Genome Browser 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

Table of Contents

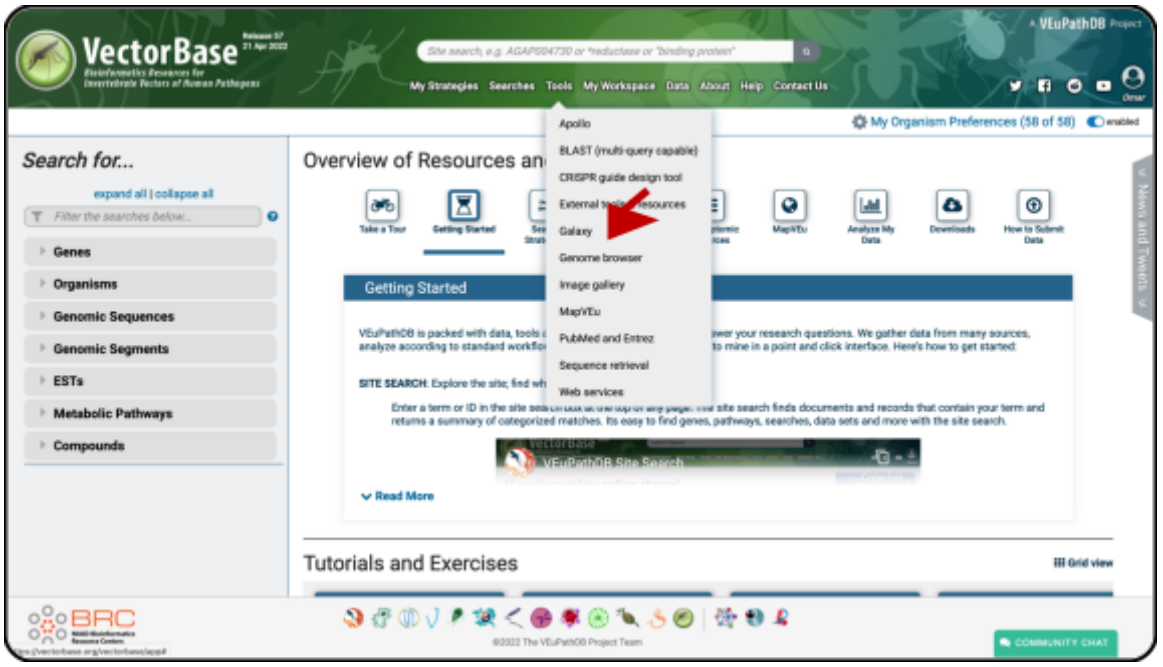
Introduction

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/.

Exercises

1. Navigating the genome browser

1. Links to the genome browser are available from multiple locations.
    a. The tools menu in the banner of any page

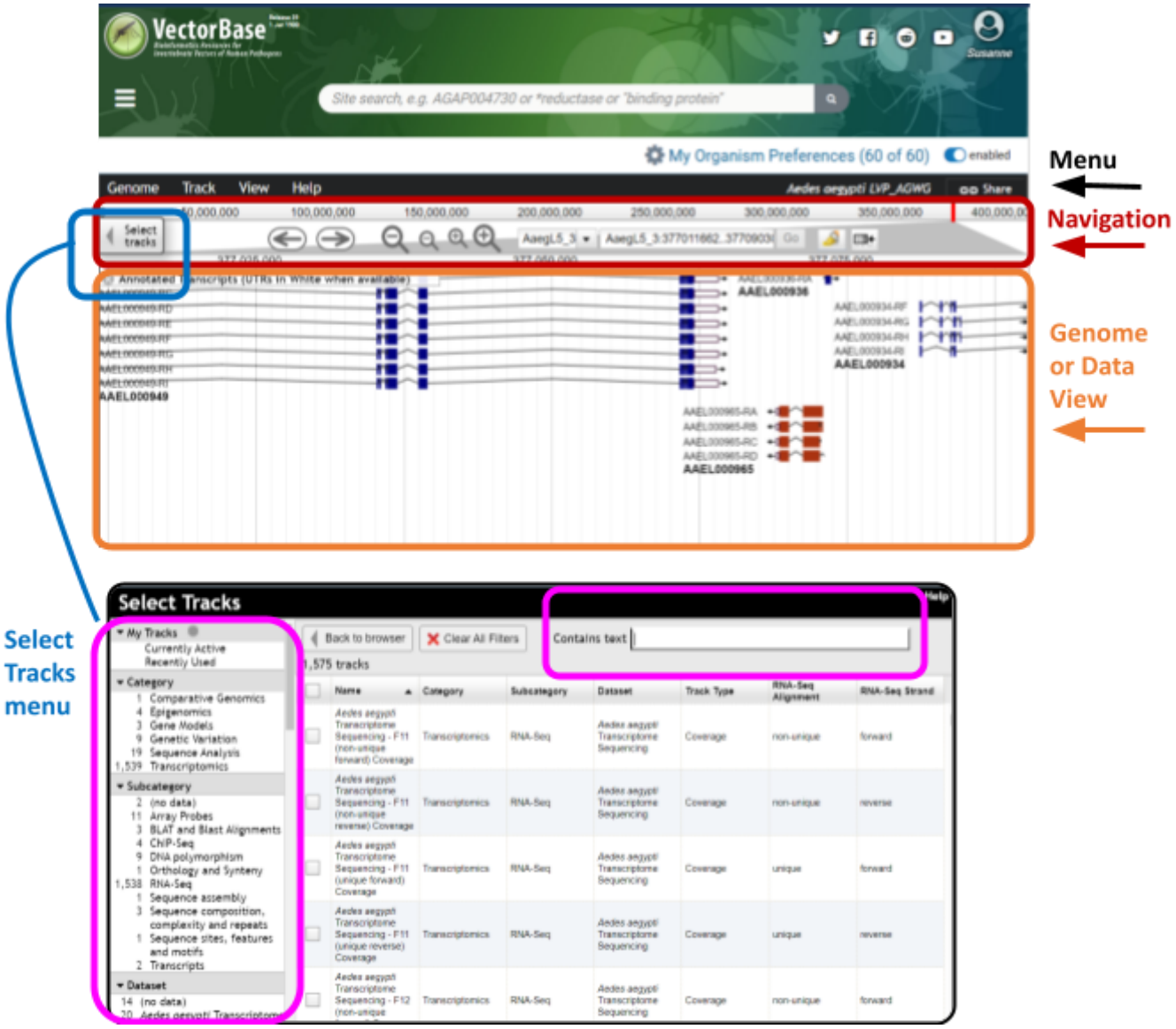


    b. Record pages such as Gene (https://vectorbase.org/vectorbase/app/record/gene/AGAP004707), or genomic sequence pages (https://vectorbase.org/vectorbase/app/record/genomic-sequence/AaegL5_1) – these links are usually to a specific JBrowse configuration that includes data relevant to the section on that 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 the “View in JBrowse genome browser” button.

View in JBrowse genome browser

c. Select any "View in JBrowse genome browser" button to proceed with the following steps, *e.g.*: https://vectorbase.org/vectorbase/app/record/gene/AAEL006498

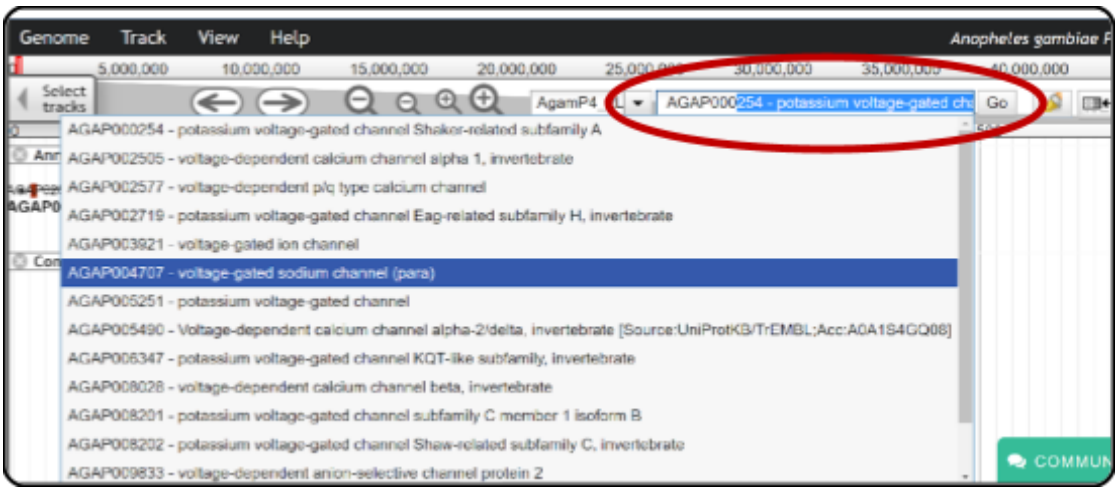
2. Use the tools menu or the View in JBrowse button to open JBrowse. Once in JBrowse examine the following features:
    a. The **menu bar**: located at the top of the JBrowse frame. This includes the Genome menu, Track menu, View menu, Help menu, and the Sharing link. What do each of these do/provide?
    b. The **navigation bar**: located below the menu bar. This includes numbers indicating the base pair position in the genomic sequence (location ruler), 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. Zoom features are also built into the scale on the top of the navigation panel. Select an area to zoom to that location.
    c. The **genome view**: this is where the data tracks are displayed. When viewing the annotation track (topmost track), you can move upstream and downstream by dragging the track features left or right.
    d. **Select tracks**: Click on the “Select track” button (top left). This menu contains all the data tracks that are aligned to the genome that you are viewing. The list of tracks can be filtered using the ‘clickable’ left panel categories or with the search/filter function above the detailed right panel.



2. Navigate to a specific gene in the genome browser

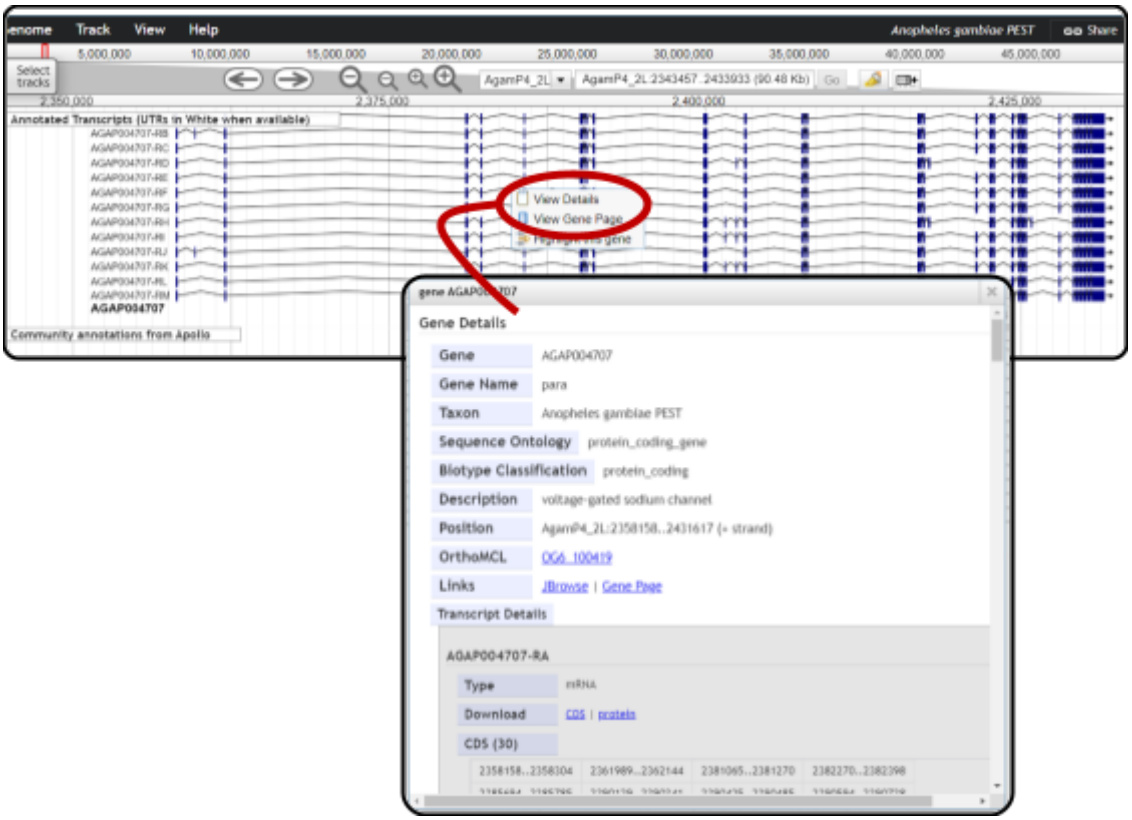
The goal of this step is to navigate to the voltage-gated sodium channel gene of *Anopheles gambiae* PEST.

1. Make sure the *Anopheles gambiae* PEST genome is selected from the genome menu.
2. 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. Then click GO.



3. Details about most features are available in pop-up panels. Click (or control-click) on the gene feature to view the details panel. You can also right-click to choose the same details panel, a link to the gene page, or highlight the gene in yellow.

    a. What information is available in the popup?



    b. Observe that the gene AGAP004707 exhibits alternative splice variants/transcripts/isoforms, such as AGAP004707-RA, *-RB, *-RC, and so on. If there is any uncertainty regarding this concept, refer back to the last page of this

tutorial for the Glossary. 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.

c. What is the difference between the small and large zoom buttons?

Tip 1: another way to zoom in and out is by clicking on shift and the up or down arrows.

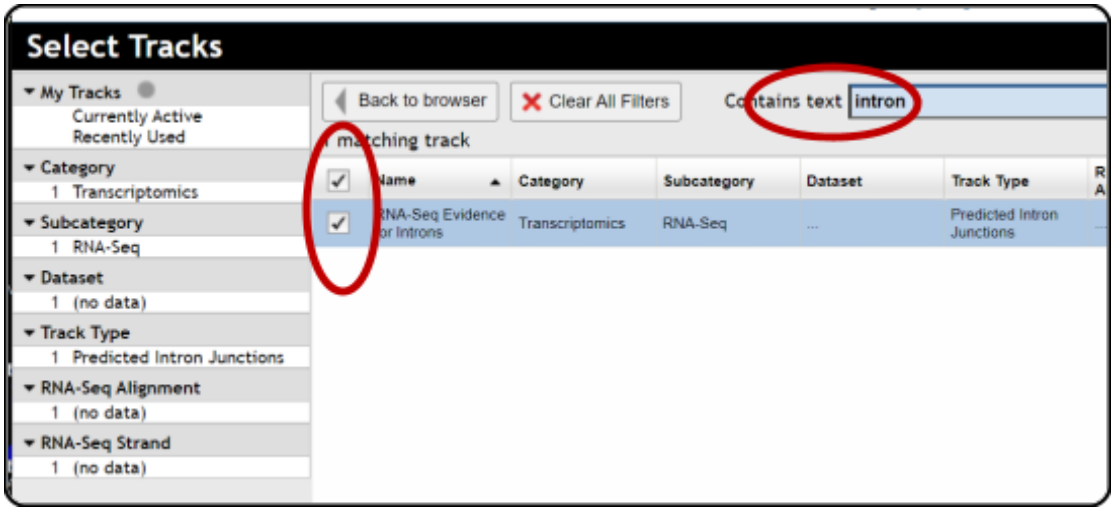
d. What happens if you click shift and left or right arrows?

Tip 2: you can also zoom in by clicking and dragging your cursor in the location ruler in the navigation bar.

3. Exploring intron evidence

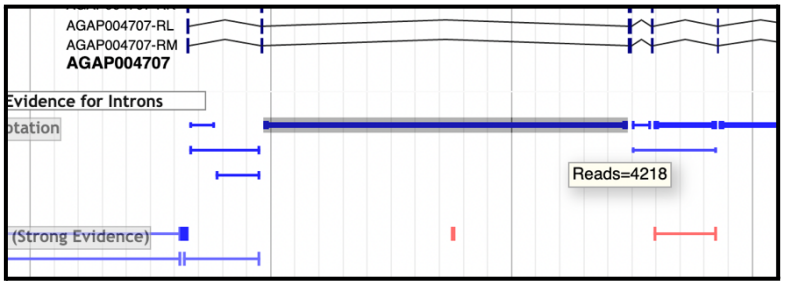
VectorBase contains many RNAseq datasets that can help you explore the gene model. VectorBase also provides a metric for all possible predicted introns based on the RNAseq data.

1. 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.

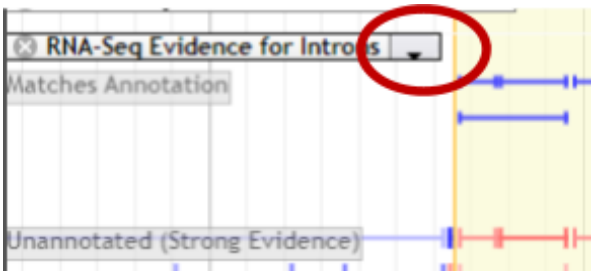


a. Explore the intron evidence tracks. Notice there are two subtracks turned on by default: Matches Annotation and Unannotated (Strong evidence). What is the difference between these subtracks?

2. How do you determine the number of reads that support an intron? Mouse over the intron span to see the number.

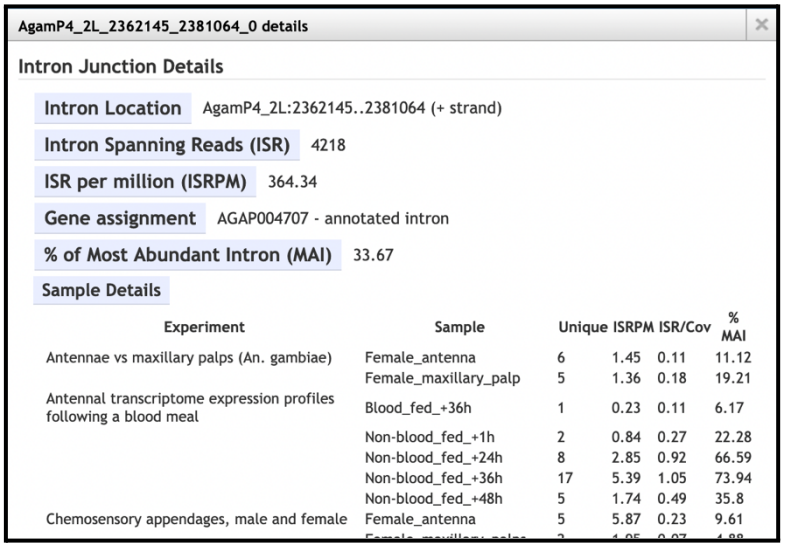


3. The track titles also contain dropdown menus with actions or information about the track. Hover over the track title and then click the down arrow that appears to access more track functions.



4. 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.
5. Are there any introns in the "unannotated (strong evidence)" subtrack that have good support?

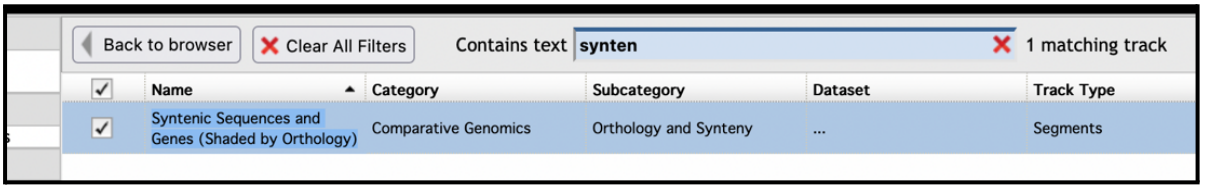
   Hint: the darker the colour, the stronger the support.



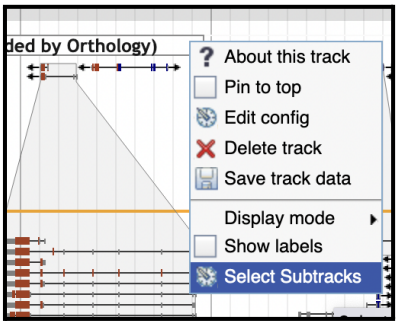
4. Examining synteny tracks

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.

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

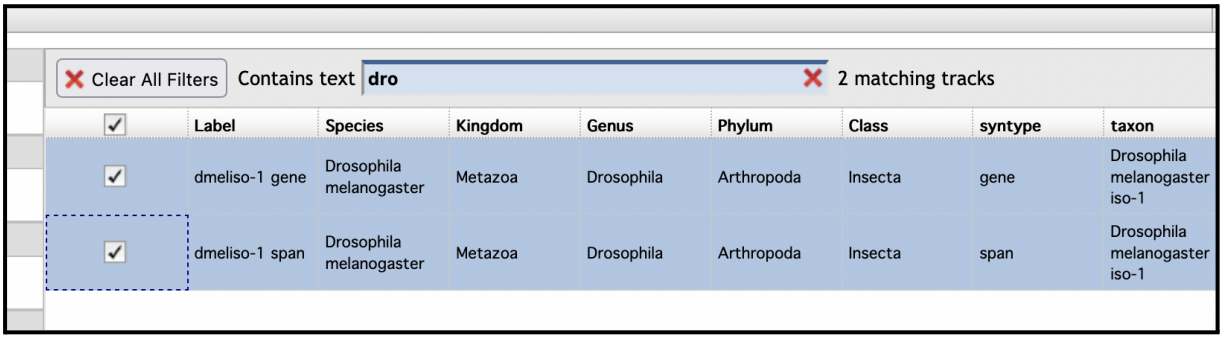


2. Select synteny subtracts by clicking on the down arrow on the track name and selecting "Select subtracts".



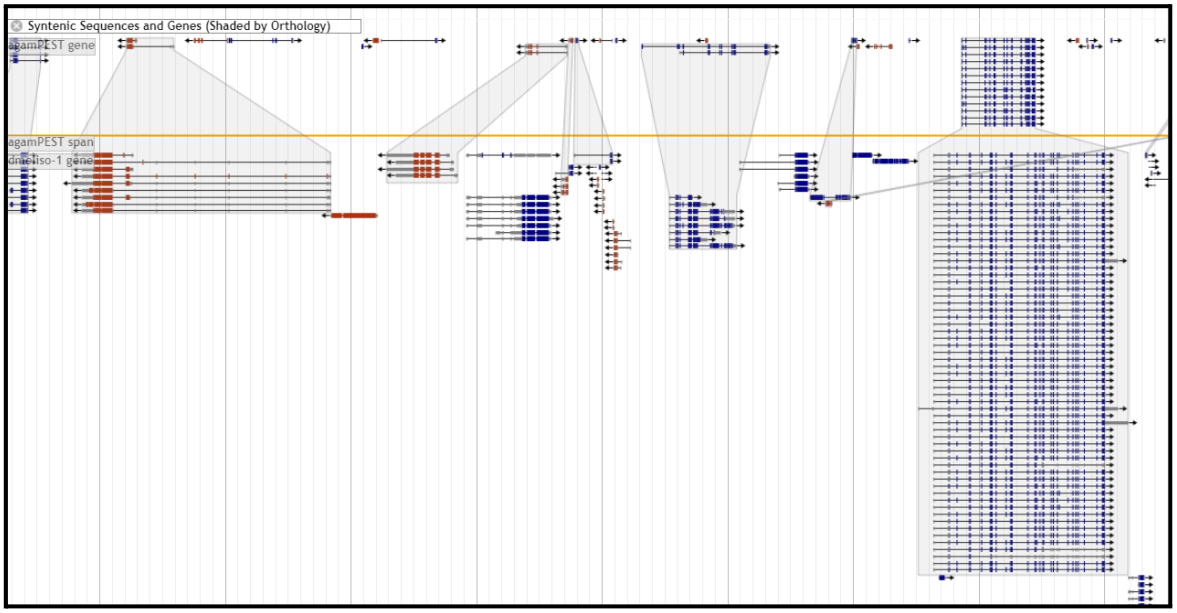
3. Unselect all the tracks (the easiest way is to use the top check box to select all then unselect all)

4. Select the tracks for *Anopheles gambiae* PEST and *Drosophila melanogaster* iso-1 and then click “SAVE” at the bottom of the “Select Subtracks” popup. (note you can use the search box at the top to find your organism of interest).



Label	Species	Kingdom	Genus	Phylum	Class	syntype	taxon
dmeliso-1 gene	Drosophila melanogaster	Metazoa	Drosophila	Arthropoda	Insecta	gene	Drosophila melanogaster iso-1
dmeliso-1 span	Drosophila melanogaster	Metazoa	Drosophila	Arthropoda	Insecta	span	Drosophila melanogaster iso-1

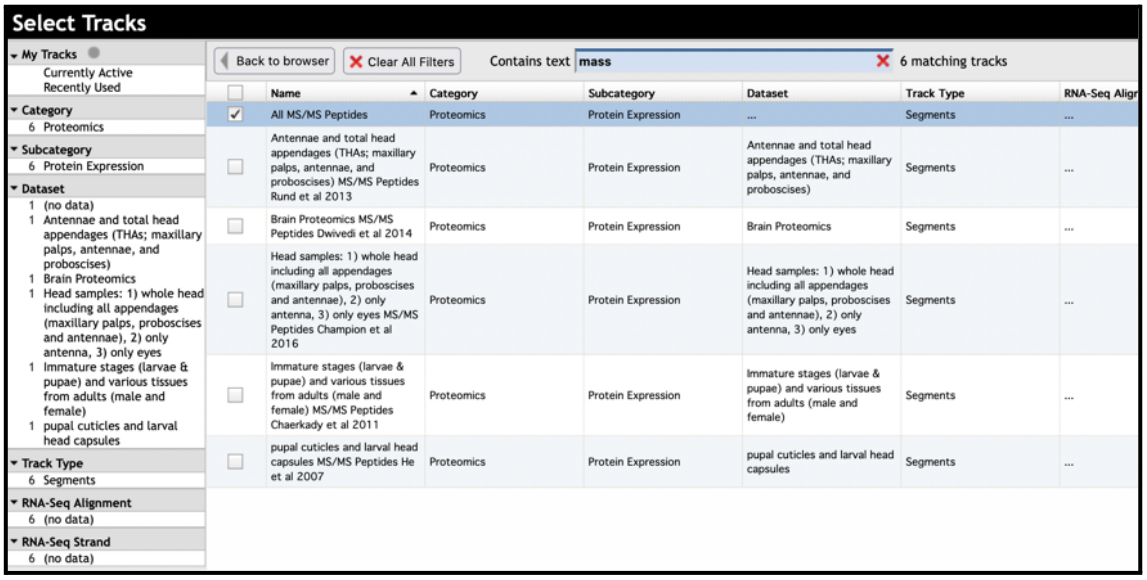
5. View the track.
    a. Does *Drosophila* have an ortholog of the *Anopheles* sodium channel?
    b. Does it have a similar number of isoforms, how many?
    c. What does synteny look like around this gene? Zoom out a bit (this might be slow to load).
    d. Is co-linearity relatively preserved between these two species?



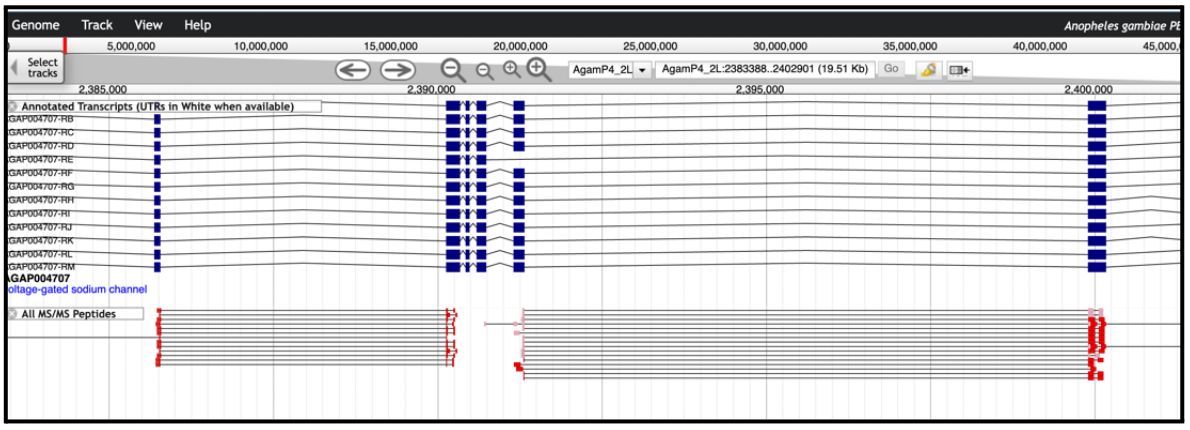
5. Adding data tracks

JBrowse contains many data tracks that you can load by selecting them from the select tracks tab. Each track represents sequence data or features aligned to the genome. Depending on the data type, viewing different data tracks in concert can reveal much.

1. 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.



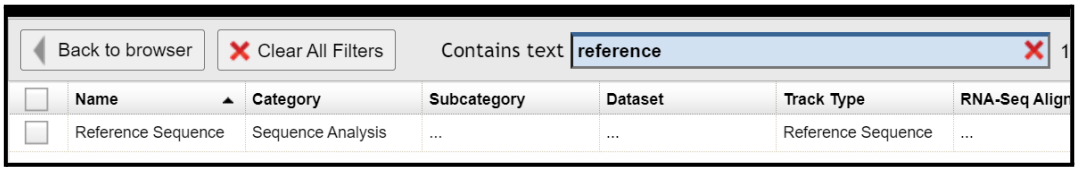
2. Does the voltage-gated channel gene you found earlier have mapped peptides? You may want to zoom in to a region with peptide evidence to see more details.



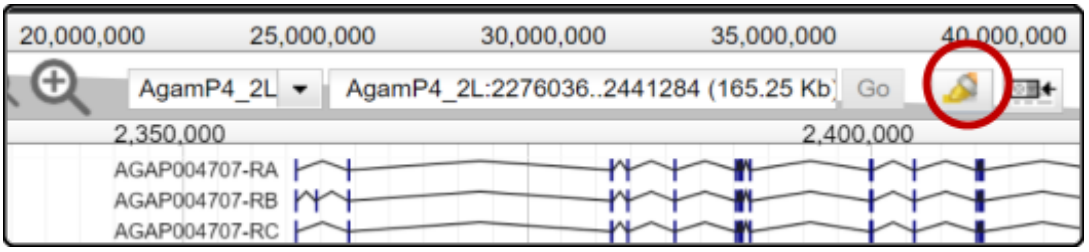
6. Retrieving sequence data from JBrowse (Optional)

Sequence data from a region of interest can be downloaded from JBrowse in FASTA format. Download is a function of the Reference Sequence track, so that must be turned on. And the region to download must be highlighted in yellow.

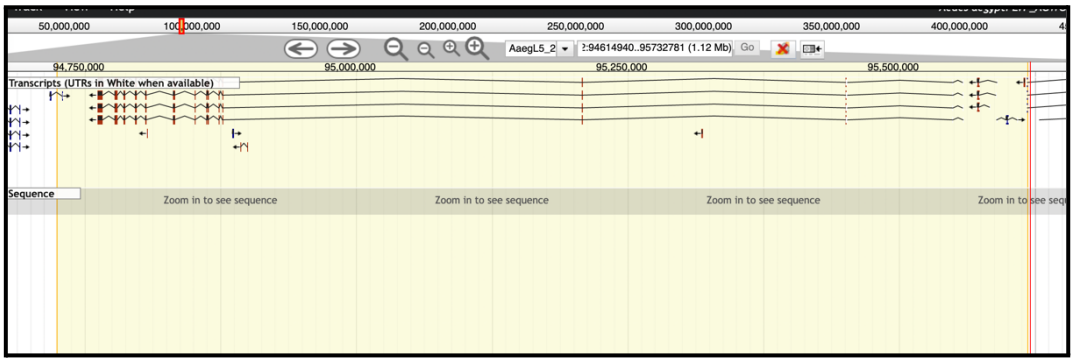
3. Make sure the “annotated transcripts” and the “reference sequence” tracks are turned on.



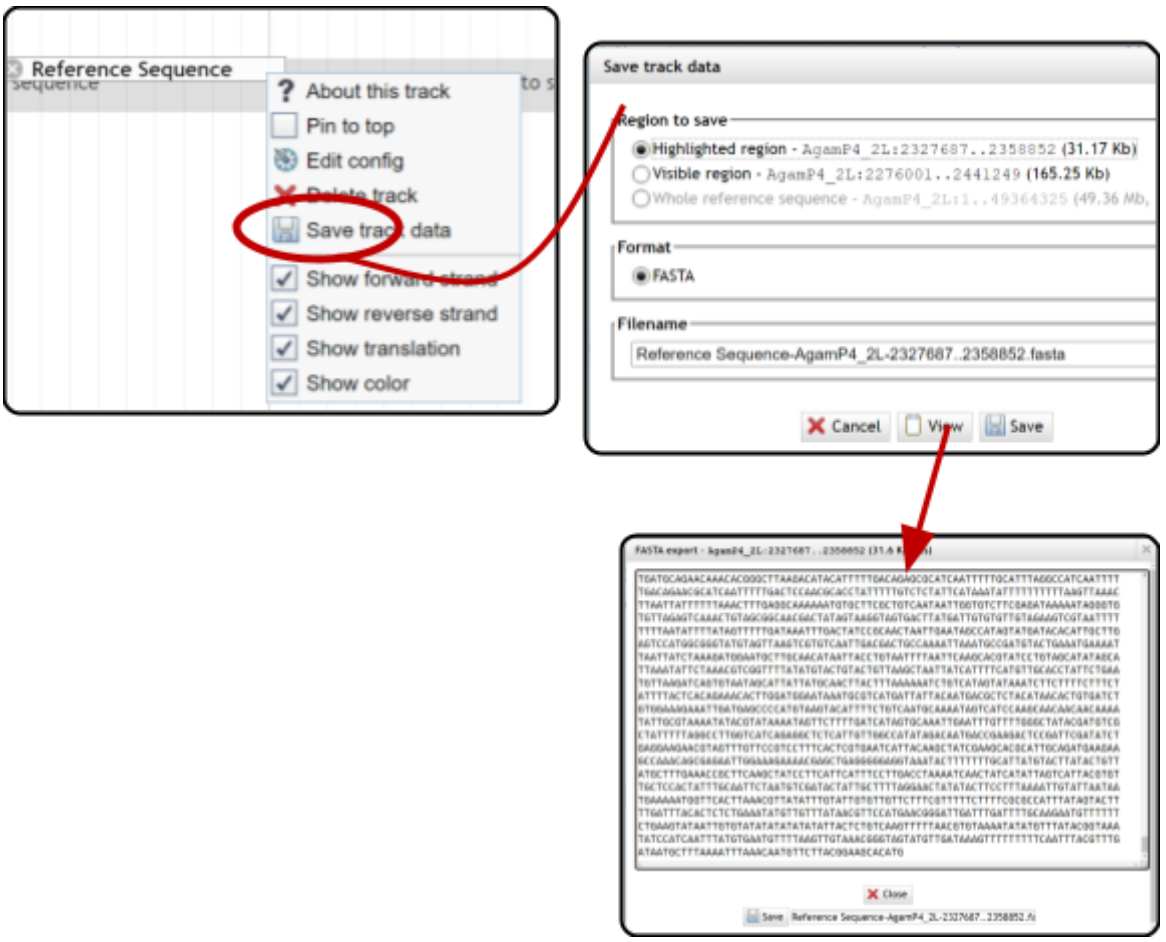
4. Click on the “highlight a region” button in the navigation bar. It should turn yellow when activated.



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



6. Click on the down arrow on the reference sequence track and select “Save track data”.
7. In the next popup window, you can keep everything as the default and either save or view the sequence.


Reference Sequence
About this track
Pin to top
Edit config
Delete track
Save track data
Show forward strand
Show reverse strand
Show translation
Show color
Save track data
Region to save
Highlighted region - AgamP4_2L:2327687..2358852 (31.17 Kb)
Visible region - AgamP4_2L:2276001..2441249 (165.25 Kb)
Whole reference sequence - AgamP4_2L:1..49364325 (49.36 Mb)
Format
FASTA
Filename
Reference Sequence-AgamP4_2L-2327687..2358852.fasta
Cancel
View
Save

Glossary

Co-linearity, the conservation of blocks of order between compared chromosomes.

Intron spanning reads, reads that cover/are divided by non-coding genomic regions.

Isoform, a member of a set of highly similar proteins that originate from a single gene and are the result of alternative splicing.

Mapping, the process of finding the location of genes in a genome.

Ortholog, homologous genes in different species that evolved from a common ancestral gene by a speciation (lineage-splitting) event.

Paralog, homologous genes created by a duplication event within the same genome.

Single Nucleotide Polymorphisms (SNPs, pronounced snips), genomic variation at a single base position.

Syntenic sequence, conserved sequence blocks. Can refer to co-linearity or the physical co-localisation of genetic loci on the same chromosome within an individual or species.