BIOL 419/519 Bioinformatics Research
Gene Finding 2
Paul Szauter – June 13, 2013

Introduction

In this exercise, we will use a custom installation of the UCSC Genome Browser at GEP to view a *D. ananassae* fosmid. The UCSC Genome Browser allows a large variety of different data types to be displayed against a physical map of a fosmid or contig. We will be using the UCSC Genome Browser extensively for annotation. In this first view, we will use it to confirm and extend your analysis of the genes present on your fosmid or contig.

1. *D. ananassae* fosmid 1475K17

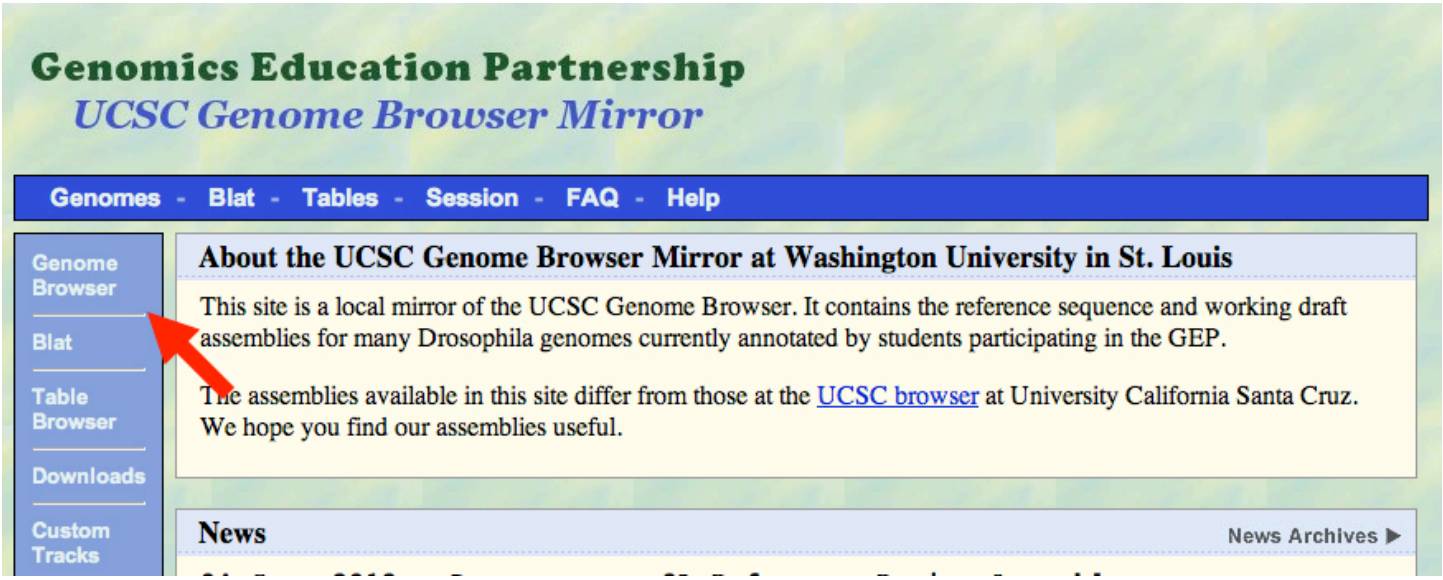
On the course website, navigate to the **Tools** page:

http://www.discoveryandinnovation.com/bioinformatics/tools.html

Click the link to the UCSC Genome Browser at GEP to go to:

http://gander.wustl.edu/

In the page that appears (shown below), click the **Genome Browser** link.



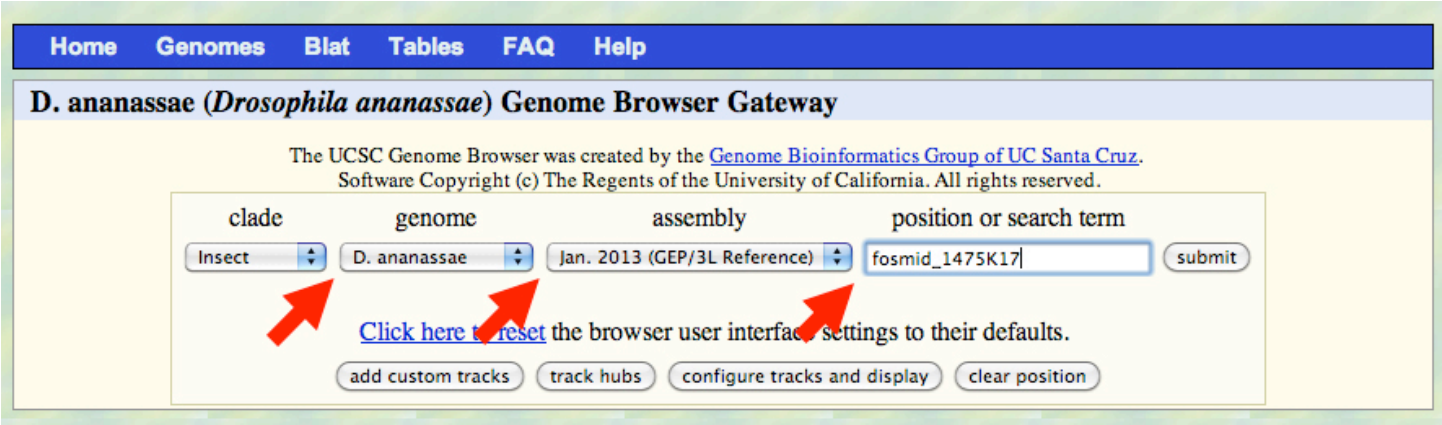
In the page that appears (shown below):

1. Select **D. ananassae** as the **Genome**.

2. Select **Jan. 2013 (GEP/3L Reference)** as the **assembly**.

3. Enter `fosmid_1475K17` as the **position or search term**.

4. Click **Submit**.



In the page that appears, there will be a visual display with many control buttons. Set the controls as described below.

1. In **Mapping and Sequencing Tracks**, set **Base Position** to full, all other tracks to **hide**.

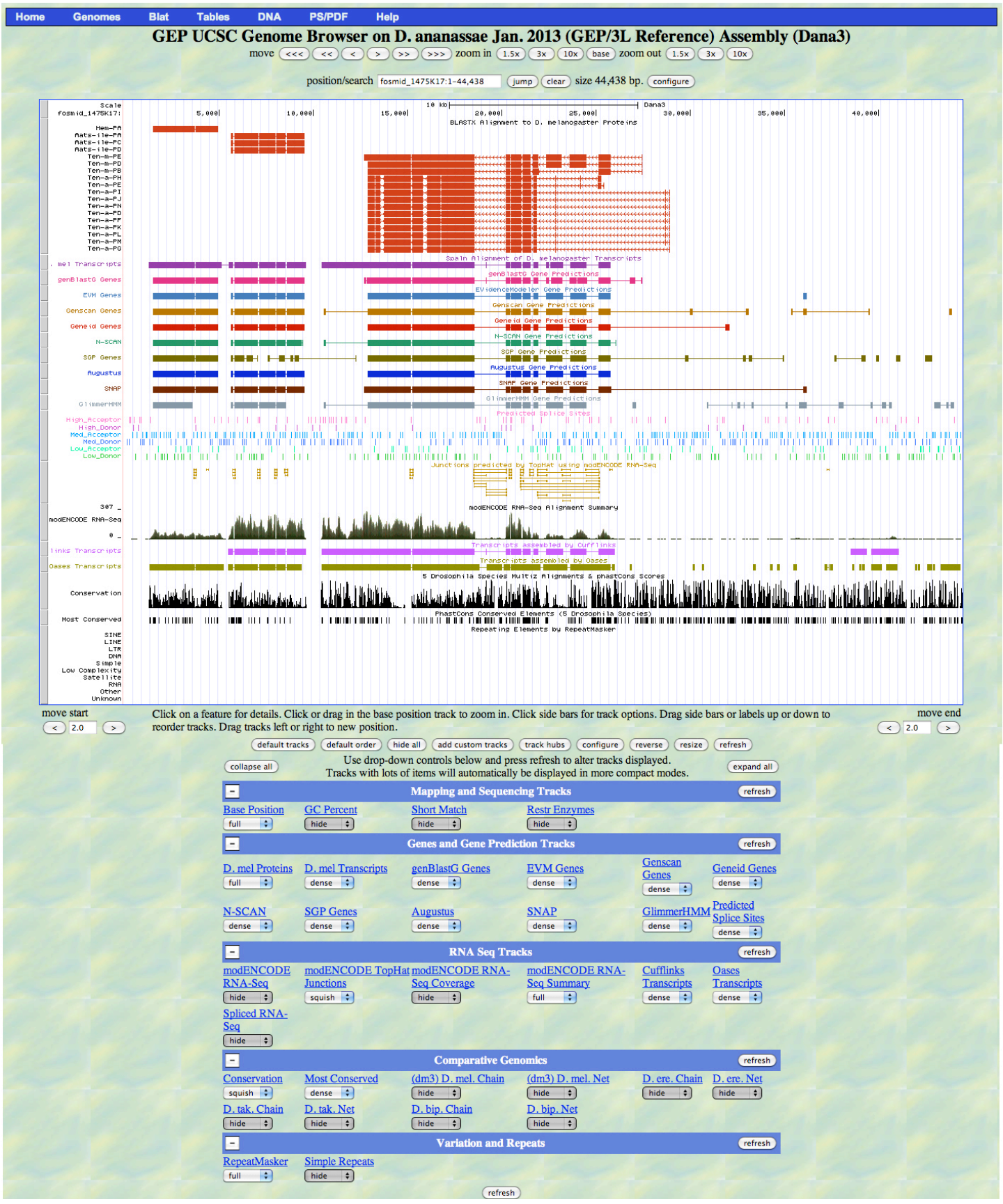
2. In **Genes and Gene Prediction Tracks**, set all to **dense**.

3. In **RNA Seq Tracks**, set **modENCODE TopHat** junctions to **squish**, **modENCODE RNA-Seq Summary** to **full**, **Cufflinks Transcripts and Oases Transcripts** to **dense**, and all others to **hide**.

4. In **Comparative Genomics**, set **Conservation** to **squish**, **Most Conserved** to **dense**, and all others to **hide**.

5. In **Variation and Repeats**, set **RepeatMasker** to **full** and **Simple Repeats** to **hide**.

Your display should now resemble the image on the next page.


Home
Genomes
Blat
Tables
DNA
PS/PDF
Help
GEP UCSC Genome Browser on D. ananassae Jan. 2013 (GEP/3L Reference) Assembly (Dana3)
move
zoom in
zoom out
position/search
fosmid_1475K17:1-44,438
jump
clear
size 44,438 bp.
configure
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
genBlastG Gene Predictions
EVidenceModeler Gene Predictions
Genscan Gene Predictions
Geneid Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
GlimmerHMM Gene Predictions
Predicted Splice Sites
Junctions predicted by TopHat using modENCODE RNA-Seq
modENCODE RNA-Seq Alignment Summary
Transcripts assembled by Cufflinks
Transcripts assembled by Oases
5 Drosophila Species Multiz Alignments & phastCons Scores
phastCons Conserved Elements (5 Drosophila Species)
Repeating Elements by RepeatMasker
move start
move end
Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.
default tracks
default order
hide all
add custom tracks
track hubs
configure
reverse
resize
refresh
collapse all
Use drop-down controls below and press refresh to alter tracks displayed.
Tracks with lots of items will automatically be displayed in more compact modes.
expand all
Mapping and Sequencing Tracks
Base Position
GC Percent
Short Match
Restr Enzymes
Genes and Gene Prediction Tracks
D. mel Proteins
D. mel Transcripts
genBlastG Genes
EVM Genes
Genscan Genes
Geneid Genes
N-SCAN
SGP Genes
Augustus
SNAP
GlimmerHMM
Predicted Splice Sites
RNA Seq Tracks
modENCODE RNA-Seq
modENCODE TopHat Junctions
modENCODE RNA-Seq Coverage
modENCODE RNA-Seq Summary
Cufflinks Transcripts
Oases Transcripts
Spliced RNA-Seq
Comparative Genomics
Conservation
Most Conserved
(dm3) D. mel. Chain
(dm3) D. mel. Net
D. ere. Chain
D. ere. Net
D. tak. Chain
D. tak. Net
D. bip. Chain
D. bip. Net
Variation and Repeats
RepeatMasker
Simple Repeats

CLASS DISCUSSION: We will discuss the meaning of the various tracks and how to use this display to confirm the results of your gene finding with BLASTX and GENSCAN.

2. Your fosmid or contig

Return to the starting page and set the UCSC Genome Browser to display your *D. ananassae* fosmid or *D. biarmipes* contig. Feel free to change the track settings (always hitting refresh) to see what happens. Consider the following questions.

a. How do the number and position of *D. melanogaster* proteins support your BLASTX results?

b. How well do the various Gene Prediction tracks agree? How does the display compare to your BLAST analysis of GENSCAN predictions?

c. What parts of your fosmid or contig have the greatest sequence conservation?

d. How are the repeated sequences distributed?

Be sure to take screenshots to support you work on the genes present on your fosmid or contig.