

An Introduction to Galaxy

Daniel Blankenberg

Postdoctoral Research Associate

The Galaxy Team

<http://UseGalaxy.org>

Overview

- What is Galaxy?
- Galaxy for Experimental Biologists
- Galaxy for Bioinformaticians

Galaxy, a web-based genome analysis platform

- An open-source **framework** for integrating various computational tools and databases into a cohesive workspace
- A web-based **service** we provide, integrating many popular tools and resources for comparative genomics
- A completely **self-contained application** for building your own **Galaxy** style sites

Overview

- What is Galaxy?
- Galaxy for Experimental Biologists
- Galaxy for Bioinformaticians

Galaxy: the one-stop shop for Genome Analysis

- Analyze
 - Retrieve data directly from popular data resources or upload your own
 - Interactively manipulate genomic data with a comprehensive and expanding “best-practices” toolset
- Visualize
 - Send data results to external Genome Browsers
 - Build reusable AJAX-based custom Genome Browsers ([Trackster](#))
- Publish and Share
 - Results and step-by-step analysis record ([Data Libraries](#) and [Histories](#))
 - Customizable pipelines ([Workflows](#))
 - Complete protocols ([Pages](#))

Galaxy's Analysis Interface

The screenshot displays the Galaxy web interface. At the top, the navigation bar includes 'Galaxy' and menu items: 'Analyze Data', 'Workflow', 'Shared Data', 'Visualization', 'Admin', 'Help', and 'User'. On the left, a 'Tools' sidebar lists various categories like 'Get Data', 'Send Data', 'ENCODE Tools', 'Filter and Sort', and 'Statistics'. The main workspace shows the configuration for the 'VCF to MAF Custom Track' tool. The 'Custom Track Name' is 'Galaxy Custom Track by Population'. The 'VCF Source Source Type' is 'Per Population (file)'. Under 'VCF population files', 'VCF population file 1' is selected with 'VCF file: 8: Concatenate queri.. and data 7' and 'Name for this population: CEU_SRP00003'. There are buttons for 'Remove VCF population file 1', 'Add new VCF population file', and 'Execute'. Below the configuration, a 'What it does' section explains the tool's function, followed by an 'Example' section with a VCF file snippet. The right sidebar shows a 'History' panel with a list of previous jobs, including '10: VCF to MAF Custom Track on data 8' and '1: UCSC Main on Human: refGene (chr21:1-46944323)'. At the bottom of the history panel, a table lists genomic regions.

Tools Options ▾

- Get Data
- Send Data
- ENCODE Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- FASTA manipulation
- Filter and Sort
- Join, Subtract and Group
- Extract Features
- Fetch Sequences
- Fetch Alignments
- Get Genomic Scores
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
 - Histogram of a numeric column
 - Scatterplot of two numeric columns
 - Plotting tool for multiple series and graph types
 - Boxplot of quality statistics
 - GMAI Multiple Alignment Viewer
 - Build custom track for UCSC genome browser
 - VCF to MAF Custom Track for display at UCSC
- Regional Variation
- Multiple regression
- Multivariate Analysis
- Evolution
- Metagenomic analyses
- EMBOSS
- NGS TOOLBOX BETA
- NGS: QC and manipulation
- NGS: Mapping
- NGS: SAM Tools
- NGS: Indel Analysis
- NGS: Peak Calling
- RGENTICS
- SNP/WGA: Data; Filters
- SNP/WGA: QC; LD; Plots
- SNP/WGA: Statistical Models

Analyze Data Workflow Shared Data Visualization Admin Help User

VCF to MAF Custom Track

Custom Track Name:
Galaxy Custom Track by Population

VCF Source Source Type:
Per Population (file) ▾

VCF population files

VCF population file 1

VCF file:
8: Concatenate queri.. and data 7 ▾

Name for this population:
CEU_SRP00003

Remove VCF population file 1

Add new VCF population file

Execute

What it does

This tool converts a Variant Call Format (VCF) file into a Multiple Alignment Format (MAF) custom track file suitable for display at genome browsers.

This file should be used for display purposes only (e.g as a UCSC Custom Track). Performing an analysis using the output created by this tool as input is not recommended; the source VCF file should be used when performing an analysis.

Unknown nucleotides are represented as '*' as required to allow the display to draw properly; these include e.g. reference bases which appear before a deletion and are not available without querying the original reference sequence.

Example

Starting with a VCF:

```
##fileformat=VCFv3.3
##fileDate=20090805
##source=nyImputationProgramV3.1
##reference=1000GenomesPilot-NCBI36
##phasing=partial
##INFO=NS,1,Integer,"Number of Samples With Data"
##INFO=DP,1,Integer,"Total Depth"
##INFO=AF,-1,Float,"Allele Frequency"
##INFO=AA,1,String,"Ancestral Allele"
##INFO=DB,0,Flag,"dbSNP membership, build 129"
##INFO=H2,0,Flag,"HapMap2 membership"
##FILTER=q10,"Quality below 10"
##FILTER=s50,"Less than 50% of samples have data"
##FORMAT=GT,1,String,"Genotype"
##FORMAT=GQ,1,Integer,"Genotype Quality"
##FORMAT=DP,1,Integer,"Read Depth"
##FORMAT=HQ,2,Integer,"Haplotype Quality"
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT NA00001 NA00002 NA00003
20 14370 rs6054257 G A 29 0 NS=3;DP=14;AF=0.5;DB;H2 GT:GQ:DP:HQ 0|0:48:1:51,51 1|0:48:8:51,51 1/1:43:5:-1,-1
20 17330 . T A 3 q10 NS=3;DP=11;AF=0.017 GT:GQ:DP:HQ 0|0:49:3:58,50 0|1:3:5:65,3 0/0:41:3:-1,-1
20 1110696 rs6040355 A G,T 67 0 NS=2;DP=10;AF=0.333,0.667;AA=T;DB GT:OQ:DP:HQ 1|2:21:6:23,27 2|1:2:0:18,2 2/2:
20 1230237 . T . 47 0 NS=3;DP=13;AA=T GT:GQ:DP:HQ 0|0:54:7:56,60 0|0:48:4:51,51 0/0:61:2:-1,-1
20 1234567 microsat1 G D4,IGA 50 0 NS=3;DP=9;AA=G GT:OQ:DP 0/1:35:4 0/2:17:2 1/1:40:3
```

Under the following conditions: VCF Source type: Per Population (file), Name for this population: CHB+JPT Results in the following MAF custom track:

```
track name="Galaxy Custom Track" visibility=pack
```

History Options ▾

Intersecting VCF with Coding Exons Demo

- 10: VCF to MAF Custom Track on data 8
- 9: VCF to MAF Custom Track on data 8
- 8: Concatenate queries on data 3 and data 7
- 7: Cut on data 6
- 6: Intersect on data 5 and data 1
- 5: Compute on data 4
- 4: Compute on data 2
- 3: Select on data 2
- 2: 2010_03/pilot1 /CEU.SRP000031.2010_03.genotype!
- 1: UCSC Main on Human: refGene (chr21:1-46944323)
3,651 regions, format: bed, database: hg18
Info: UCSC Main on Human: refGene (chr21:1-46944323)
| display at UCSC main | view in GeneTrack | display at Ensembl May 2009

1. Chrom	2. Start	3. End	4. Name
chr21	9928775	9928911	NM_199260_cds
chr21	9930695	9930766	NM_199260_cds
chr21	9932177	9932270	NM_199260_cds
chr21	9936233	9936313	NM_199260_cds
chr21	9938240	9938346	NM_199260_cds
chr21	9941954	9942035	NM_199260_cds

Tools and Datasources

• Datasources

- Upload File from your computer
- UCSC table browser
- BioMart
- interMine / modMine
- EuPathDB server
- EncodeDB at NHGRI
- EpiGRAPH server

• Tool Suites

- Text Manipulation
- Format Converters
- Filtering and Sorting
- Join, Subtract, Group
- Sequence Tools
- Multiple-species Alignment Tools
- Genomic Interval Operations
- Summary Statistics
- Graphing / Plotting
- Regional Variation
- EMBOSS
- Evolution / Phylogeny
- NGS
- RGenetics
- ...and more

Visualize

- Send data results to external Genome Browsers
- Build reusable and sharable custom Genome Browsers (**Trackster**)

External Genome Browsers

- UCSC
- Ensembl
- GBrowse
- Adding more is easy!
 - <https://bitbucket.org/galaxy/galaxy-central/wiki/ExternalDisplayApplications/Tutorial>

Visualize at UCSC

14: Tag Counts (bigWig)   

2.4 Gb, format: bigwig, database: mm9

Info:

[display at UCSC main](#)

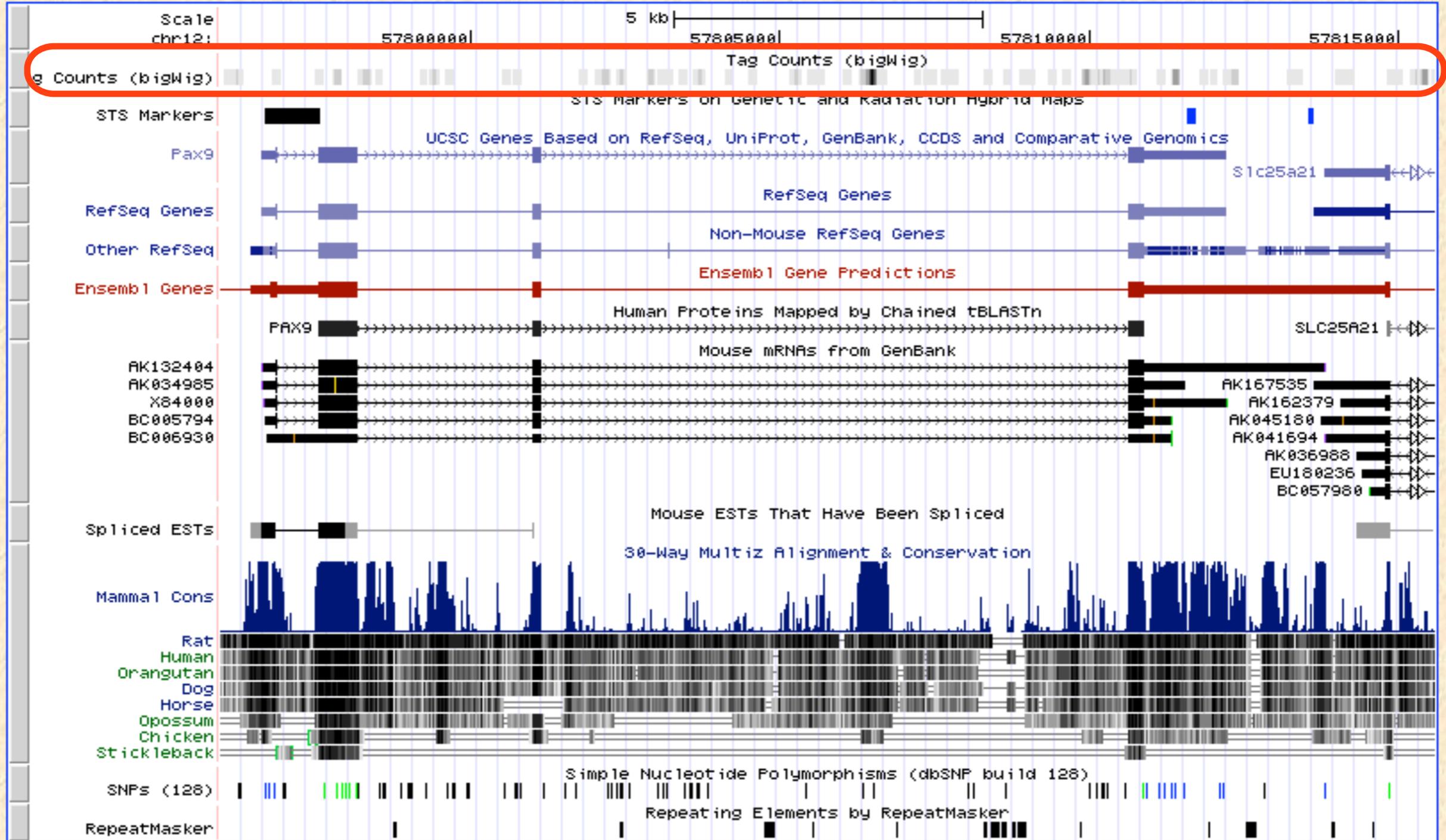
Binary UCSC BigWig file

UCSC Genome Browser on Mouse July 2007 (NCBI37)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out

position/search

chr12 (qC1) 12qA1.1 qA2 12qA3 qB1 12qB3 12qC1 qC2 12qC3 qD1 D2 12qD3 12qE 12qF1 qF2



Trackster

- Track/data viewer in web browser
 - HTML5 Canvas, jQuery
 - Renders in browser, not on server
- View your data from within Galaxy
 - No file transfers to third party
 - Use it locally, even without internet access
- Fast, responsive, interactive UI

Wig, Bedgraph (Line Tracks)



Regular line graph display

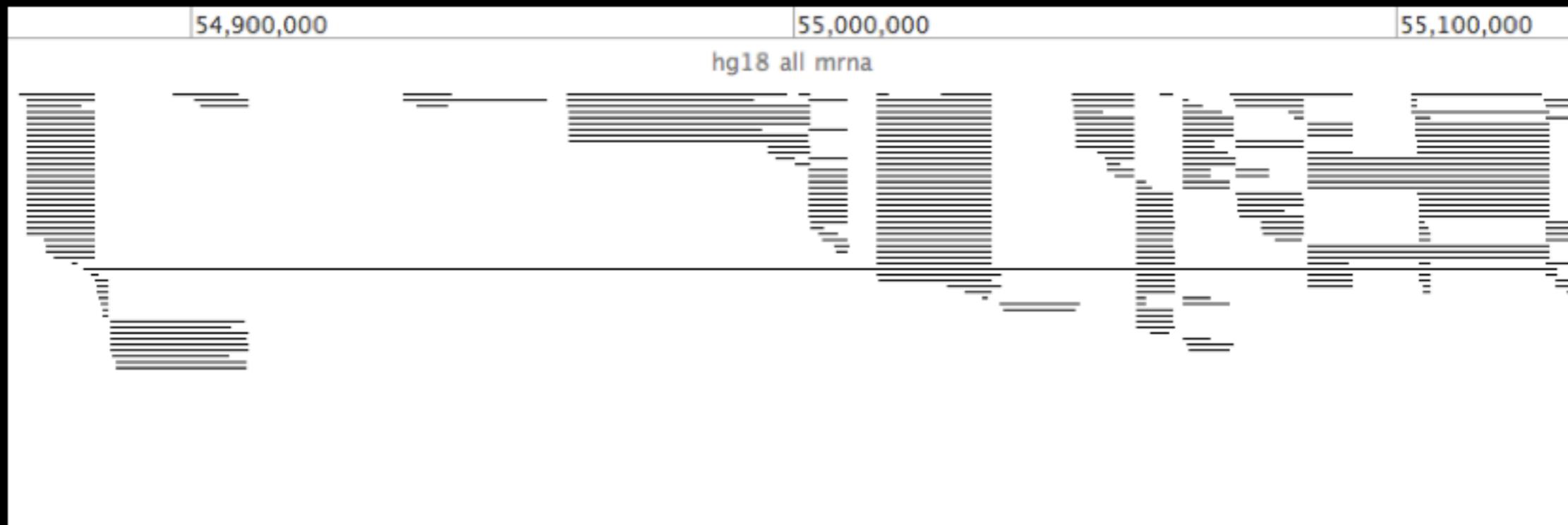


Intensity display (shades of gray)



Filled line graph display

Bed (Feature Track)

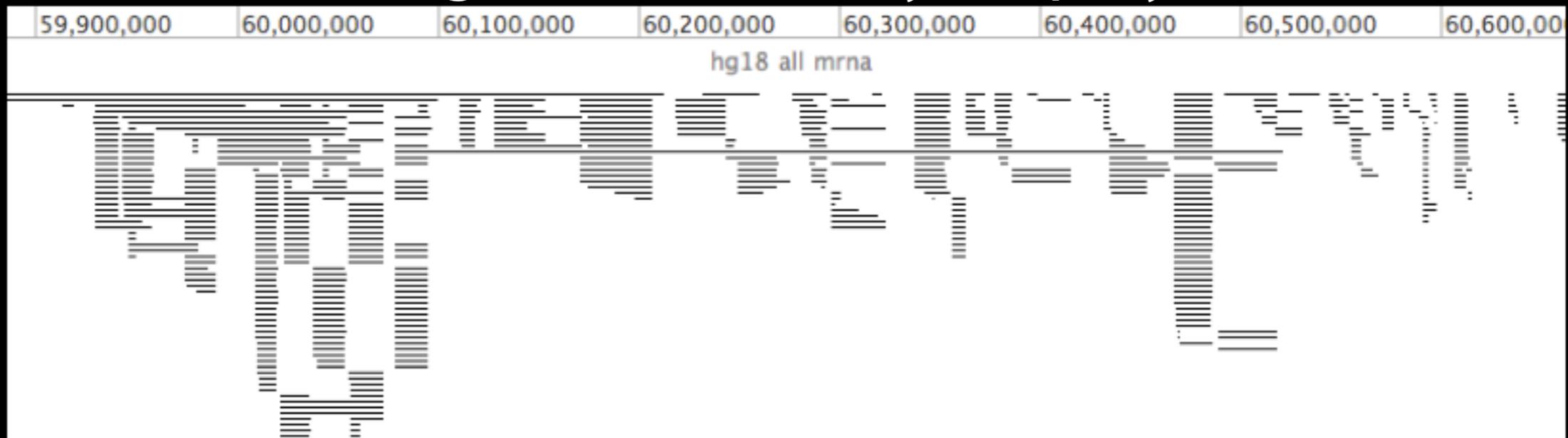


Snippet of hg18 all_mrna feature track

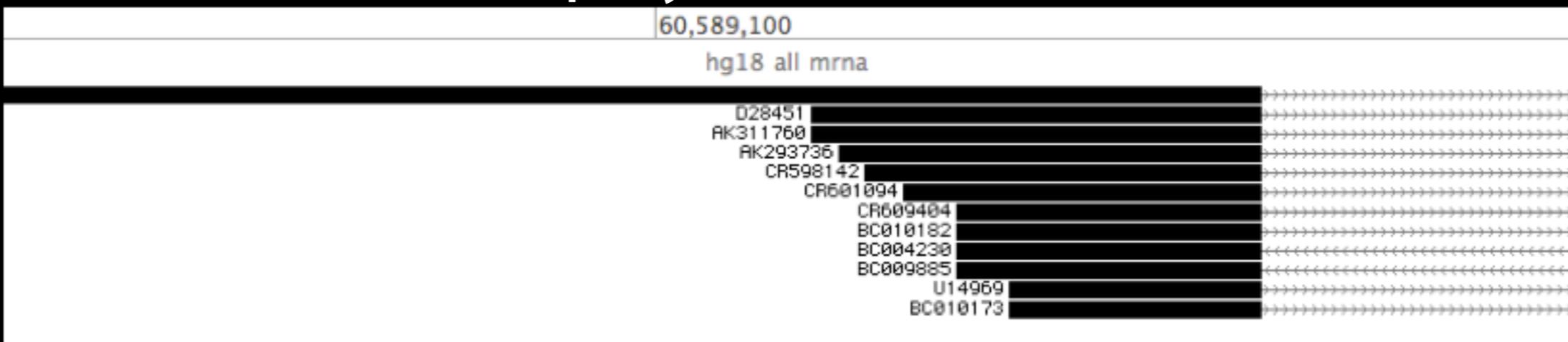
3 levels of detail: automatically adjusts based on what can fit on the screen



High level density display

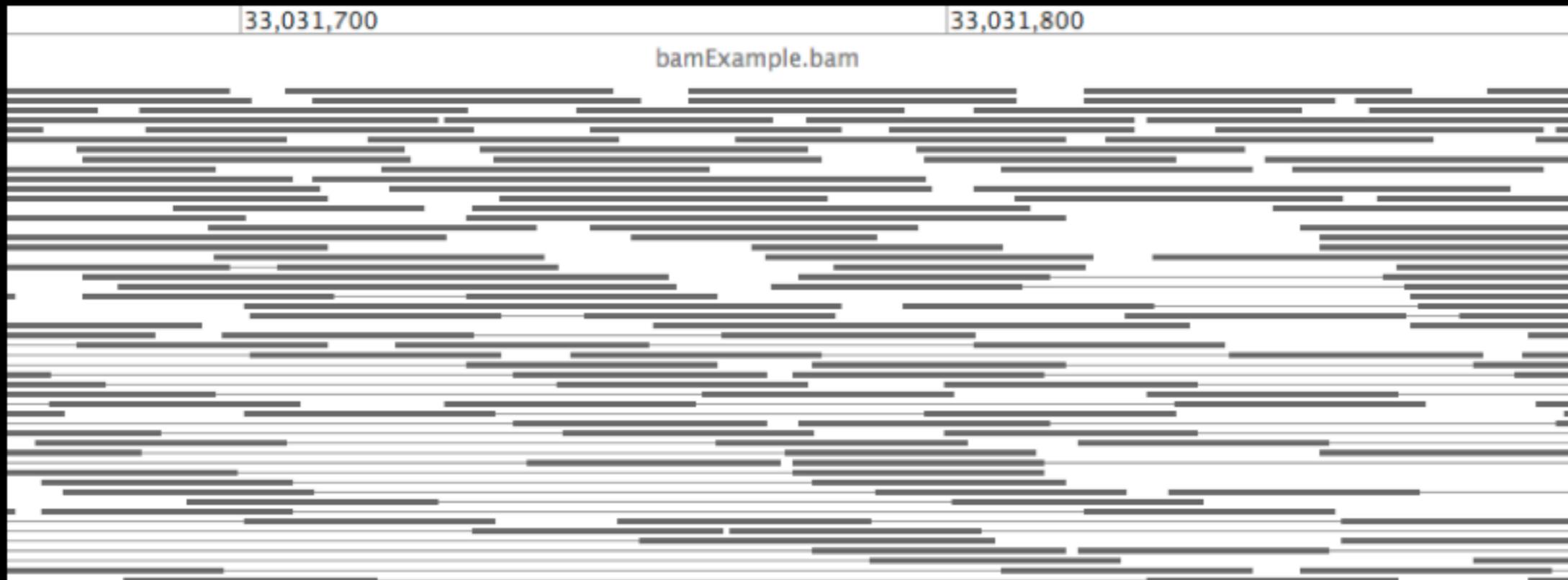


Feature display with no labels/detail



Feature display with labels, intron indicators, exon indicators

BAM (Aligned Reads)



33,031,740	33,031,750	33,031,760	33,031,770	33,031,780	33,031,790	33,031,800	33,031,810
bamExample.bam							
ATCTGACTTCCTACATT							
ATCTGACTTCCTACATTAACCTAACCAATATACCTAAT							
ATCTGACTTCCTACATTAACCTAACCAATAT							
ATCTGACTTCCTACATTAACCTAACCAATATACCTAATTTTTTA							
AAAAAACTTCATACATTAACCTAACCAATATACCTAATTTTTTACC							
ACATTAACCTAACCAATATACCTAATTTTTTCCCC							
ATCTGACTTCCTACATTAACCTAACCAATA							
208BEAAXX:7:23:910:1072 CATTAACTAACCAATATACCTAATTTTTTACCCCA							
SRR001117.6801930 CATTAACTAATTCATATACCTAATTTTTTACCCCAACCAATGGGGTT							
SRR001753.175081 CTAACCAATATACCTAATTTTTTACCCCAACCN							
-XAH_0001_FC2037KAAXX:7:328:422:542 ATATACCTAATTTTTTACCCCAACCAATGGGGTTAGGCACTCCCTCC							
CCTAATTTTTTACCCCAACCAATGGGGTTAGGCACTC							
-XAT_0001_FC208BFAAXX:8:282:898:586 TAAATTTTTTACCCCAACCAATGGGGTTAGGCACTCAAGCAAA							
-XAF_0002_FC205V7AAXX:7:112:400:398 ATAGACCGAATTTTTTACCCCAACCAATGGGGTTAGGCACTCAAGCA							
ERR001269.11128388 CTATTTTTTTTTTTCGTCTTTCTTCTCGCACTGAAAG							
L-XAT_0005_FC208EJAAXX:5:92:935:293 TTTTTACCCCAACCAATGGGGTTAGGCACTCAAGCAAA							
S322_0002_FC208CCAAXX:6:83:991:1993 AACCAATGGGGTTAGGCACTCAAGCAAA							

Data Libraries

Data Library "Bushman"

Library Actions ▾

These are the data underlying the analyses reported in the paper "Complete Khoisan and Bantu genomes from southern Africa" by S. C. Schuster et al., published in the journal Nature, February 18, 2010. Each data set can be downloaded and/or imported into a Galaxy history. Data will be updated as the project progresses.

Name	Information	Uploaded By	Date	File Size
<input type="checkbox"/> All SNPs in personal genomes ▾	Summary table of SNPs in all individuals	greg@bx.psu.edu	2010-01-28	676.8 Mb
<input type="checkbox"/> Alu insertions in KB1 ▾		greg@bx.psu.edu	2010-02-10	14.9 Kb
<input type="checkbox"/> Alu insertions in NB1 ▾		greg@bx.psu.edu	2010-02-10	6.5 Kb
<input type="checkbox"/> KB1 microsatellites.txt ▾		greg@bx.psu.edu	2010-02-15	3.5 Mb
<input type="checkbox"/> NB1 microsatellites.txt ▾		greg@bx.psu.edu	2010-02-15	828.5 Kb
<input type="checkbox"/> amino acid differences with functional predictions ▾		greg@bx.psu.edu	2010-02-05	1.1 Mb
<input type="checkbox"/> gene copy numbers in KB1 and other personal genome ▾		greg@bx.psu.edu	2010-02-15	2.1 Mb
<input type="checkbox"/> indels in ABT ▾		greg@bx.psu.edu	2010-02-03	105.3 Kb
<input type="checkbox"/> indels in KB1 ▾		greg@bx.psu.edu	2010-02-03	14.2 Mb
<input type="checkbox"/> indels in MD8 ▾		greg@bx.psu.edu	2010-02-03	109.8 Kb
<input type="checkbox"/> indels in NB1 ▾		greg@bx.psu.edu	2010-02-03	519.5 Kb
<input type="checkbox"/> indels in TK1 ▾		greg@bx.psu.edu	2010-02-03	123.2 Kb
<input type="checkbox"/> novel SNPs in ABT ▾		greg@bx.psu.edu	2010-02-09	9.4 Mb
<input type="checkbox"/> novel SNPs in KB1 ▾		greg@bx.psu.edu	2010-02-09	16.9 Mb
<input type="checkbox"/> novel SNPs in MD8 ▾		greg@bx.psu.edu	2010-02-09	594.1 Kb
<input type="checkbox"/> novel SNPs in NB1 ▾		greg@bx.psu.edu	2010-02-09	4.1 Mb
<input type="checkbox"/> novel SNPs in TK1 ▾		greg@bx.psu.edu	2010-02-09	722.6 Kb
<input type="checkbox"/> sequenced exon-containing intervals ▾		greg@bx.psu.edu	2010-02-03	3.1 Mb

For selected items: ▾

<http://usegalaxy.org/bushman>

Managing Libraries

- Loading Data
 - Upload a single file
 - Import datasets from a Galaxy history
 - Upload a directory of files
 - Directly from Sequencer using **Sample Tracking System**
- Accessing Data
 - **Data contents on disk are not copied**
 - Dataset security
 - Public
 - Role-based access control (RBAC)
- Annotating Library Data: Library Templates
 - Build user fillable forms
 - Associate at Library, Folder or Dataset level

Workflows

The screenshot displays the Galaxy workflow editor interface. The main canvas shows a workflow titled "Clone of 'metagenomic analysis' shared by 'anton@bx.psu.edu'". The workflow consists of several interconnected steps:

- Input dataset** (output) feeds into **Select high quality segments** (output1 (fasta)).
- Input dataset** (output) feeds into **FASTA-to-Tabular** (output (tabular)).
- FASTA-to-Tabular** feeds into **Add column** (out_file1).
- Add column** feeds into **Tabular-to-FASTA** (output (fasta)).
- Tabular-to-FASTA** feeds into **Megablast** (output1 (tabular)).
- Select high quality segments** feeds into **Megablast** (output1 (tabular)).
- Megablast** feeds into **Concatenate queries** (out_file1).
- Concatenate queries** feeds into **Join two Queries** (out_file1).
- Join two Queries** feeds into **Filter** (out_file1).
- Filter** feeds into **Fetch taxonomic representation** (out_file1 (taxonomy)).
- Fetch taxonomic representation** feeds into **Summarize taxonomy** (out_file1 (tabular)).
- Summarize taxonomy** feeds into **Draw phylogeny** (out_file1 (pdf)).
- Draw phylogeny** feeds into **Find lowest diagnostic rank** (out_file1 (taxonomy)).
- Find lowest diagnostic rank** feeds into **Draw phylogeny** (out_file1 (taxonomy)).

The right sidebar shows the details for the selected "Draw phylogeny" tool, including options for "show ranks from root to", "Class", "select font size", "maximum number of leaves", and "Edit Step Actions".

<http://main.g2.bx.psu.edu/u/aun1/w/metagenomic-analysis>

Pages

Galaxy Analyze Data Workflow Shared Data Visualization Admin Help User

Published Pages | aun1 | Windshield Splatter

Windshield splatter analysis with the Galaxy metagenomic pipeline: A live supplement

SERGEI KOSAKOVSKY POND^{1,2,*}, SAMIR WADHAWAN^{3,6*}, FRANCESCA CHIAROMONTE⁴, GURUPRASAD ANANDA^{1,3}, WEN-YU CHUNG^{1,3,7}, JAMES TAYLOR^{1,5}, ANTON NEKRUTENKO^{1,3} and THE GALAXY TEAM^{1*}

Correspondence should be addressed to SKP, JW, or AN.

How to use this document

This document is a live copy of supplementary materials for [the manuscript](#). It provides access to the exact analyses and workflows discussed in the paper, so you can play with them by re-running, changing parameters, or even applying them to your own data. Specifically, we provide the two histories and one workflow found below. You can view these items by clicking on their name to expand them. You can also import these items into your Galaxy workspace and start using them; click on the green plus to import an item. To import workflows you must [create a Galaxy account](#) (unless you already have one) – a hassle-free procedure where you are only asked for a username and password.

This is the Galaxy history detailing the comparison of our pipeline to MEGAN:

Galaxy History | Galaxy vs MEGAN
Comparison of Galaxy vs. MEGAN pipeline.

This is the Galaxy history showing a generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3A**):

Galaxy History | metagenomic analysis

10: Concatenate queries on data 8 and data 7	Merge Megablast runs to produce a single dataset for reads compared to both WGS and NT
11: Join two Queries on data 9 and data 10	Combine sequence length data with results from Megablast runs
12: Filter on data 11	Filter suboptimal sequence alignments using the expression $(\text{sequence_alignment_length}/\text{sequence_read_length} > 0.5)$; sequences that do not meet this criterion are filtered out
13: Fetch taxonomic representation on data 12	Get taxonomic representation for filtered, aligned sequences
14: Find lowest diagnostic rank on data 13	Get reads specific to ranks below Kingdom level
15: Summarize taxonomy on data 13	Tabulate list of taxonomic groups contained in reads from dataset 14
16: Draw phylogeny on data 14	Build and draw phylogenetic tree from ranks in dataset 14

This is the Galaxy workflow for generic analysis of metagenomic data. (This corresponds to the "A complete metagenomic pipeline" section of the manuscript and **Figure 3B**):

Galaxy Workflow | metagenomic analysis
Generic workflow for performing a metagenomic analysis on NGS data.

Supplemental Analysis

Comparison between Galaxy pipeline and Megan

(Use [this link](#) to see Galaxy history representing this analysis. Individual elements of this history are referred to as **History Item1, 2 and so on** using bold typeface)

The first step of a homology-based metagenomic analysis is to contrast a collection of sequencing reads against a database whose entries are assigned to taxonomic ranks. Following the procedure of (Huson et al. 2007) we used the non-redundant protein database (NR) from the [National Center for Biotechnology Information](#). There are several avenues for importing large sets of alignments into Galaxy. First, alignments can be generated directly within Galaxy (see the following section). Alternatively, alignments generated elsewhere (e.g., using local BLAST installations) or from public repositories (e.g., SRA) can be imported into Galaxy. To demonstrate this feature, we imported alignments from SRA into Galaxy using the SRA tool.

About this Page

Author
aun1

Related Pages
[All published pages](#)
[Published pages by aun1](#)

Rating
Community (2 ratings, 5.0 average) ★★★★★
Yours ★★★★★

Tags
Community:
paper galaxy

Yours:

Sharing

The screenshot shows the Galaxy web interface. At the top, there is a navigation bar with the Galaxy logo and several menu items: Analyze Data, Workflow, Shared Data (highlighted), Visualization, Admin, Help, and User. A dropdown menu is open under 'Shared Data', listing: Data Libraries, Published Histories, Published Workflows, Published Visualizations, and Published Pages (which is highlighted by a mouse cursor). Below the navigation bar, there is a 'Published Pages' section with a search box and a link to 'Advanced Search'. The main content is a table with the following columns: Title, Annotation, Owner, Community Tags, and Last Updated. The table lists several published pages, including 'bushman', 'mtDEMO: Heteroplasmy', 'mtDEMO: Mapping Cheek Reads', 'mtDEMO: Estimating Error', 'mtDEMO: Getting Things Mapped', 'Finding Heteroplasmic Sites', 'FASTQ manipulation tools', 'Windshield Splatter', 'pe', 'NGS Analysis Service', and 'Screencasts'. Each row includes a title, a brief annotation, the owner's name, a star rating, community tags, and the last updated date.

Title	Annotation	Owner	Community Tags	Last Updated ↑
bushman		aun1	genomics paper nature	Jul 21, 2010
mtDEMO: Heteroplasmy	Part D of the mtDNA analysis demo	aun1		Jul 10, 2010
mtDEMO: Mapping Cheek Reads	Part C of the mtDNA analysis demo	aun1		Jul 10, 2010
mtDEMO: Estimating Error	Part B of the mtDNA analysis tutorial	aun1		Jul 10, 2010
mtDEMO: Getting Things Mapped	Part A of the mtDNA Analysis tutorial	aun1		Jul 10, 2010
Finding Heteroplasmic Sites		aun1		Jul 10, 2010
FASTQ manipulation tools	Supplementary material for FASTQ manipulation tools	dan		May 24, 2010
Windshield Splatter	Live supplement for Genome Research windshield splatter paper.	aun1	paper galaxy	Mar 19, 2010
pe		aun1	workflow pe	Mar 12, 2010
NGS Analysis Service	Description of Galaxy main's NGS services and tools.	aun1	screencasts ngs galaxy tutorial	Mar 06, 2010
Screencasts		aun1	screencasts galaxy help	Feb 17, 2010

Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25;11(8):R86.

Galaxy 101

<http://usegalaxy.org/galaxy101>

- A simple question...
- Which coding exons have highest number of single nucleotide polymorphisms?

Galaxy 101

<http://usegalaxy.org/galaxy101>

- Overview
 - Interactively Analyze Data
 - Create reusable generic Workflow
 - Share analysis Results, History, Workflow
- Required Data
 - Genomic Coordinates of
 - Coding Exons
 - SNPs

Galaxy 101

<http://usegalaxy.org/galaxy101>

Interactive Analysis Steps

- Get Genomic data from UCSC Table Browser
- Determine each SNP that overlaps with a specific coding exon
- Calculate count of overlapping SNPs for each exon
- Sort and select exons by greatest SNP counts

Overview

- What is Galaxy?
- Galaxy for Experimental Biologists
- Galaxy for Bioinformaticians

Galaxy: the instant web-based tool and data resource integration platform

- Open Source downloadable package that can be deployed in individual labs
- Zero Configuration, but highly configurable
- Modularized
 - Easy to plug in your own components
- Straightforward to run your own private Galaxy Server

Why a private Galaxy instance?

- Add new Tools
- Integrate new Data Sources
- Add new display applications
- Secure your private instance for working with sensitive data

The Problem

- You have written a Python script to analyze genomic data and you want to share it with command-line averse colleagues

The Galaxy Solution

- Solution: Integrate the script as a new Tool into your own Galaxy server
- Steps:
 - Obtain and install Galaxy source code (GetGalaxy.org)
 - Write an XML file describing the inputs and outputs and how to execute the script
 - Instruct Galaxy to load the tool

Quick Install

1. Get the latest copy from the repository:

The latest source code can be downloaded from the anonymous [Mercurial](#) repository with this command:

```
1 % hg clone http://www.bx.psu.edu/hg/galaxy galaxy_dist
```

If you don't have Mercurial, tarballs can be downloaded instead: [zipped](#), [bzipped](#) or [gzipped](#). However, this makes it more difficult to stay updated in the future since there's no simple way to update your copy.

2. Enable configuration files and download eggs:

Once the source code is downloaded, cd to the `galaxy_dist` directory and run the `setup.sh` script. This will copy sample configuration files and download the proper eggs for your platform:

```
1 % cd galaxy_dist
2 % sh setup.sh
```

This step requires Internet access to download the eggs. If the system on which you are installing Galaxy does not have Internet access, please follow the instructions for offline systems on [Config/Eggs](#) before attempting this step.

3. Start it up:

At this point Galaxy is ready to run. Simply run the following command:

```
1 % sh run.sh
```

This will start up the server on localhost and port 8080, so Galaxy can be accessed from your web browser at <http://localhost:8080> . To stop the Galaxy server, just hit `ctrl-c` in the terminal from which Galaxy is running.

Cluster

Cluster intervals of:

max distance between intervals: (bp)

min number of intervals per cluster:

Return type:

TIP: If your query does not appear in the pulldown menu -> it is not in interval format. Use "edit attributes" to set chromosome, start, end, and strand columns

Screencasts!

See Galaxy Interval Operation [Screencasts](#) (right click to open this link in another window).

Syntax

- **Maximum distance** is greatest distance in base pairs allowed between intervals that will be considered "clustered". **Negative** values for distance are allowed, and are useful for clustering intervals that overlap.
- **Minimum intervals per cluster** allow a threshold to be set on the minimum number of intervals to be considered a cluster. Any area with less than this minimum will not be included in the output.
- **Merge clusters into single intervals** outputs intervals that span the entire cluster.
- **Find cluster intervals; preserve comments and order** filters out non-cluster intervals while maintaining the original ordering and comments in the file.
- **Find cluster intervals; output grouped by clusters** filters out non-cluster intervals, but outputs the cluster intervals so that they are grouped together. Comments and original ordering in the file are lost.

Example



```
cluster.xml
1 <tool id="gops_cluster_1" name="Cluster">
2   <description>[[Cluster]] the intervals of a query</description>
3   <command interpreter="python2.4">
4     gops_cluster.py $input1 $output -1 $input1_chromCol,$input1_startC
5       -d $distance -m $minregions -o $returntype
6   </command>
7   <inputs>
8     <param format="interval" name="input1" type="data">
9       <label>Cluster intervals of</label>
10    </param>
11    <param name="distance" size="5" type="integer" value="1" help="(bp
12      <label>max distance between intervals</label>
13    </param>
14    <param name="minregions" size="5" type="integer" value="2">
15      <label>min number of intervals per cluster</label>
16    </param>
17    <param name="returntype" type="select" label="Return type">
18      <option value="1">Merge clusters into single intervals</option>
19      <option value="2">Find cluster intervals; preserve comments and
20      <option value="3">Find cluster intervals; output grouped by clus
21      <option value="4">Find the smallest interval in each cluster</opt
22      <option value="5">Find the largest interval in each cluster</opt
23    </param>
24  </inputs>
25  <help>
26
27  .. class:: infomark
28
29  **TIP:** If your query does not appear in the pulldown menu -> it is n
30
31  -----
32
33  **Screencasts!**
34
35  See Galaxy Interval Operation Screencasts (right click to open this l
36
37  .. \_Screencasts: http://www.bx.psu.edu/cgi-bin/trac.cgi/wiki/GopsDesc
38
39  -----
40
41  **Syntax**
42
43  - Maximum distance is greatest distance in base pairs allowed betw
44  - Minimum intervals per cluster allow a threshold to be set on the
45  - Merge clusters into single intervals outputs intervals that span
46  - Find cluster intervals; preserve comments and order filters out
47  - Find cluster intervals; output grouped by clusters filters out n
48
49  Line: 87 Column: 8 XML Soft Tabs: 2
```

Input Parameter types

Basic

- Text
- Integer
- Float
- Select
 - Static
 - Dynamic
- Boolean
- Genome build
- Data column
- Data
- Hidden
- Base URL
- File
- Drill down
- Grouping
 - Conditional
 - Repeat
- Config Files

Running a Production Server

- Use a *real* database server: PostgreSQL, MySQL
- Run on compute cluster resources
- External Authentication: LDAP, Kerberos, OpenID
- Load balancing; proxy support

Lack IT knowledge or resources?

- No problem, just use the **Cloud**

Galaxy on the Cloud

- Availability of Resources are not a Problem
 - Virtually unlimited resources: storage, computing, services
 - No need to maintain machines or personnel
 - Only pay for what you use
- Amazon Elastic Compute Cloud (EC2) and Eucalyptus
- **Web-based Galaxy instantiation**

Point, Click, Cloud

Galaxy Info: [report bugs](#) | [wiki](#) | [screencasts](#) [GC Home](#)

Galaxy Cloud Console

The Galaxy cloud console allows you to manage this instance of Galaxy. From here you can start the main Galaxy interface (including an initial set of "worker" nodes on which jobs will be run), as well as add and remove workers while the main interface is running.

Terminate Galaxy

Access Galaxy

Scale

+ Add more instances **- Remove idle instances**

Status

Cluster name: galaxy-cluster
Cluster status: Ready
Disk status: 59G / 100G (59%)
Instance status: Idle: 9 Available: 12 Requested: 12



i-5d065036
State: Ready
Alive: 11m 19s

- Filesystems
- Permissions
- JobScheduler

● Filesystems ● Database ● Scheduler ● Galaxy

Cluster status log

```
21:20:21 - Instance 'i-5d065036' ready
21:20:28 - Ready for use
21:20:29 - Instance 'i-59065032' ready
21:20:29 - Instance 'i-5f065034' ready
21:22:40 - Instance 'i-5b065030' ready
21:23:32 - Instance 'i-e9e8bf82' not responding, rebooting instance...
21:23:32 - Instance 'i-efe8bf84' not responding, rebooting instance...
21:23:32 - Instance 'i-ed8bf86' not responding, rebooting instance...
21:25:23 - INSTANCE_ALIVE private_dns:ip-10-243-21-219.ec2.internal
public_dns:ec2-174-129-174-158.compute-1.amazonaws.com zone:us-east-1d type:m1.large ami:ami-ed03ed84
21:25:23 - Sent master public key to worker instance 'i-e9e8bf82'.
21:25:29 - Waiting on worker instance 'i-e9e8bf82' to configure itself...
```

You added a tool, now what?

- Share it with the community!
- Galaxy Tool Shed
 - Upload and Download contributed tools
 - Rate and provide comments and feedback

Get and Contribute Tools

Galaxy Tool Shed / (beta) Tools Help User

Community

Tools

- [Browse by category](#)
- [Browse all tools](#)
- [Login to upload](#)

Categories

 [Advanced Search](#)

<u>Name ↓</u>	<u>Description</u>	<u>Tools</u>
Convert Formats	Tools for converting data formats	4
Data Source	Tools for retrieving data from external data sources	1
Fasta Manipulation	Tools for manipulating fasta data	5
Next Gen Mappers	Tools for the analysis and handling of Next Gen sequencing data	5
Ontology Manipulation	Tools for manipulating ontologies	1
SAM	Tools for manipulating alignments in the SAM format	0
Sequence Analysis	Tools for performing Protein and DNA/RNA analysis	7
SNP Analysis	Tools for single nucleotide polymorphism data such as WGA	1
Statistics	Tools for generating statistics	1
Text Manipulation	Tools for manipulating data	3
Visualization	Tools for visualizing data	1

<http://usegalaxy.org/community>

Using Galaxy

- Use public Galaxy server: UseGalaxy.org
- Download Galaxy source: GetGalaxy.org
- Galaxy Wiki: GalaxyProject.org
- Screencasts: GalaxyCast.org
- Public Mailing Lists
 - galaxy-bugs@bx.psu.edu
 - galaxy-user@bx.psu.edu
 - galaxy-dev@bx.psu.edu

Acknowledgments

- All Members of the Galaxy Team (see them at <https://bitbucket.org/galaxy/galaxy-central/wiki/GalaxyTeam>)
- Thousands of our users
- GMOD Team
- UCSC Genome Informatics Team
- BioMart Team
- FlyMine/InterMine Teams
- Funding sources
 - NSF-ABI
 - NIH-NHGRI
 - Beckman Foundation
 - Huck Institutes at Penn State
 - Pennsylvania Department of Public Health
 - Emory University

Galaxy Team



Enis Afgan | Emory



Guru Ananda | Penn State



Dannon Baker | Emory



James Taylor | Emory



Jeremy Goecks | Emory



Sergei Kosakovsky Pond | UCSD



Greg von Kuster | Penn State



Dave Clements | Emory



Nate Coraor | Penn State



Ross Lazarus | Harvard | BakerID



Kanwei Li | Emory



Anton Nekrutenko | Penn State



Kelly Vincent | Penn State

+ Jennifer
Jackson

Two full days of presentations,
workshops, and conversations by
and for Galaxy community members

Galaxy 2011



Community Conference

25-26 May Lunteren, The Netherlands

<http://galaxy.psu.edu/gcc2011>