# Visiting a Gene Page & the Genome Browser (JBrowse) Basics

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

Answer Key (release 47 - June 2020)[[1]](#footnote-1)

**Learning objectives**

Gene pages:

* Become familiar with the information in gene pages
* Navigate to and from the gene pages

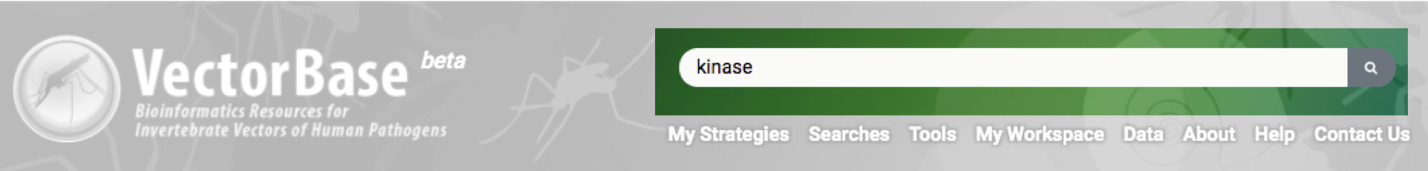
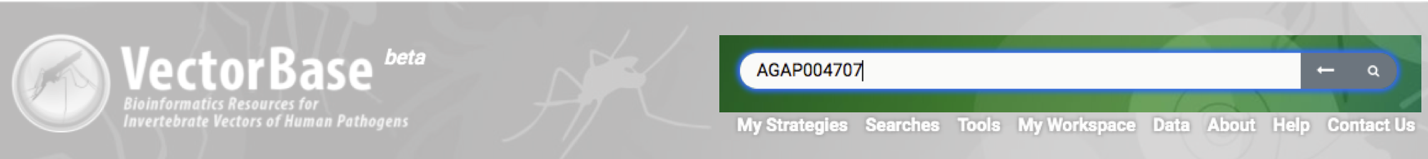
Genome Browser:

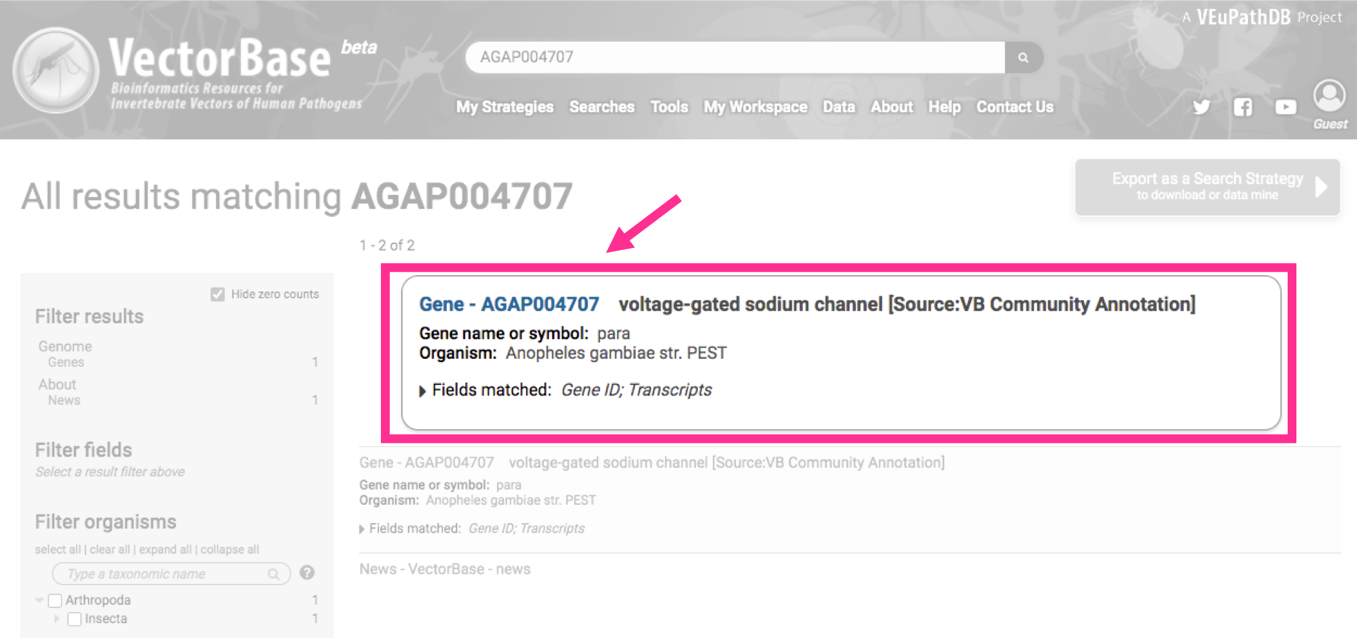
* Navigate to the genome browser
* Become familiar with the different menus and features
* Find specific genes
* Add preloaded data tracks
* Download track data

You can read more about JBrowse and its features here: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4830012/>

1. **Navigation to the Gene pages**

In Site Search use a keyword (gene name/symbol or function/description) or a gene ID; or click on the results from a Search Strategy, one of the columns has the gene IDs.

* 1. With keywords (e.g., kinase) or gene IDs (e.g., AGAP004707) from “Site Search”

If you use a gene ID, in the center of the results page a featured a box will take you to the gene page

1. Examine the information at the top of the gene page AGAP004707:

What is the gene name?

|  |
| --- |
| voltage-gated sodium channel gene |

What chromosome is this gene on?

|  |
| --- |
| 2L |

1. Examine the “1. Gene Models” section of the gene page:

How many exons does this gene have?

|  |
| --- |
| 39 |

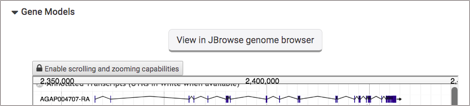
How many transcripts does this gene encode?

|  |
| --- |
| 13 |

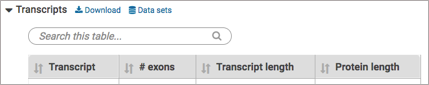
What direction are the transcripts relative to the chromosome?

|  |
| --- |
| Forward |

1. Try interacting with the gene models image view. Now click on “Enable scrolling and zooming capabilities”. What functions are you able to do now?

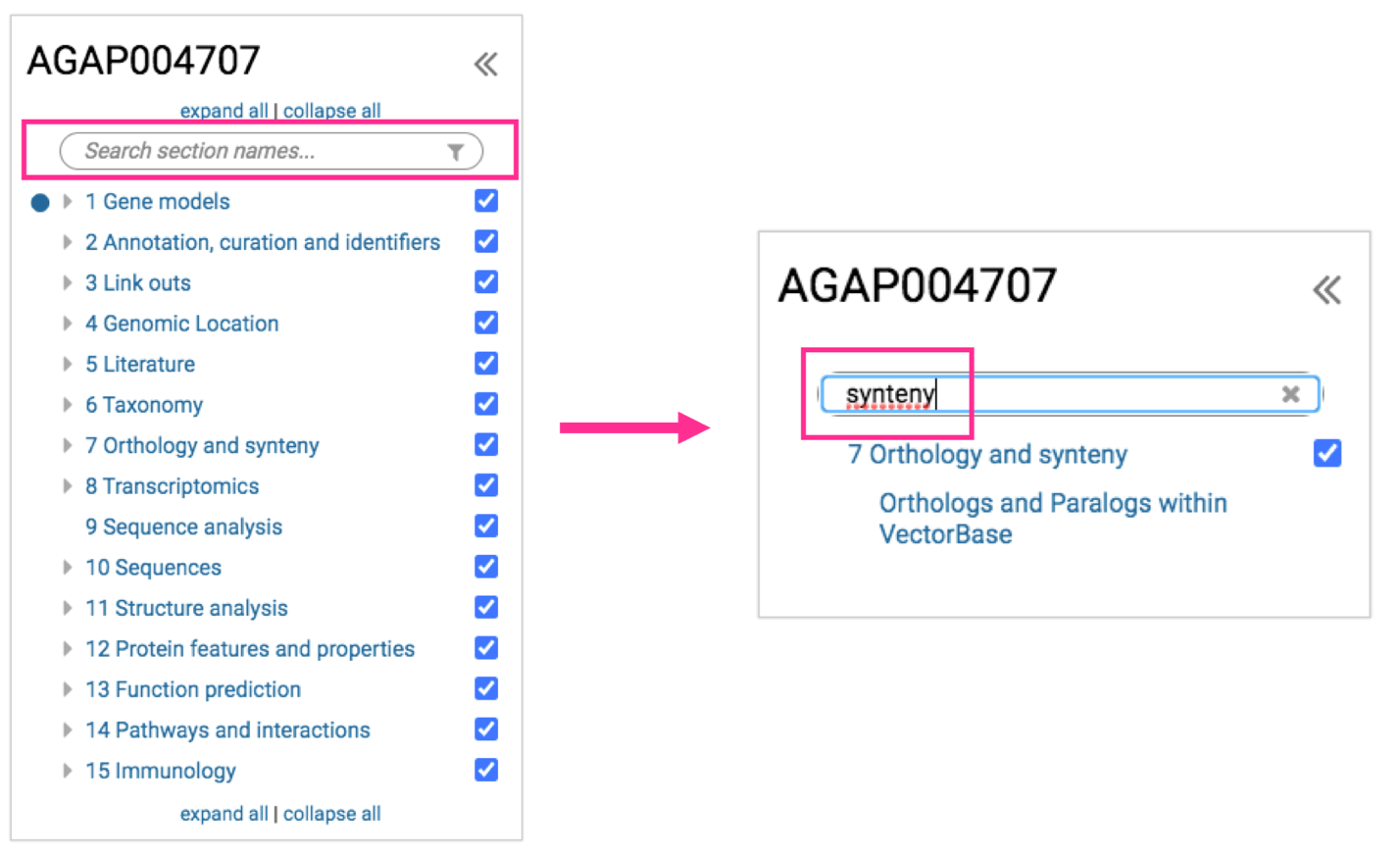


|  |
| --- |
| scroll up and down |

1. How many nucleotides is the largest transcript? Click on the column arrows to sort

|  |
| --- |
| 6,387 |

1. **Content Navigation**

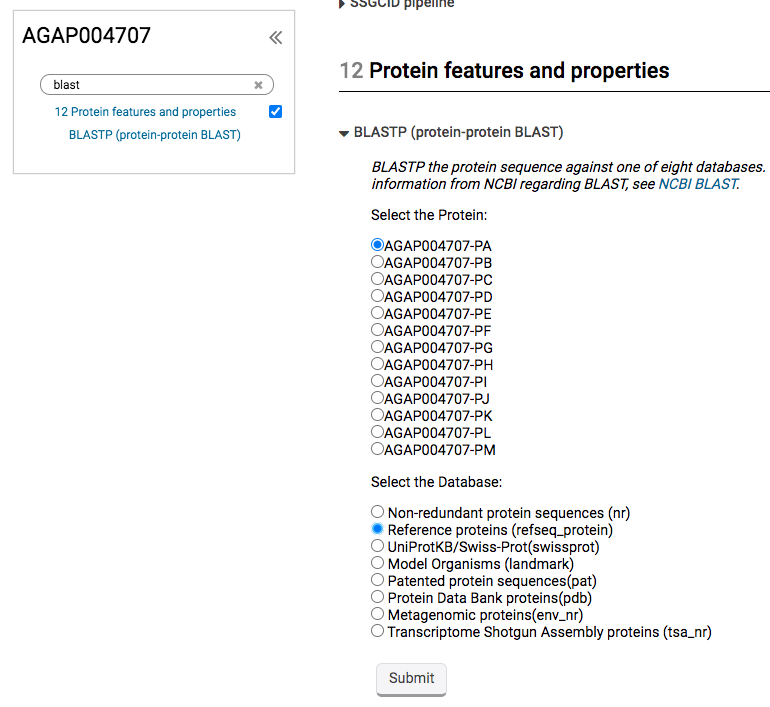
How do you find/navigate to the different sections of the page? Use the “Contents” menu on the left side, type a keyword and click on the menu, click on the work to navigate to it on the page. In the example below the word “synteny” is used. You can also click on the images in the Shortcuts section in the top of the page.

1. **Run jobs such us BLAST and Clustal**

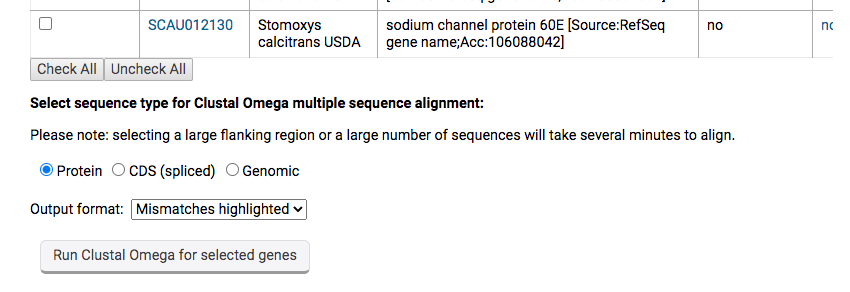
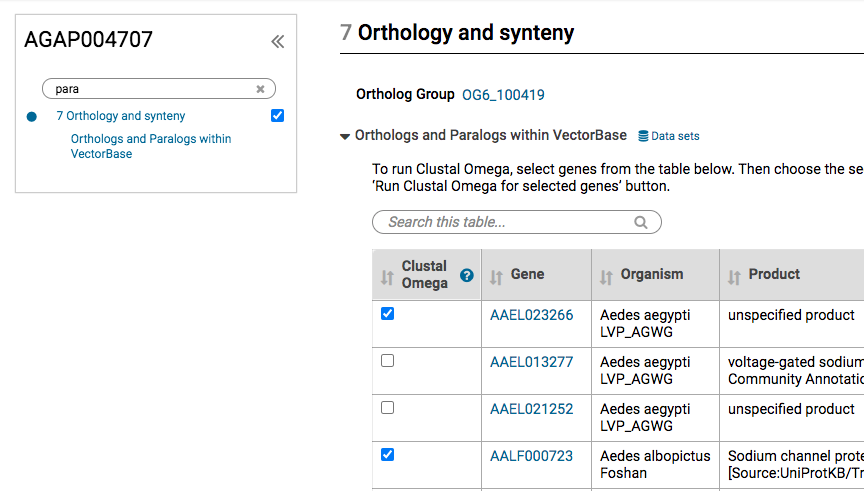
You can run BLAST similarity search or a multiple sequence alignment job with Clustal

1. For BLAST use the “Contents” menu on the left side, same as above

1. For Clustal use the “Contents” menu on the left side, same as above. The tool is part of the “Orthologs & Paralogs” table. The Clustal parameters are at the bottom of the table. When looking at the results mismatches can help you identify candidate genes to be manually annotated

**BLAST**

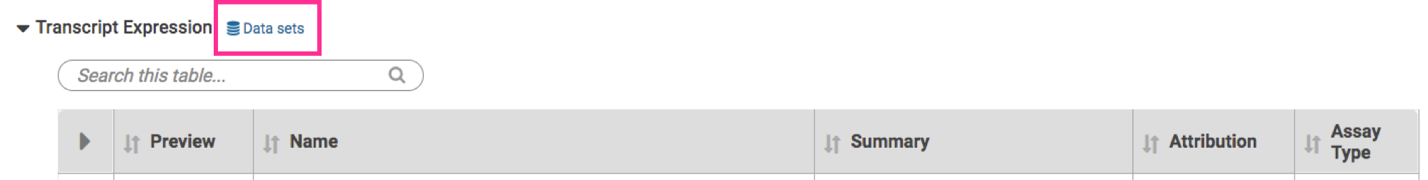
**Clustal**



1. **Is this gene expressed at the transcript level?**[[2]](#footnote-2)

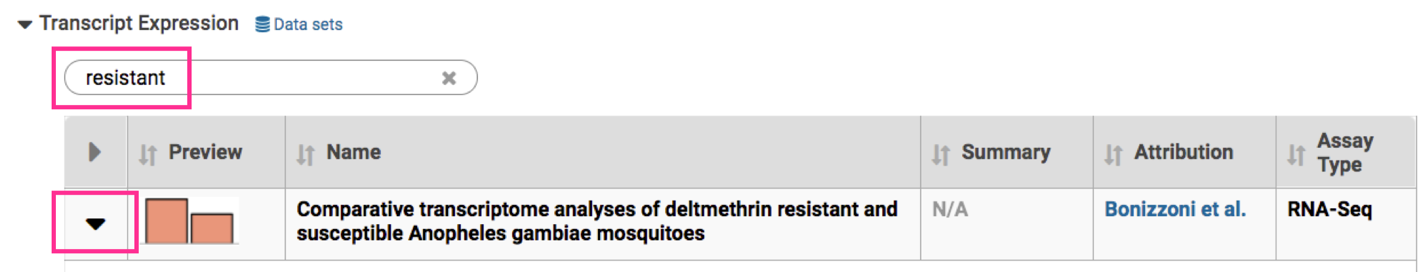
1. Look at the gene page section entitled “Transcriptomics”. Again, use the “Contents” menu on the left side. What kinds of assays provide evidence for transcript expression? View the “Transcription Expression” table

|  |
| --- |
| Array and RNA-Seq |

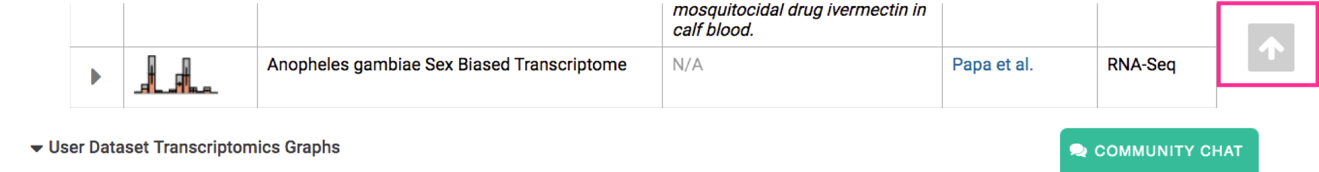
1. Can you link to the data set record for the transcriptomics experiments? In this new page, how many data sets are shown?

|  |
| --- |
| 23 |

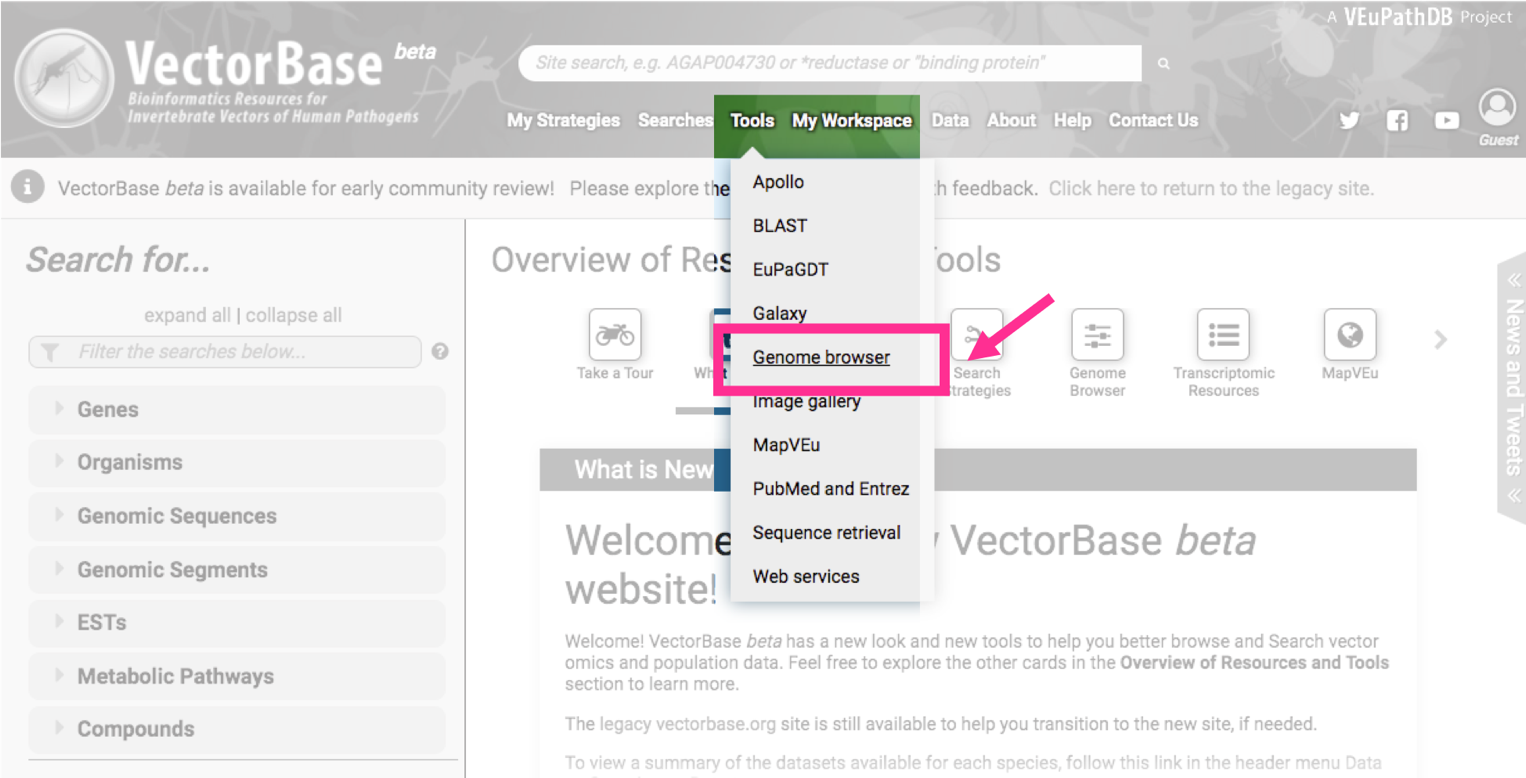
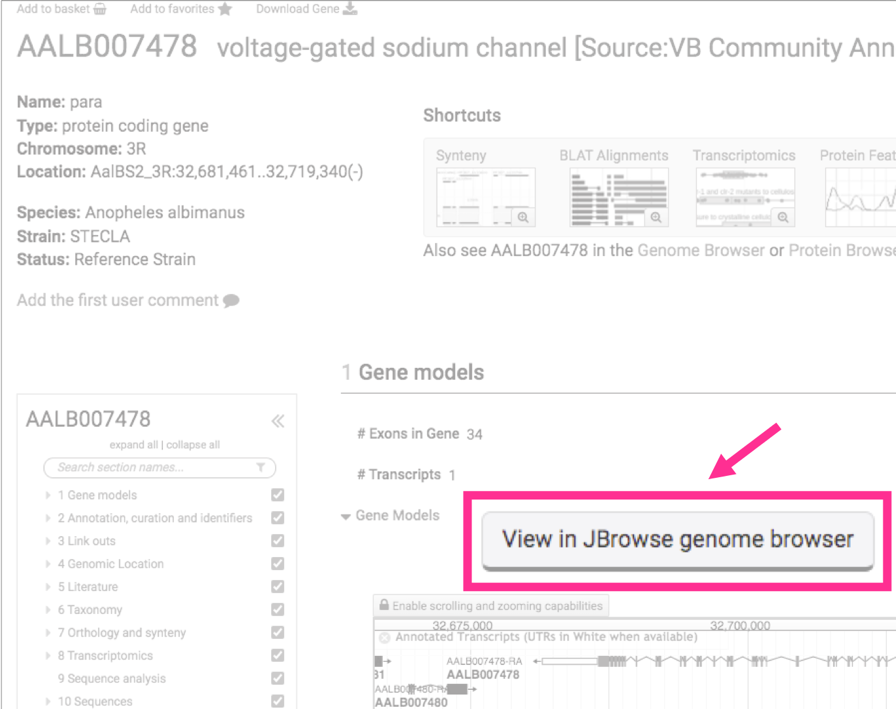
1. Use a keyword to find an experiment related to your question, can be either in the title, description or the author. Is this gene transcript more abundant in insecticide resistant or susceptible mosquitoes? Justify your answer (hint: also look below in the “Data Table”)



|  |
| --- |
| Is more abundant in resistant mosquitoes because it has ~7.5 transcripts per million (TPM) compared to ~5 TPM in susceptible mosquitoes. |

1. In the bottom of the screen click on the arrow facing up, it will take you to the top of the page. Scroll down to the “ 1 Gene models” section of the gene page
2. **Navigating to the Genome Browser (JBrowse)**

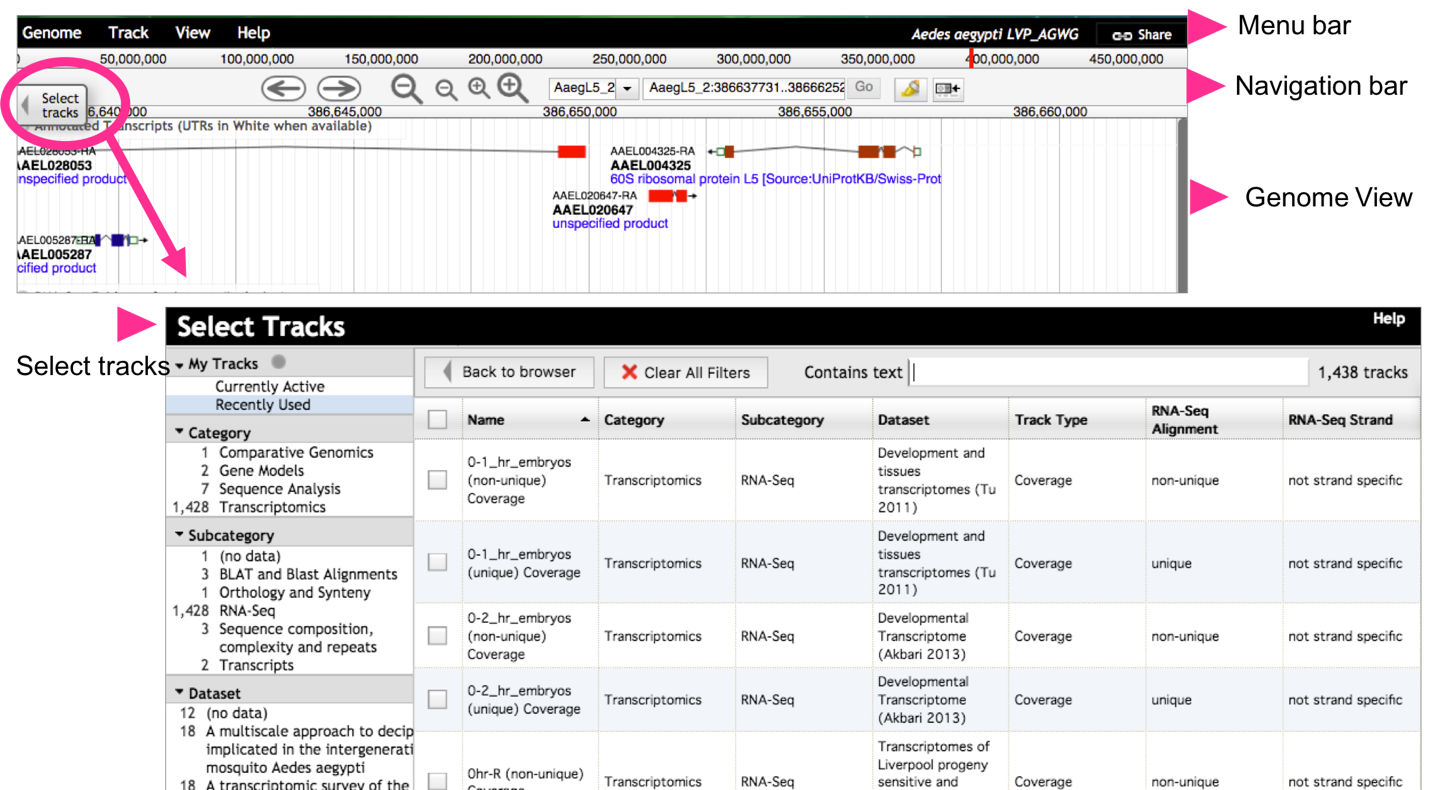
Links to the genome browser are available from multiple locations:

* 1. The tools menu in the header of any page.
  2. From record pages such as gene (and others SNP or genomic sequence) pages – these usually have links to a specific JBrowse configuration that includes data relevant to the section on that record page. For example (image below), a JBrowse link from the “1 Gene models” section on the gene page would display the gene of interest along with the “RNA-Seq Evidence for introns (Refined)”.

1. **Getting around JBrowse.**
   1. Use the gene AAEL004325 and get to the genome browser using any of the options mentioned above.
   2. Once in JBrowse examine the following features:

The **menu bar**: located at the top of the JBrowse frame. This includes five menus. What do these three of them do/provide?

|  |
| --- |
| * Genome menu: We are currently in *Aedes aegypti* genome, from here we can change to any of the 44 VectorBase genomes * Help menu: quick guide and documentation about JBrowse * Share link: link to share the current view with a friend |

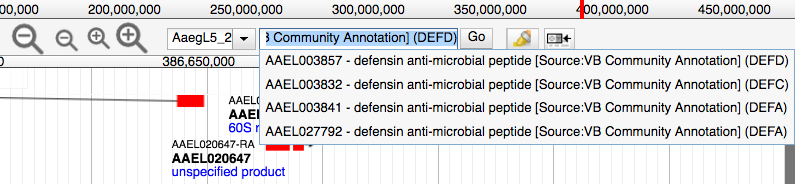
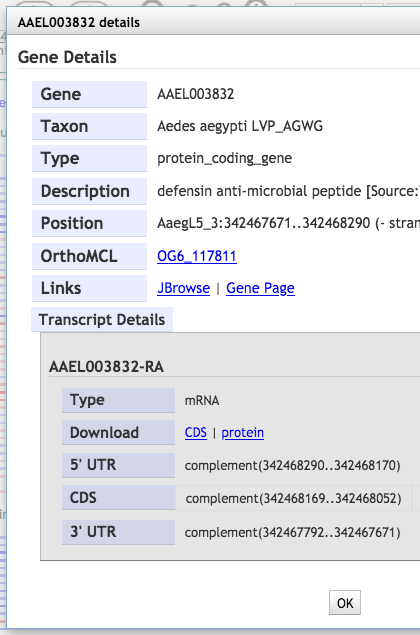
The **navigation bar**: located below the menu bar. This contains zooming (magnifying glass icons), panning (left/right arrows) and highlighting (yellow highlighter) buttons, reference sequence selector (drop down with sequences from the selected genome sorted by length), a text box to search for features such as gene IDs (image below) and overview bar which shows the location of the region in view.

* + 1. The **genome view**: this is where the data tracks are displayed.
  1. 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.

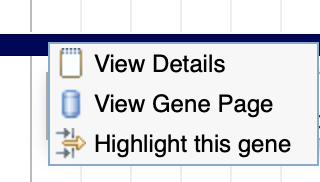
1. **Navigating to a specific gene in JBrowse**.

The goal of this step is to navigate to the anti-microbial peptide defensin C (DEFC) gene of *A. aegypti LVP\_AGWG*.

* 1. Make sure the *A. aegypti* genome is selected/shown in the menu bar.

* 1. Start typing the word *defensin* in the genome browser search box. After a few seconds you should see the result of the search (do not hit enter). Select the gene DEFC from the search dropdown. This will take you to gene AAEL003832
  2. 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?

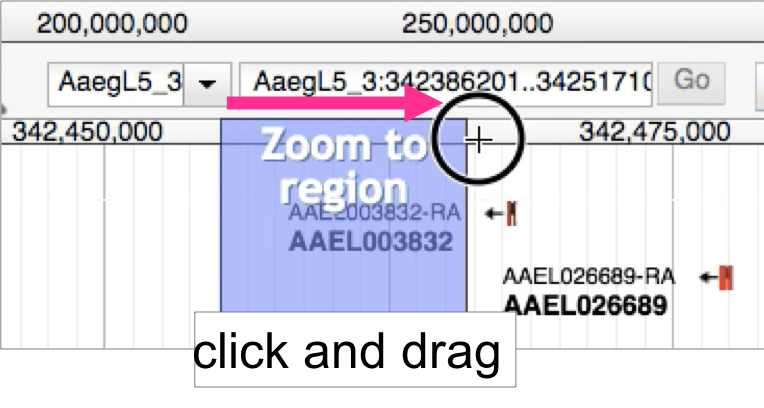
|  |
| --- |
| Gene details: gene, taxon, type, description, position, OrthoMCL, links  Transcript details: type, download, 5’UTR, CDS, 3’UTR |

* 1. 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).
  2. What genes are immediately upstream and downstream of DEFC? Provide gene IDs (Hint: use the zoom out button in the navigation bar).

|  |  |
| --- | --- |
| upstream: AAEL019922 | downstream: AAEL026689 |

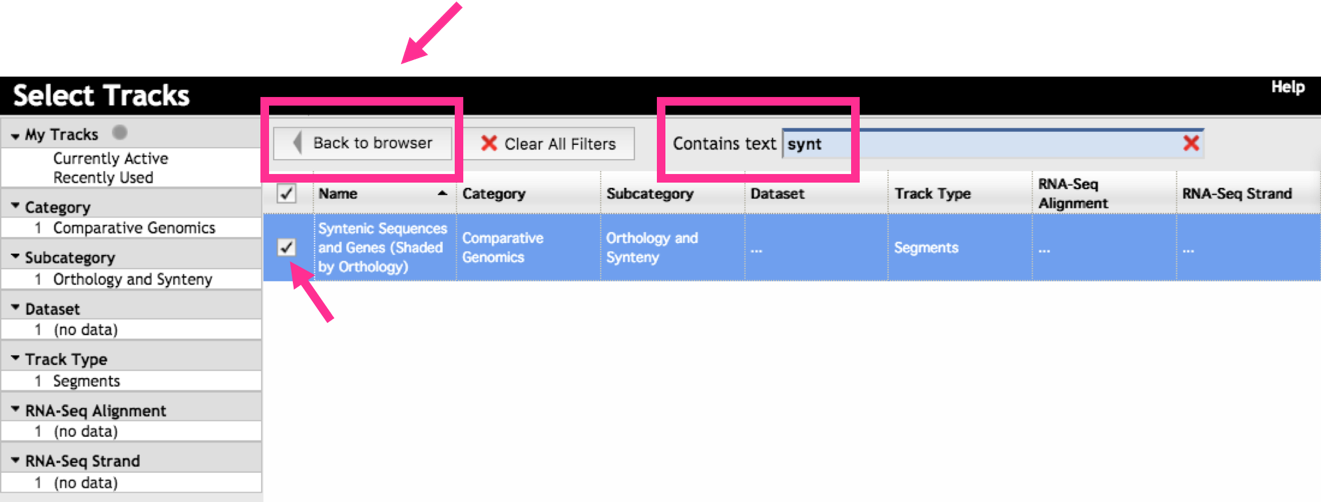
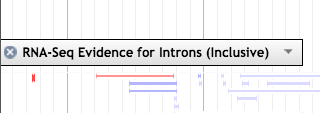
* 1. What is the difference between the small and large zoom buttons?

|  |
| --- |
| The magnitude of the zooming, it goes to a bigger or smaller region |

* 1. Another way to zoom in and out is by clicking on shift and the up or down arrows. You can also zoom in by clicking and dragging your cursor in the location ruler in the navigation bar).

1. **Exploring synteny between genomes.**

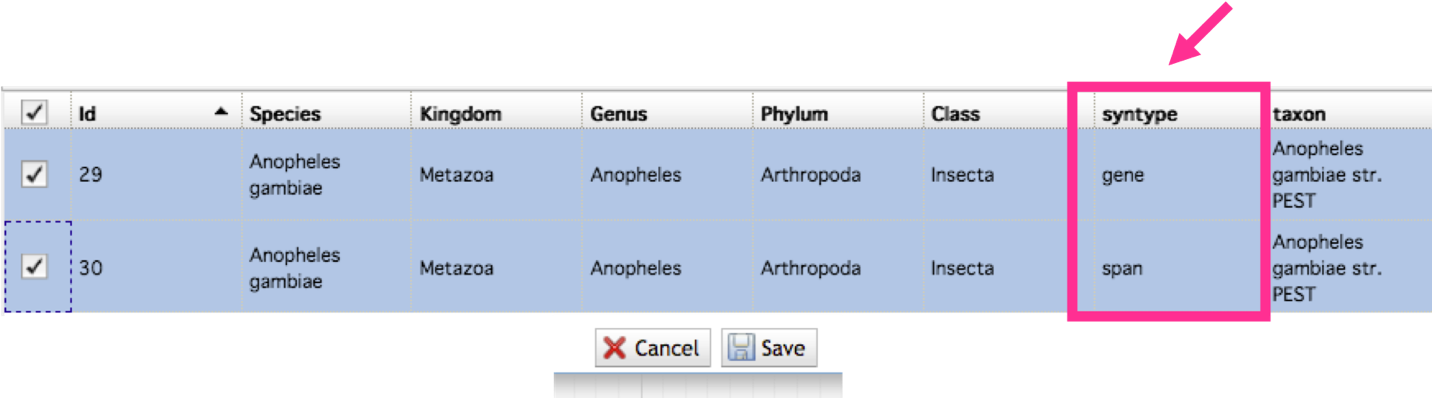
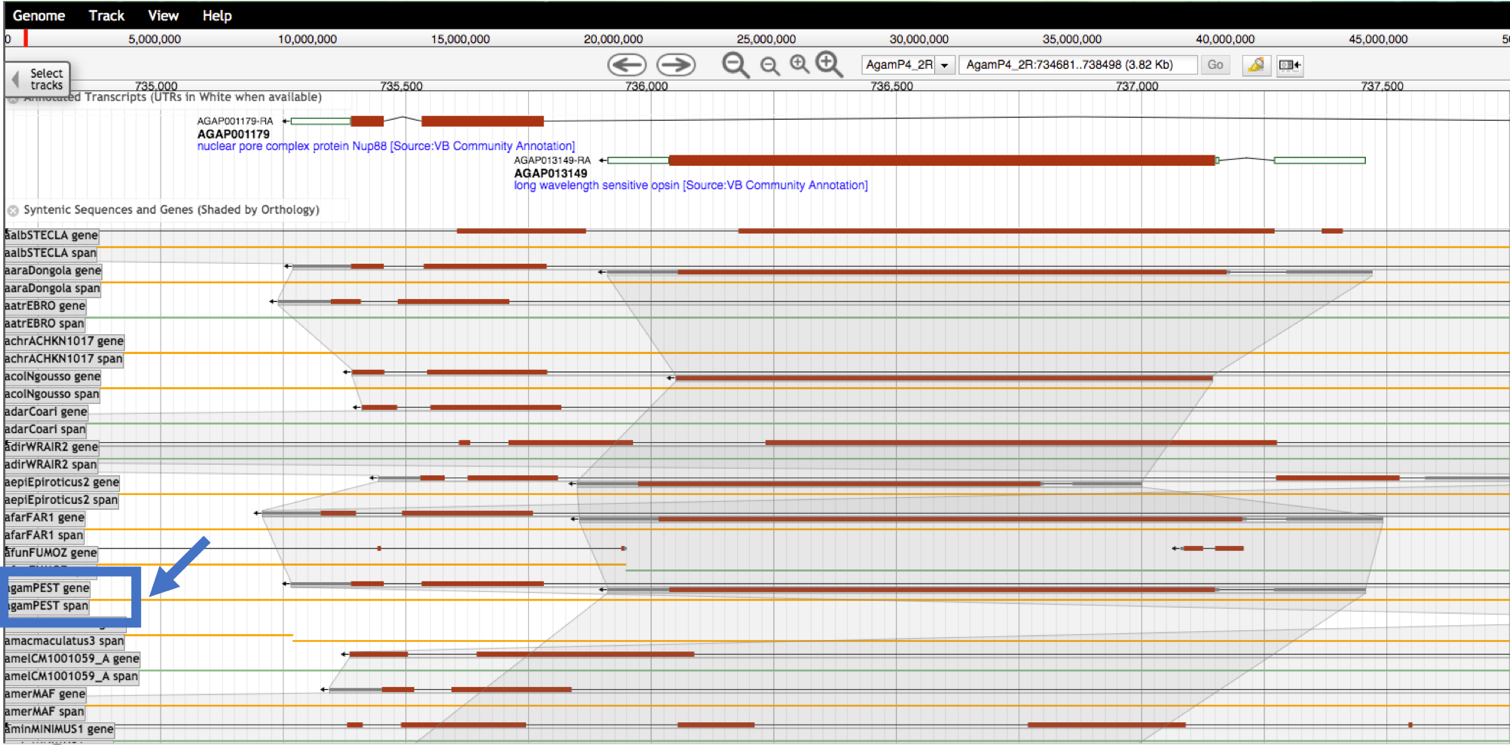
Synteny helps define conservation of homologous genes and gene order between genomes.

* Change to the genome of *Anopheles gambiae* PEST
* Go to the gene AGAP013149, long wavelength opsin, GPRop1
* Go to the “Select Tracks” tab on the left of the page and turn on the track called “Syntenic Sequences and Genes”. How did you find this track? One option is to click on the “Comparative Genomics” category on the left side to filter the tracks. Another option is to type a keyword syntenic in this page search box
* Return to the browser by clicking “Back to Browser” (image above)
* Zoom out so you can see a couple of genes on either side of GPRop1 (does not have to be exact)
* Delete other visible tracks, if available. Hover with the mouse over it, with a click on close/the X
* Which genus are currently available in the synteny track?

|  |
| --- |
| *Species from the genus Anopheles* |

* Configure the synteny track to delete *Anopheles sinensis* strain Chinaand strain SINENSIS.
  + To configure the subtracks:
    - Hover with the mouse over it, click on the down arrow in the track name



* + - Select the option called “Select Subtracks” from the menu
    - In the next popup delete the species mentioned above and click save (image below)
    - Note: In the opposite case, if you were to add a species, you should select both the “gene” and “span” subtracks for each new species; and click on the save button at the bottom of the popup.
* What does the synteny track in this region look like?
* Zoom out some more to a region of about ~7.4 kb with GPRop1 in the center. Are genes (in general) similarly organized between these species?

|  |
| --- |
| yes |

* What does the shading between genes mean?

|  |
| --- |
| shows the ortholog relationships |

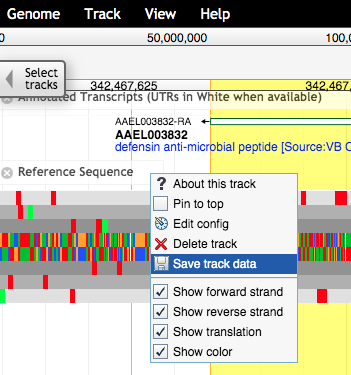
* In what direction is the GPRop1 gene relative to the chromosome?

|  |
| --- |
| Reverse |

* Go to a new gene, AGAP001181. Observe the synteny. What does synteny look like if you add more distantly related species? Does AGAP001181 appear to have orthologs in distantly related species such as *Cimex* and *Pediculus*?

|  |
| --- |
| synteny is is not as conserved with *Cimex* and *Pediculus*, there are not gray shading connections |

1. **Retrieving data from tracks** 
   1. Downloading sequence in FASTA format from a region of interest:

Go to *A. aegypti* gene DEFC (defensin C), AAEL003832

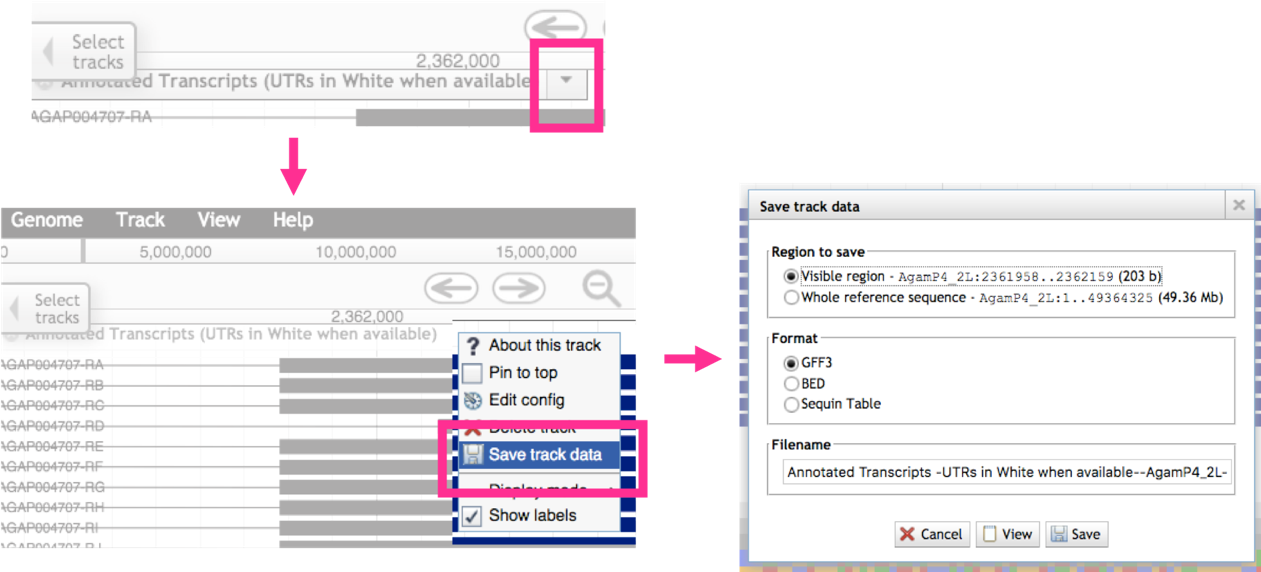
Make sure the “annotated transcripts” and the “reference sequence” tracks are turned on.

Turn off any other tracks

Select the gene, right click and select “highlight this gene”. It should turn yellow when activated.

Click on the down arrow on the reference sequence track and select “Save track data”.

In the next popup window you can keep everything as the default and either save or view the sequence. Note: If the popup window does not show for you please try page reload or clear browser cache

Alternatively, hover with the mouse over the annotated transcripts track > a drop-down arrow will appear > click on save track data. If you only want a specific region, exon or a complete gene, right click and select “highlight this gene” first.

1. The answers might change in future releases. VectorBase (and all other VEuPathDB sites) data are updated every two months. For information about the most recent release please visit: <https://beta.vectorbase.org/vectorbase.beta/app/static-content/VectorBase/news.html#VectorBase47Released> [↑](#footnote-ref-1)
2. Expression at the protein level and genetic variation/single nucleotide polymorphisms (SNPs), will be activated in VectorBase beta in a future release.  [↑](#footnote-ref-2)