Stacks, Manual Annotation, Consensus Gene Set
Galaxy Automates Software Pipelines

- Sequencing
- Experimental design, sampling
- Create Assembly
- HiSAT
- StringTie
- Automated Annotation
- Manual Annotation
- Comparative analyses
- Consensus Gene Set
- Gene Set
- FGENESH
- Automated Annotation
- Synthesis & dissemination
- Galaxy
Apollo is a Manual Annotation Tool

- Sequencing
- Experimental design, sampling
- Create Assembly
- HiSAT
- StringTie
- FGENESH
- Automated Annotation
- Annotation
- Manual Annotation
- Consensus Gene Set
- Comparative analyses
- Synthesis & dissemination
Automated Identification is not Perfect

- Automation of Gene Models
  - Find ORFs, multiple rounds of gene prediction, etc.
- 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 expertise

Experimental Evidence

cDNAs, HMM domain searches, RNAseq, genes from other species.

Human Analysis
Apollo is a Tool for Manual Annotation

- Web-based Editor
- Real-time collaborative
- Built on REST API
- Genomic browser

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.
Genome Browser - Evidence Viewer

- Transcripts (GFF3, GBK)
- BAM Reads
- Automated Annotation
- Transcripts (GFF3, GBK)
- BigWig XY
- BigWig HeatMap
- Themes (dark/light)
- Color CDS Frame
Genome Annotation Editor

Exported Refined Genomic Elements

Transcripts (GFF3, GBK)

BAM Reads

Automated Annotation

Transcripts (GFF3, GBK)

BigWig XY

BigWig HeatMap
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
- ncRNA
- rRNA
- miRNA
- repeat_region
- transposable_element
Edit Annotation Structure

Adjust exon by dragging

Sequence Alterations

Undo / Redo Operations
Editing Annotations

Change Annotation Type

Edit Additional Structural Data (right-click popup)

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

History of Structural Edits
Export Annotations and Navigate Sequences

Export 3 sequences from Honeybee as GFF3

- GFF3
- GFF3 with FASTA

Export 2 sequences from Honeybee as FASTA

- Genomic
- cDNA
- CDS
- Peptide

Search

Navigation
Import and Share Organisms

Import organisms from existing JBrowse directory

Share “Public” organisms

Choose Organism to View:
- Cat
- Celegans
- Chicken
- Cow
- Cow2
- Dog
- Honeybee

Import organisms from existing JBrowse directory.
Manage Who Sees What

- Add / Search Users
- Edit User Permission
- User Can “Admin” an Organism

Use Groups to Manage Bulk Permissions
Other Integrations

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

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

- Automated Annotations
  GFF3 FASTA BAM, etc

- Manual Annotation

- GFF3 FASTA CHADO

- Next Workflow

- TRIPAL?

- Built on REST API

- BASE URL: http://icebox.lbl.gov/Apollo2
Example Integrations

- **Workspace@NAL**: REST API
- **GenSAS**: GFF3 output, scripts
- **VectorBase**: DNA SUBWAY
- **Galaxy**: Docker + REST API
Galaxy Integration

To Build
- [https://github.com/GMOD/docker-compose-galaxy-annotation](https://github.com/GMOD/docker-compose-galaxy-annotation)
- docker-compose up

Based on
- REST API Integration (Eric Rasche) - TAMU-CPT:
- Docker Tools:
  - Galaxy: [https://github.com/bgruening/docker-galaxy-stable](https://github.com/bgruening/docker-galaxy-stable)
  - JBrowse: [https://hub.docker.com/r/jbrowse/gmod-jbrowse/](https://hub.docker.com/r/jbrowse/gmod-jbrowse/)

Docker Build
- [https://github.com/GMOD/docker-apollo](https://github.com/GMOD/docker-apollo)
- docker-compose up

Standard Build Quick-start
- [https://github.com/GMOD/Apollo](https://github.com/GMOD/Apollo)
- ./apollo run-local (dev)
Galaxy Integration
Galaxy / Apollo Integrated Workflow

Automated Annotations
GFF3 FASTA BAM, etc

Import into JBrowse

Import Organism

Refine Annotations
Manual Annotation
GFF3 FASTA CHADO

Export Annotations

Next Workflow

TRIPAL

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
Import into JBrowse

- Add JBrowse Genome Browser
- Add FASTA Sequence
- Add one or more Tracks (e.g. GFF3) to JBrowse
View in 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
Import Organism to Apollo

- Import Organism from existing JBrowse View

---

1. Get Data
2. Lift–Over
3. Text Manipulation
4. Filter and Sort
5. Join, Subtract and Group
6. Convert Formats
7. Extract Features
8. Fetch Sequences
9. Fetch Alignments
10. Statistics
11. 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

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

3. Retrieve Data from Apollo into Galaxy
4. Retrieve JBrowse for an organism, from Apollo
5. Annotate opens an IFrame to Apollo
6. List Organisms in Apollo
Annotate in Apollo

- View directly in `<server>:`:<port>/apollo/`
Export Manual Annotations

• Export Annotated Sequences
• Both GFF3 and FASTA at once

Retrieve Data from Apollo into Galaxy (Galaxy Version 2.0)

Organism Common Name Source
- Direct Entry

Organism Common Name
Voivox Sample Data

refseqs
- Insert

What it does
Exports the GFF3/Fasta sequence from Apollo.

Citation
If you use this tool in Galaxy, please cite: Eric Rasche (2016), Galaxy wrapper
Summary

• Apollo refines genomic annotations, collaboratively

• Apollo integrates with Galaxy using Docker and Web Services
Future Apollo Work

1.0
Mavenize
Web Apollo
Desktop Apollo

2.0
DB backend, Sidebar, Grails, Multi-organism, WS

2.1
Coordinate Transformation

2.2
Variant annotation and visualization

2.3
Phenotype annotation

Collapse introns and space between features
Bring scaffolds together
View and link genes across scaffolds
Questions?

BBOP

Apollo
Nathan Dunn
Monica Munoz-Torres
Deepak Unni §
Colin Diesh §

Texas A&M University
Eric Rasche
(Galaxy / Docker integration)

JBrowse
Eric Yao

*Nathan Dunn

Gene Ontology
Chris Mungall
Seth Carbon
Jeremy Nguyen

• 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.

• Stephen Ficklin, GenSAS, Washington State University

• 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

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

Automated Annotations
GFF3 FASTA BAM, etc

JBrowse

Manual Annotation

GFF3 FASTA CHADO

Next Workflow

TRIPAL
Galaxy Integration

- Web Services Integration (Eric Rasche) - TAMU-CPT:
  - Web services pip library:
- Docker Tools (Show Docker Apollo Image)
  - Galaxy: [https://github.com/bgruening/docker-galaxy-stable](https://github.com/bgruening/docker-galaxy-stable)
  - Apollo Docker library: [https://github.com/GMOD/docker-apollo](https://github.com/GMOD/docker-apollo)
  - Use with JBrowse one: [https://github.com/erasche/docker-jbrowse](https://github.com/erasche/docker-jbrowse)
  - Integration: [https://github.com/erasche/docker-recipes](https://github.com/erasche/docker-recipes)
- Tool uses Web Services and external volumes for integration
Scriptable Web Services

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

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

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

- Get GFF3
- Zoom to Base Level
- Edit Information (alt-click)

- Delete
- Merge
- Split
- Duplicate
- Make Intron
- Move to Opposite Strand

- Set Translation Start
- Set Translation End
- Set Longest ORF
- Set Readthrough Stop Codon

- Set as 5' end
- Set as 3' End
- Set both Ends

- Set to Downstream Splice Donor
- Set to Upstream Splice Donor
- Set to Downstream Splice Acceptor
- Set to Upstream Splice Acceptor

- Undo
- Redo
- Show History

Retrieve Individual Sequences

Retrieve Entire Scaffold
Server Criteria

- Quick setup and easy to deploy
- Plugins (websockets, security, etc.)
- Scaleable
- Rapid Application Development
  - Integrated testing
- Bonus: Minimal changes
- Mature and well-supported

Considerations
- Node.js
- Ruby on Rails
- Grails
- Struts, Spring MVC, etc.
- Django, Php, etc.

DataStore
- Thread-safe transactions
- Transparent / Queryable
- Structured
- Simple Setup
Backend Technology

Grails

Quick setup
No XML
Extensible

Web Services
Permissions
Database integration

{ REST }

Backend Technology

Grails

Quick setup
No XML
Extensible

Web Services
Permissions
Database integration

{ REST }
Grails Server

- Quick Setup
- Plugins
- Scaleable (Spring / Hibernate / JVM)
- Rapid Application Development
- Integrated testing
- Existing JBrowse tracks

DataStore
- Thread-safe transactions
- POJO (No XML)
- Multiple DB support
- Structured

Clients
Sidebar Client Criteria

- Integrate with JBrowse
- Uses exposed REST
- Build your own interface
- Build Complex UI
- Well-supported
- Is JavaScript at some level

Of the Million Available Options
- Angular + Bootstrap
- jQuery all the way
- GWT + Bootstrap
- Dojo
- Other Ideas: React, Polymer, etc.
Frontend Technology

• Sidebar - GWT (Java -> JS)
  • Gmail, Google Sheets, etc.
  • Built over existing web services
  • Allows significant complexity

• Can use without sidebar
• Roll your own front-end
GWT Sidebar Client

JBrowse Genomic Viewer

- Websockets improved reliability / performance

GWT

- Java -> JS Optimized Cross-Compiler
- Google uses for Complex UI (Gmail, Sheets, Adwords)
- Uses exposed REST
- Build your own interface (more here)
- Can still use JS / HTML wherever
Projection: What’s it Do?

Removing introns increases information

View Scaffolds Together
- View highly fragmented scaffolds together

View Genes across Scaffolds Together
- Annotate Genes across fragmented scaffolds
How to get Apollo into your WorkFlow?