

# The Apollo Genome Annotation and Curation Tool



**Nomi L. Harris** ([nomi@fruitfly.org](mailto:nomi@fruitfly.org)), **Suzanna Lewis**,  
**Mark Gibson**, **Colin Weil**, **John Richter**  
Berkeley Drosophila Genome Project

**Stephen Searle**, **Michele Clamp**  
Ensembl Project, Sanger Institute

*Bioinformatics Open Source Conference (BOOSC2002)*  
August 2, 2002



# What is Apollo?

- Open source tool for viewing and making annotations on genomic sequence
- Annotation editor for GMOD
- Gbrowse provides simpler way to view (but not edit) annotations
- Apollo lets human experts (biologists) create and refine annotations
- Developed by FlyBase-BDGP and Ensembl
  - Ensembl: standalone human annotation browser
  - BDGP: Drosophila annotation editor

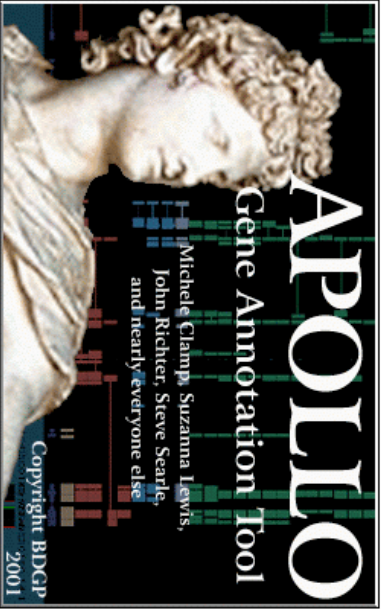
# Advantages

- Integrate large amounts of data in visually efficient manner
- Browser and editor
- Customizable
- Open Source, extensible
- Multi-platform (Java)
- Install from Web or CD

# Installer

Location: <http://www.fruitfly.org/annot/apollo/download>

## Download Apollo Genome Annotation Tool







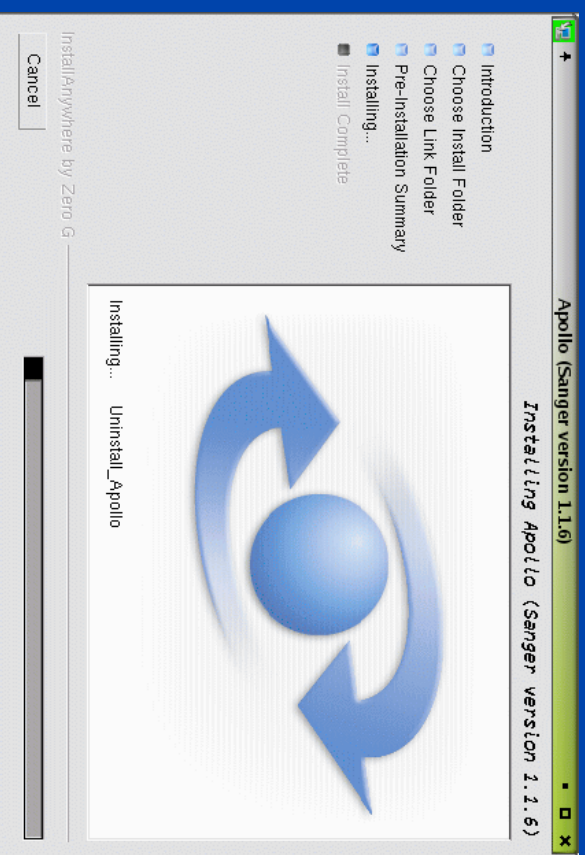
Alpha Version 1.1.7 (July 29, 2002)  
Requires JDK 1.2 or 1.3

Recommended Installation for Your Platform:  
 Start Installer for Linux...  Include VM in download

Installer created with [InstallAnywhere](#)® by Zero G Software, Inc. Copyright 2002. www.ZeroG.com

### Available Installers

Platform	Includes Java VM	without Java VM	Instructions
 Windows	<a href="#">Download (16.5M)</a>	<a href="#">Download (8.2M)</a>	<a href="#">View</a>
 Mac OS X	<a href="#">Download (7.9M)</a>	<a href="#">Download (8.3M)</a>	<a href="#">View</a>
 Linux	<a href="#">Download (30.2M)</a>	<a href="#">Download (8.3M)</a>	<a href="#">View</a>
 UNIX	Any Unix Platform	<a href="#">Download (8.3M)</a>	<a href="#">View</a>



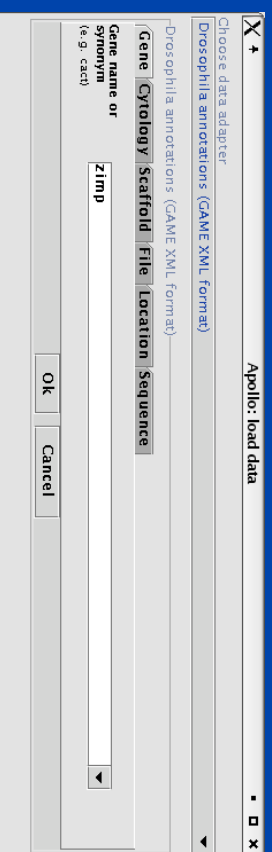
Apollo (Sanger version 1.1.6)  
Installing Apollo (Sanger version 1.1.6)

Installing... Uninstall\_Apollo

Cancel

# Loading annotation data

- Where do you want to go?
  - Data source (Drosophila, Ensembl human, etc.)
  - Chromosome arm, start, end
  - Cytological band
  - Gene
  - Sequence
  - Filename or URL



- Transparent access to GadFly annotation database via CGI
- Navigation bar



# Main display

The screenshot displays a genomic browser interface. At the top, the window title is "AE003603.Sept from 3R.3:969901-1262360". The menu bar includes "File", "Edit", "View", "Tiers", "Analysis", "Bookmarks", "Gene", "Window", and "Help". The "Chromosome" dropdown is set to "3R". The "Start" position is 969901 and the "End" position is 1262360. The "Expand" and "Load" buttons are visible. The main display area shows a chromosome view with a scale from 1.0Mb to 1.25Mb. Below this, there are several tracks showing genomic features. A detailed view of the gene "RH03539:contig1" is shown, including its structure and associated data.

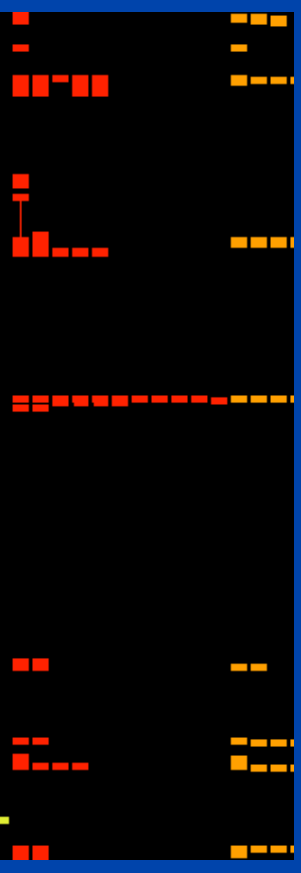
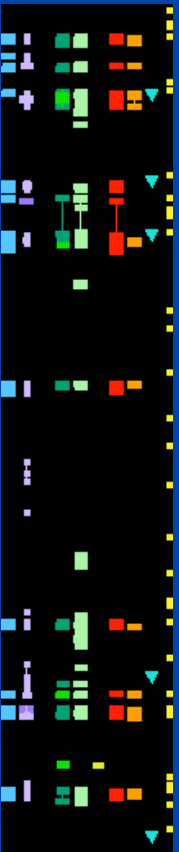
Type	Name	Range	Score
Gene	21125.2...	991026 - 9... 687...	
DGC	RH03539...	976627 - 9... 100.0	
FLY EST	LD24906...	999325 - 1... 100.0	
Relea...	CT34444.2	991026 - 9... 100.0	
Trans...	JOSHTRA...	993297 - 9... 0.0	

Name	score	descripti...	Genomic...	Match R...	Genomic...	Match Le...
5prime 520 bp, TN_G9325.28 735 bp, TN_G9325.41 686 bp, TN_G9325.37 694 bp, TN_G9325.45 694 bp, TN_G9325.44 602 bp						
:734724... 99.0	5prime ...	976627 ...	1 - 379	379	379	
:734724... 100.0	5prime ...	979242 ...	380 - 520	141	141	
:734724... 99.0	TN_G93...	995431 ...	521 - 930	410	410	

Position: 1167484 Feature: Action

# How much detail?

- Show/hide data types
- Change score cutoffs
- Expand/collapse tiers



- Text info detail panels
  - IDs hyperlinked to Web pages
- Semantic zooming
  - Start/stop codons
  - Sequence





# Many ways to customize

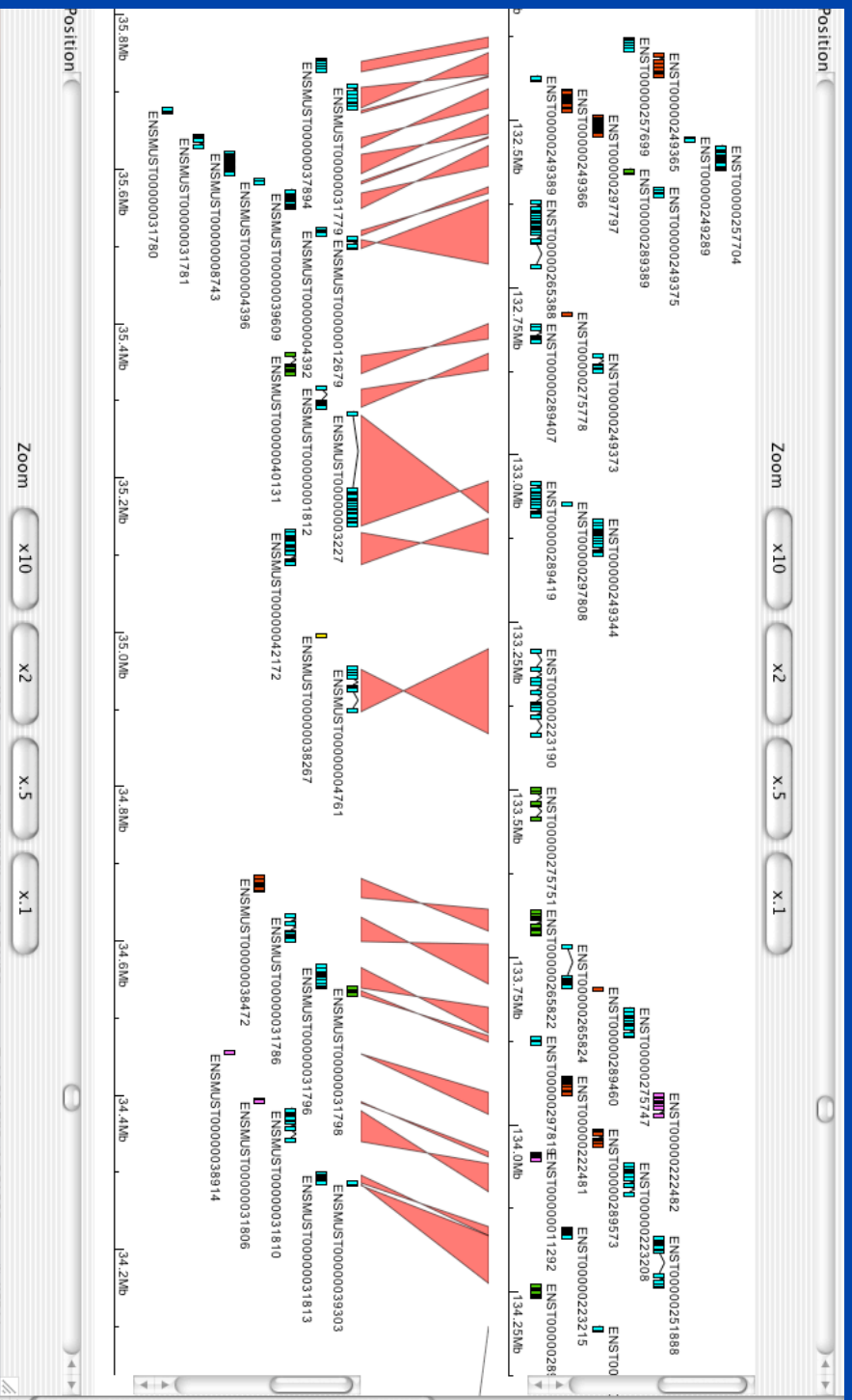
- Config file
  - Background color, 3D rectangles, user annotation colors, etc.
- Types panel
  - Show, hide, set cutoffs
  - Color chooser
- Drag tiers
- Define new data types in config file





# Synteny Viewer

- Cross-species comparisons (mouse/human)



# Data adapters

- Parser -> data models -> display
- Existing data adapters
  - GAME XML
  - GFF
  - Ensembl CGI server
  - DAS
- Write your own data adapter!
  - Extend AbstractDataAdapter class
  - Display options defined in config file

# How other groups are using Apollo

- Bristol-Myers Squibb
  - Launching Apollo from web browser via mime types
- GNF
  - JDBC adapter layer over BioSQL
- Biogen
  - View human genome alignment between public and Biogen internal database
  - Connected BLAT pipeline to Apollo
- HGMP-RC Fugu Genomics group
  - Displaying annotations on fugu scaffolds

# Coming (very) soon

- Style preferences per data adapter
- Intron editor
- Load & layer multiple datasets
- Dynamic analysis launch (BLAST, etc.)
- DAS adapter (almost works)
- GenBank adapter
  - Can write GenBank but not read

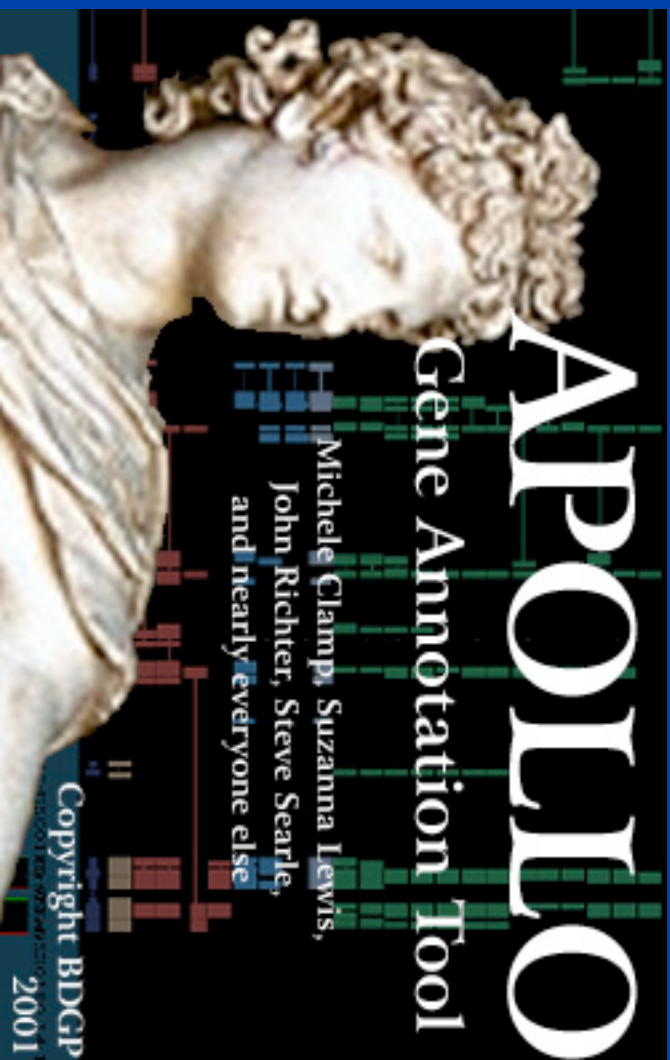
# Conclusions

- **It works:** ten biologists used Apollo to annotate 120Mb *Drosophila melanogaster* genome
- User-customizable
- Modular and extensible
- Continuing to evolve and improve

# Acknowledgements

- FlyBase-BDGP
  - Software: Suzanna Lewis, Mark Gibson, Nomi Harris, Colin Weil, Chris Mungall
  - Biologists: Sima Misra, Jon Tuppy, Simon Prochnik, Josh Kaminker, Chris Smith
- Ensembl: Steve Searle, Michele Clamp
- Howard Hughes Medical Institute/BDGP: John Richter (Apollo and org.bdgp)
- Tim





- Code available at SourceForge: [www.sourceforge.net/gmod](http://www.sourceforge.net/gmod)
- Installer available at [www.fruitfly.org/annot/apollo/download.html](http://www.fruitfly.org/annot/apollo/download.html)