JBrowse Basics

Note: *this exercise uses VectorBase (https://VectorBase.org) as an example database, but the same functionality is available on all VEuPathDB resources.*

Learning objectives:

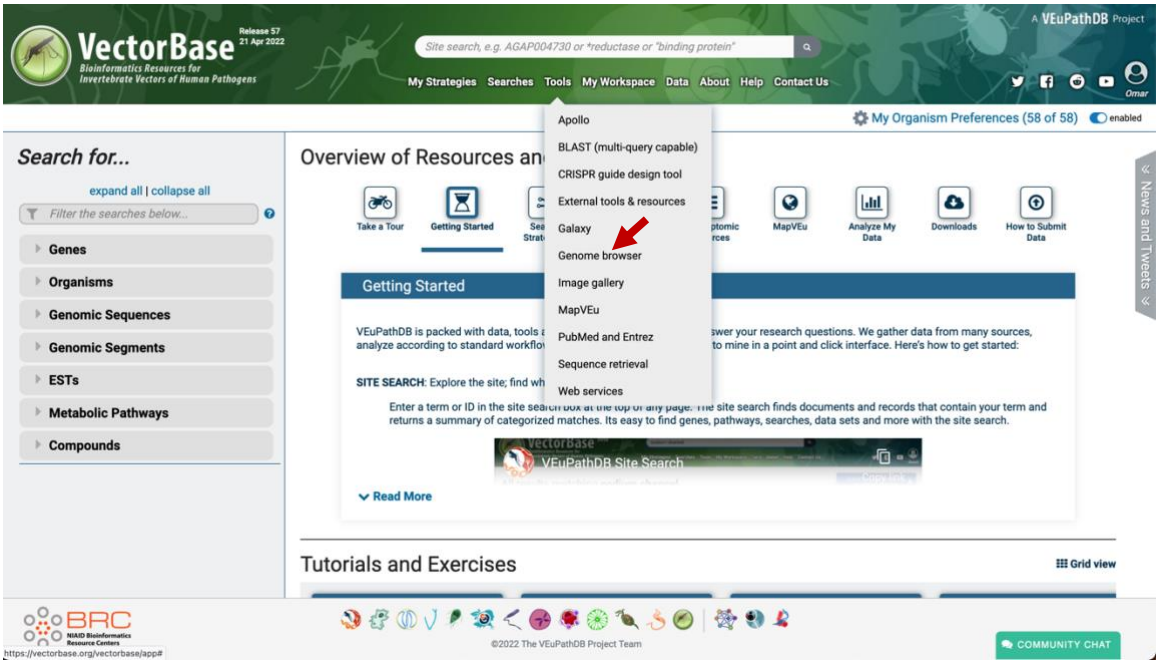
- Navigate to the genome browser
- Use the menu and navigation bars
- Run searches
- Add pre-loaded data tracks
- Upload your own data tracks
- Configure tracks
- Download track data

1. Navigating to the Genome Browser (JBrowse)

JBrowse is a fast and full-featured genome browser built with JavaScript and HTML5. You can read more about JBrowse and its features here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4830012/

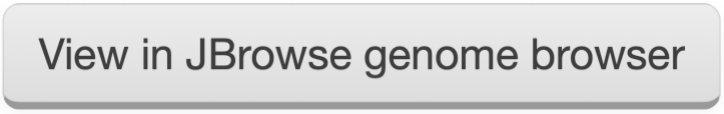
Links to the genome browser are available from multiple locations:

a. The tools menu in the banner of any page.



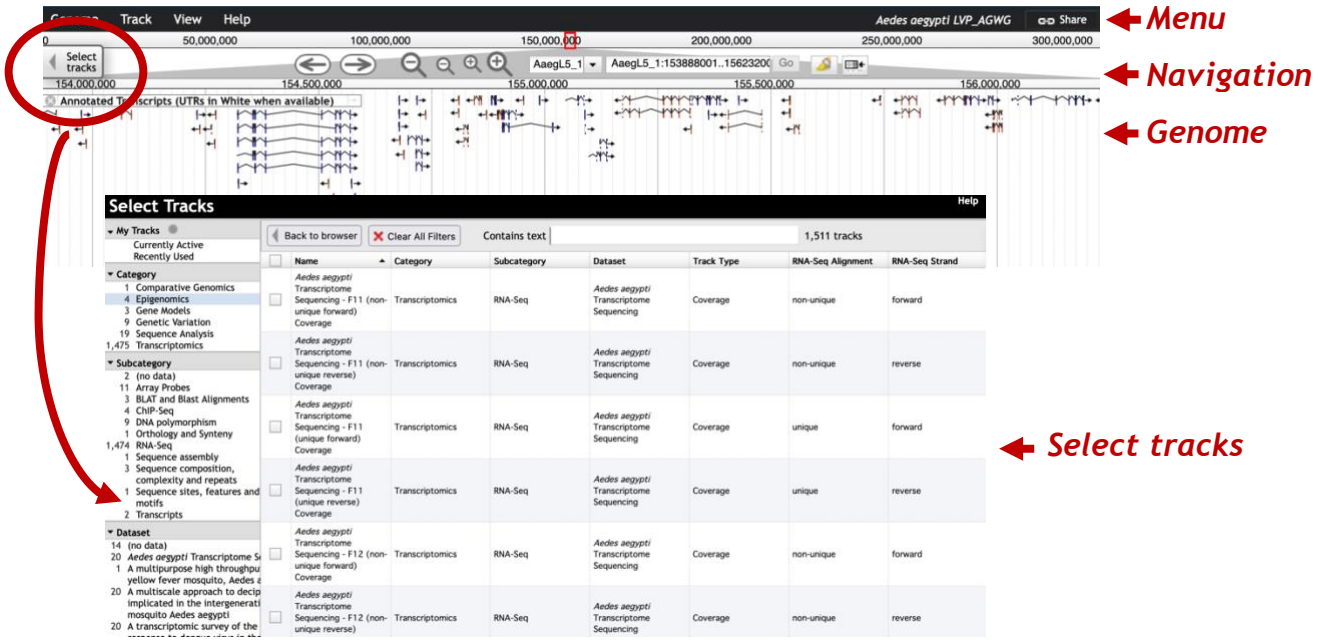
b. From record pages such as gene, SNP or genomic sequence pages – these links are usually to a specific JBrowse configuration that includes data relevant to the section on that record page. For example, a JBrowse link from an RNAseq dataset

on the gene page would display the gene of interest along with the RNAseq data as density or coverage plots. These links are usually indicated by "View in JBrowse



genome browser" button.

2. **Getting around JBrowse.**
   a. Use any of the above described JBrowse linking strategies to get to the genome browser.
   b. Once in JBrowse examine the following features:
      i. The **menu bar**: located at the top of the JBrowse frame. This includes the Genome menu, Track menu, View menu, Help menu and the Sharing link. What do each of these do/provide?
      ii. The **navigation bar**: located below the menu bar. This contains zooming (magnifying glass icons), panning (left/right arrows) and highlighting (yellow highlighter) buttons, reference sequence selector (drop down with sequences from the selected genome sorted by length), a text box to search for features such as gene IDs and overview bar which shows the location of the region in view.

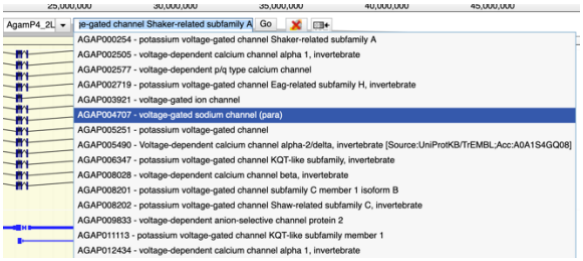


      iii. The **genome view**: this is where the data tracks are displayed.
   c. Selecting tracks: click on the "select track" button (top left). You can search/filter for tracks and then select them for display by checking the check box next to the track name.

3. Navigating to a specific gene in JBrowse.

The goal of this step is to navigate to the voltage-gated sodium channel gene of *Anopheles gambiae* PEST
Make sure the *Anopheles gambiae* PEST genome is selected from the genome menu.

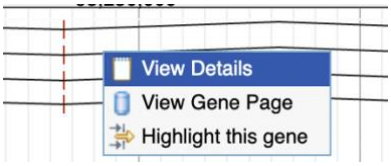
a. Start typing the word voltage in the search box. After a few seconds you should see a list of results (do not hit enter). Select the gene called AGAP004707-voltage-gated sodium channel (para) from the search



dropdown.

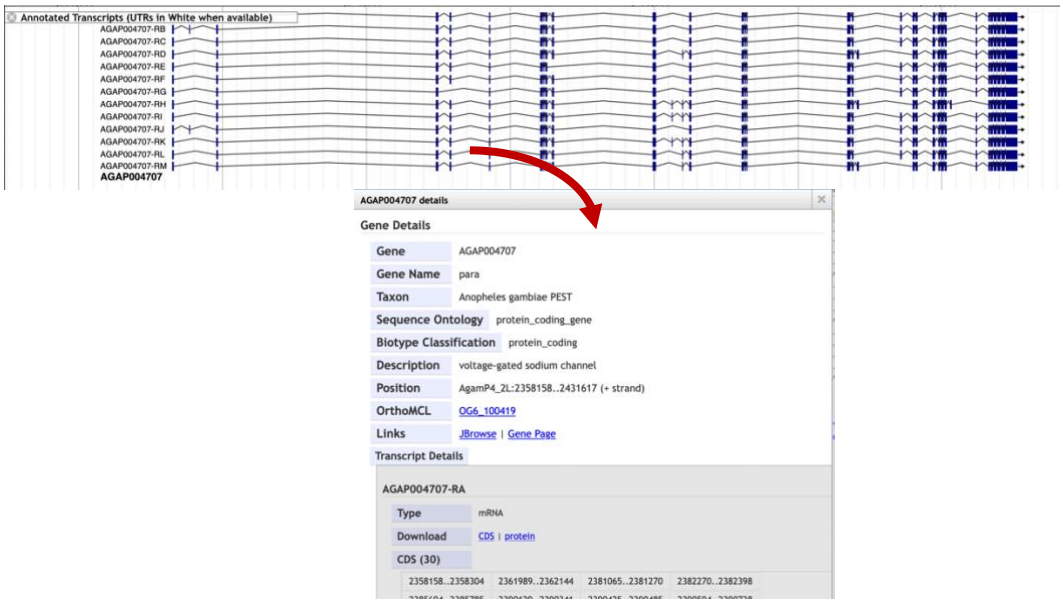
b. You can get information about any feature in the genome view window by clicking on it. Click on the gene feature. What information is available in the popup?

c. You can also right click (or control click) on a feature to display the context menu which provides quick links to highlight a feature, go to the feature page (like the gene page) or get the



info popup (the same one you get when you click on the feature).

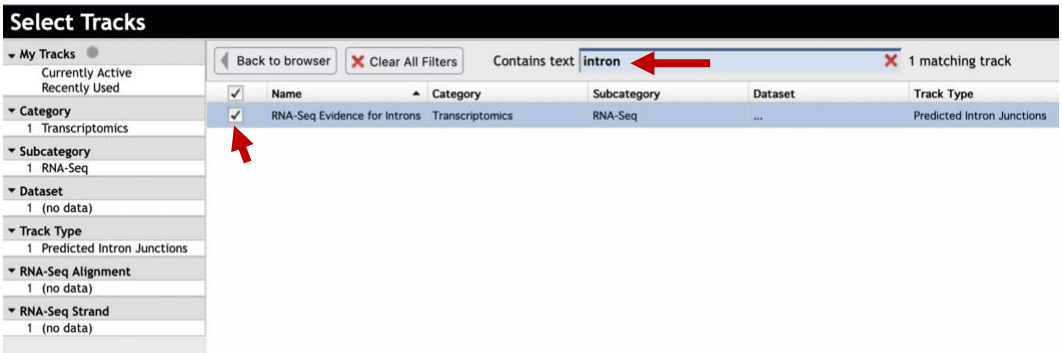
d. How many isoforms does this gene have? What genes are immediately upstream and downstream of this gene? (Hint: if needed use the zoom out button in the navigation bar). What is the difference between the small and large zoom buttons? (*Tip1:* another way to zoom in and out is by clicking on shift and the up or down



arrows. What happens if you click shift and left or right arrows? *Tip2:* you can also zoom in by clicking and dragging your cursor in the location ruler in the navigation bar).

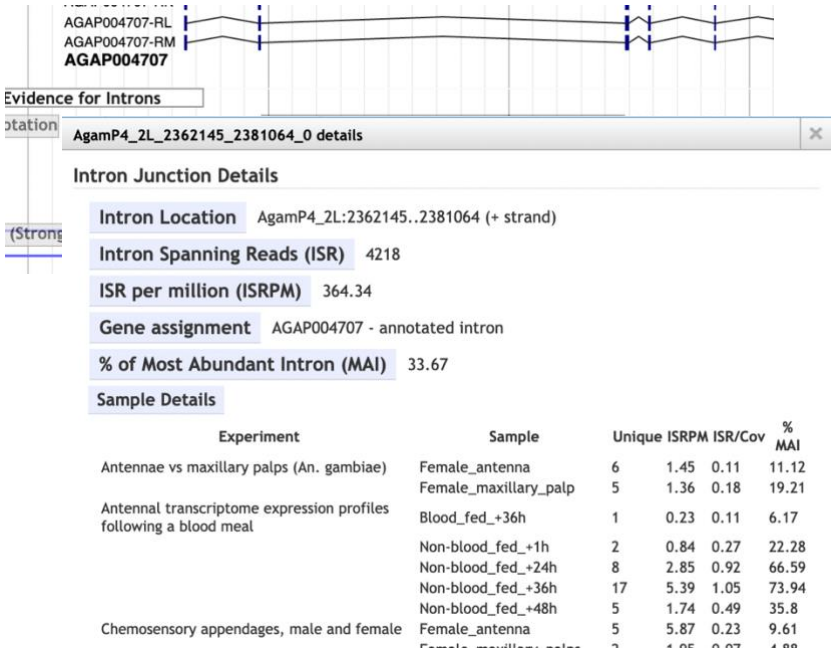
4. Exploring intron evidence

VectorBase contains many RNAseq datasets which can help you explore the gene model. VectorBase also provides a metric for all possible predicted introns based on the RNAseq data. Turn on the track called "RNA-Seq Evidence for Introns". To do this click on the "Select Tracks" tab then type the word intron in the search box – this will filter the tracks to the ones that contain the word "intron". Select this track and



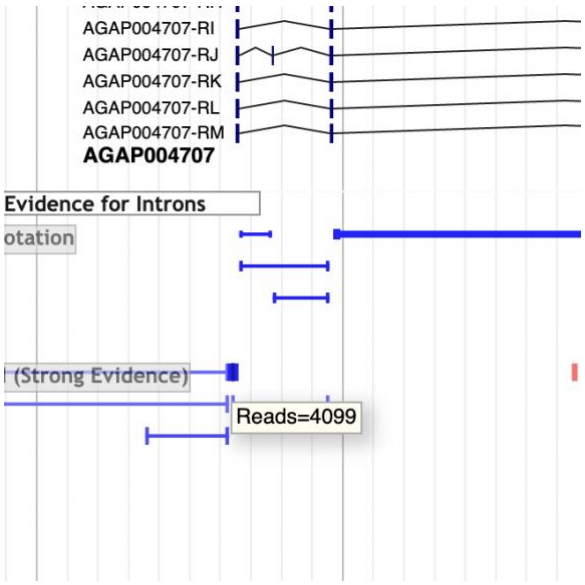
close the "Select Tracks" window.

- Explore the intron evidence tracks. Notice there are two subtracks turned on by default: Matches Annotation and Unannotated (Strong evidence). What is the difference between these subtracks?
- How do you determine the number of reads that support an intron? Mouse over the intron span to see the number.



Experiment	Sample	Unique	ISRPM	ISR/Cov	% MAI
Antennae vs maxillary palps (An. gambiae)	Female_antenna	6	1.45	0.11	11.12
	Female_maxillary_palp	5	1.36	0.18	19.21
Antennal transcriptome expression profiles following a blood meal	Blood_fed_+36h	1	0.23	0.11	6.17
	Non-blood_fed_+1h	2	0.84	0.27	22.28
	Non-blood_fed_+24h	8	2.85	0.92	66.59
	Non-blood_fed_+36h	17	5.39	1.05	73.94
	Non-blood_fed_+48h	5	1.74	0.49	35.8
Chemosensory appendages, male and female	Female_antenna	5	5.87	0.23	9.61

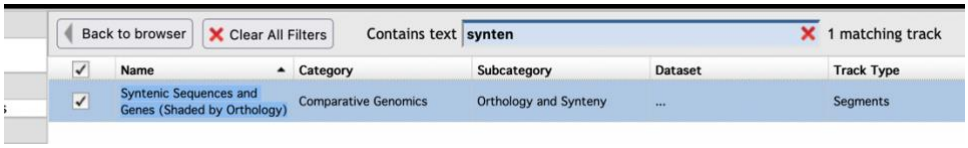
- What happens if you click on one of the intron spanning reads? Notice the popup which contains a lot of information about this intron and the experiments that support it.
- Are there any introns in the "unannotated (strong evidence)" subtrack that have good support? Hint: the darker the color, the stronger the support.



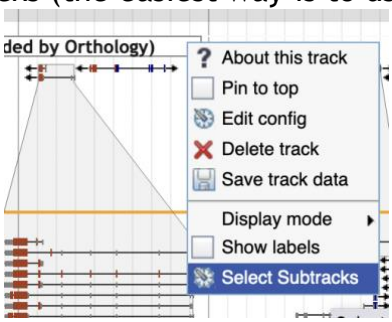
5. **Examining synteny tracks**

JBrowse in VEuPathDB includes a configurable track containing sequence alignment between genomes where genes are shaded between genomes based on orthology. This provides a nice way to examine co-linearity between genomes. In this example we will set up the synteny track to display synteny between *Anopheles gambiae* PEST and *Drosophila melanogaster* iso-1.

- Turn on the track called "Syntenic Sequences and Genes (Shaded by Orthology)"

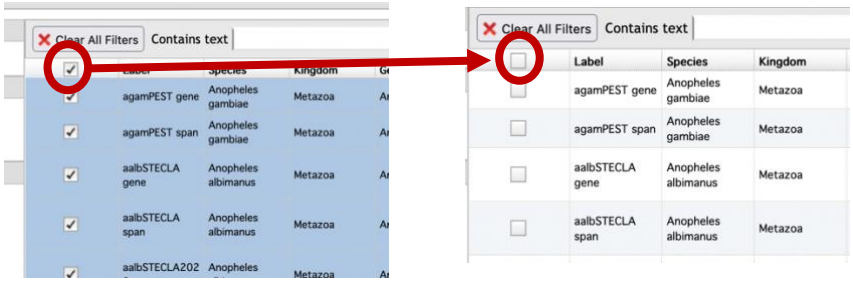


- Select synteny subtracts by clicking on the down arrow on the track name and selecting "Select subtracts".
- Unselect all the tracks (the easiest way is to use the top check box to

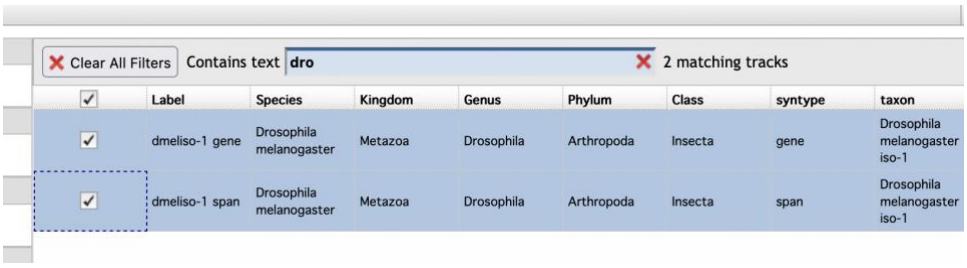


select all then unselect all)

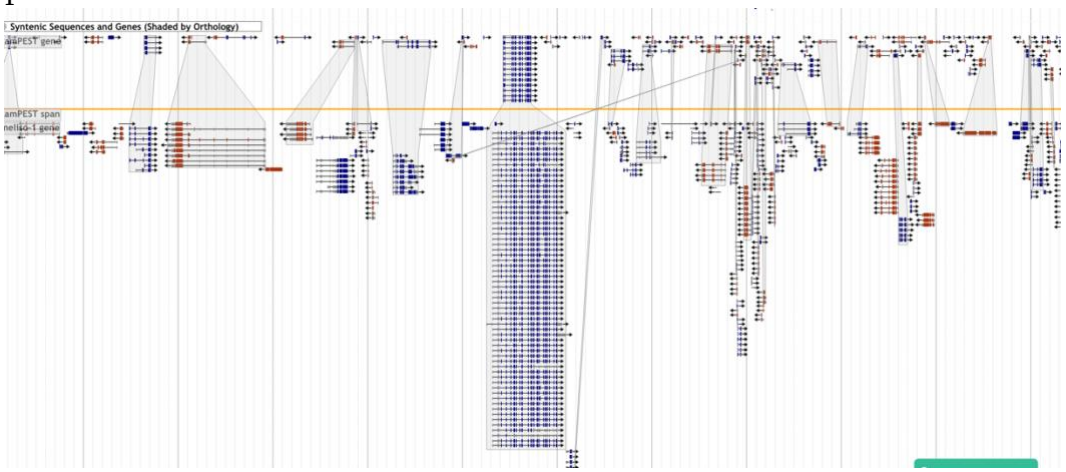
- Select the tracks for *Anopheles gambiae* PEST and *Drosophila*



melanogaster iso-1 (note you can use the search box at the top to find your organism of interest).



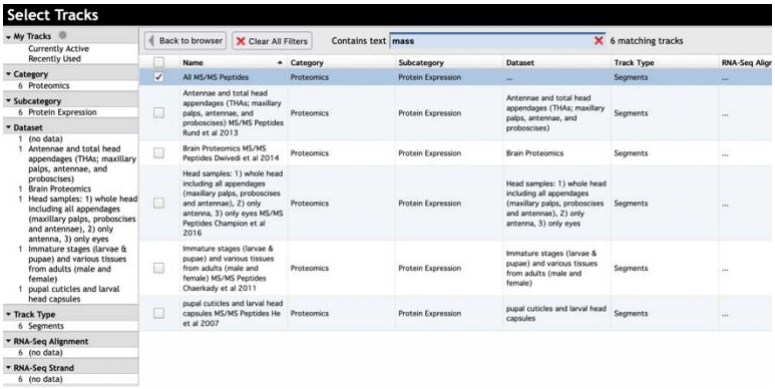
- Does *Drosophila* have an ortholog of the *Anopheles* sodium channel? Does it have a similar number of isoforms or does it have many more?
- What does synteny looked like around this gene? Zoom out a bit (this might be slow to load). Is co-linearity relatively preserved between these two species?



6. Adding more data tracks.

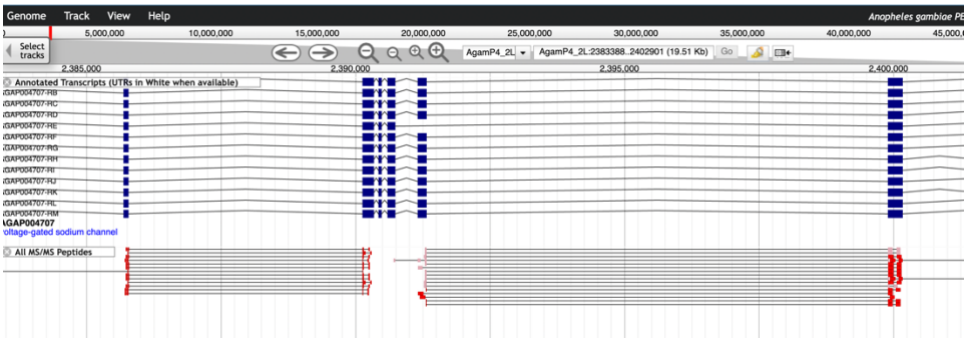
JBrowse contains many data tracks that you can load by selecting them from the select tracks tab. Try this:

- Load the track called All MS/MS Peptides



- This track displays all peptides from mass spectroscopy experiments we have in the database mapped to the genome.

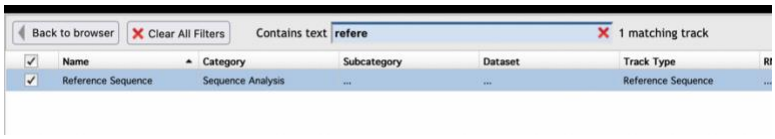
- Does the voltage-gated channel gene have mapped peptides? You may want to zoom in to a region with peptide evidence to see more details.



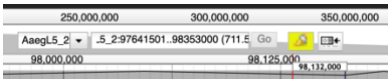
7. (Optional) Retrieving sequence data from JBrowse

a. Downloading sequence in FASTA format from a region of interest:

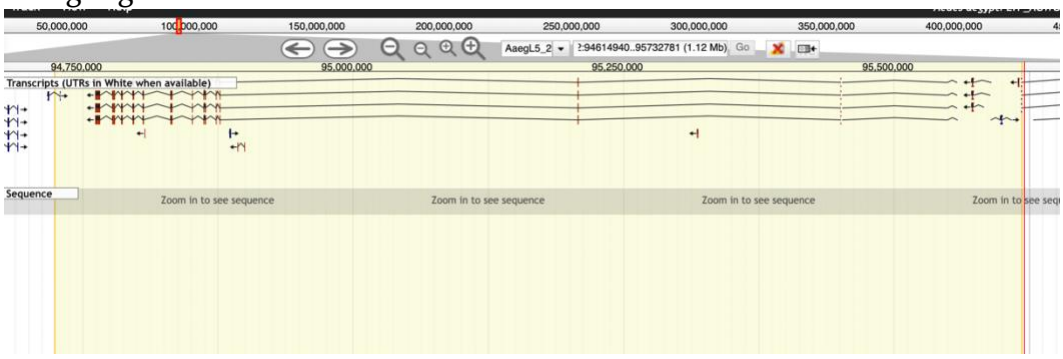
i. Make sure the "annotated transcripts" and the "reference sequence" tracks are turned on.



ii. Click on the "highlight a region" button in the navigation bar. It should turn yellow when activated.



iii. Click and drag in the genome view region and select the area you would like to highlight.



iv. Click on the down arrow on the reference sequence track and select "Save track data".

v. In the next popup window you can keep everything as the default and either save or view the sequence.