

## JBrowse Basics

*Note: this exercise uses TriTrypDB (<https://TriTrypdb.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

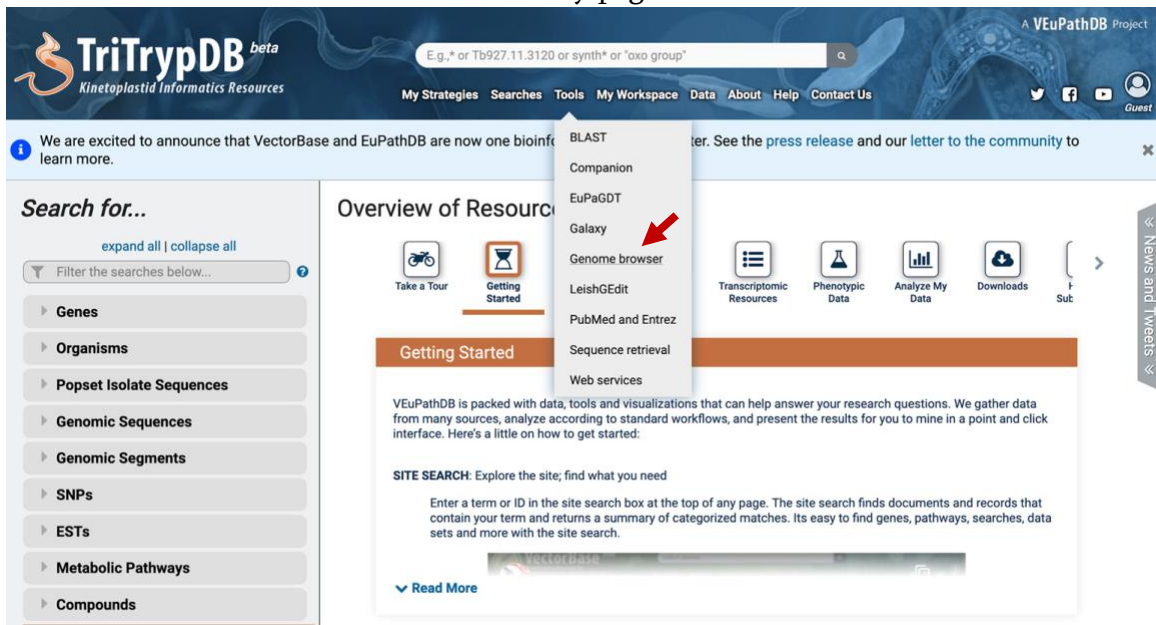
### 1. Navigating to the Genome Browser (JBrowse)

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”

## View in JBrowse genome browser

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.

The screenshot shows the JBrowse genome browser interface for *Trypanosoma brucei*. The top menu bar includes Genome, Track, View, and Help. The navigation bar below it contains zooming and panning tools, a reference sequence selector, and a search box. The genome view displays various tracks, including mRNA and ncRNA. A red circle highlights the 'Select tracks' button, and a red arrow points to the 'Select Tracks' panel at the bottom. The 'Select Tracks' panel shows a list of tracks with checkboxes and a table of track details.

Name	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand
3-end oligo-dT (non-unique) Coverage	Transcriptomics	RNASeq	Procyclic form 5 and 3 end-enriched transcriptomes	Coverage	non-unique	not strand specific
3-end oligo-dT (unique) Coverage	Transcriptomics	RNASeq	Procyclic form 5 and 3 end-enriched transcriptomes	Coverage	unique	not strand specific
3-end random (non-unique) Coverage	Transcriptomics	RNASeq	Procyclic form 5 and 3 end-enriched transcriptomes	Coverage	non-unique	not strand specific
3-end random (unique) Coverage	Transcriptomics	RNASeq	Procyclic form 5 and 3 end-enriched transcriptomes	Coverage	unique	not strand specific
427 Coverage and Alignments	Genetic Variation	DNASeq	Aligned HTS SNP calls for Lister strain 427	Coverage (Read Alignments zoomed)	...	...
5-SL-end (non-unique) Coverage	Transcriptomics	RNASeq	Procyclic form 5 and 3 end-enriched transcriptomes	Coverage	non-unique	not strand specific
5-SL-end (unique) Coverage	Transcriptomics	RNASeq	Procyclic form 5 and 3 end-enriched transcriptomes	Coverage	unique	not strand specific
5-triphosphate			Procyclic form 5			

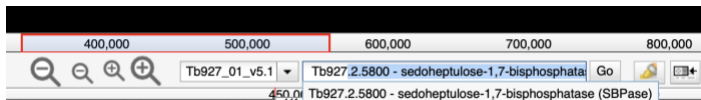
- c. Selecting tracks: click on the “select track” button (top left). You can search/filter for tracks and then select them for display by checking the check box next to the track name.

### 3. Navigating to a specific gene in JBrowse.

The goal of this step is to navigate to the sedoheptulose-1,7-bisphosphatase (SBPase) gene of *T. brucei* 927.

- a. Make sure the *Trypanosoma brucei brucei* TREU927 genome is selected from the genome menu.
- b. Start typing the word sedoheptulose in the search box. After a few seconds you should see the result of the search (do not hit enter). Select the gene from the search dropdown. This will take you to Tb927.2.5800.

Genome	Track	View	Help
<input type="checkbox"/>	Leishmania major strain Friedlin		
<input type="checkbox"/>	Leishmania major strain LV39c5		
<input type="checkbox"/>	Leishmania major strain SD 75.1		
<input type="checkbox"/>	Leishmania mexicana MHOM/GT/2001/U1103		
<input type="checkbox"/>	Leishmania panamensis MHOM/COL/81/L13		
<input type="checkbox"/>	Leishmania panamensis strain MHOM/PA/94/PSC-1		
<input type="checkbox"/>	Leishmania sp. MAR LEM2494		
<input type="checkbox"/>	Leishmania tarentolae Parrot-Taril		
<input type="checkbox"/>	Leishmania tropica L590		
<input type="checkbox"/>	Leishmania turanica strain LEM423		
<input type="checkbox"/>	Leptomonas pyrrhocoris H10		
<input type="checkbox"/>	Leptomonas seymouri ATCC 30220		
<input type="checkbox"/>	Paratrypanosoma confusum CUL13		
<input type="checkbox"/>	Trypanosoma brucei Lister strain 427		
<input type="checkbox"/>	Trypanosoma brucei Lister strain 427 2018		
<input checked="" type="checkbox"/>	Trypanosoma brucei brucei TREU927		
<input type="checkbox"/>	Trypanosoma brucei gambiense DAL972		
<input type="checkbox"/>	Trypanosoma congolense IL3000		
<input type="checkbox"/>	Trypanosoma cruzi Bug2148		
<input type="checkbox"/>	Trypanosoma cruzi CL Brener Esmeraldo-like		
<input type="checkbox"/>	Trypanosoma cruzi CL Brener Non-Esmeraldo-like		
<input type="checkbox"/>	Trypanosoma cruzi Dm28c 2014		
<input type="checkbox"/>	Trypanosoma cruzi Dm28c 2017		



- c. 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

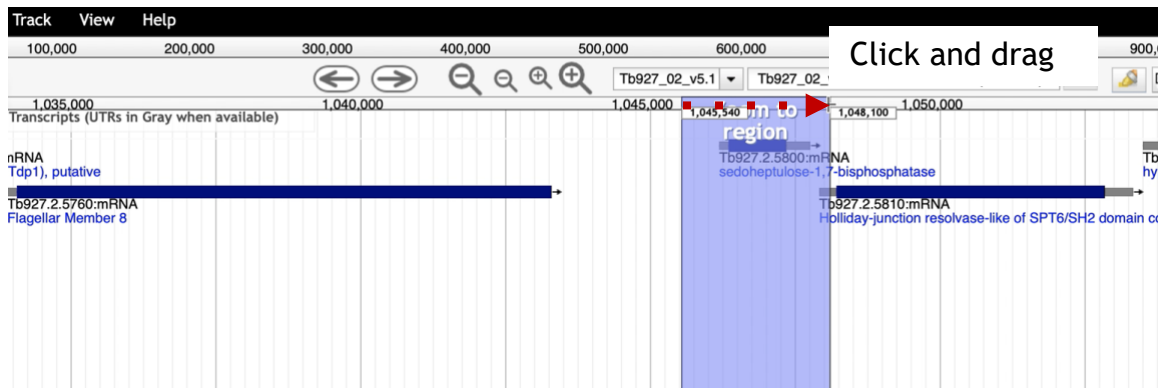
**Tb927.2.5800:mRNA details**

- Species: Trypanosoma brucei brucei TREU927
- ID: Tb927.2.5800:mRNA
- Gene ID: Tb927.2.5800
- Gene Type: Protein Coding
- Description: sedoheptulose-1,7-bisphosphatase
- 5' UTR: 1046195..1046356
- CDS: 1046357..1047355
- 3' UTR: 1047356..1047765
- Download: [CDS](#) | [protein](#)
- OrthoMCL: [OG5\\_134853](#)
- Links: [JBrowse](#) | [Gene Page](#)

- d. 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).
- e. What genes are immediately upstream and downstream of SBP? (Hint: 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

- [View Details](#)
- [View Gene Page](#)
- [Highlight this gene](#)

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 transcription start sites.

Are you confident about the gene transcription start? (Note: gene features are in blue (left to right) or red (right to left) with untranslated regions (UTRs) in grey). What additional data track would be useful for you to assess this? (hint: Click on the “Select Tracks” button to reveal all available tracks. Now type the word “splice” in the “contains text” box. This will filter all tracks that contain the word splice. Find the one called “Unified Splice Leader Addition Sites” and select it. Click on the “Back to browser” button). What do the different diamond colors mean? Click on them and see

if you can figure this out from the popups? Which color provides the most evidence

**Select Tracks** Help

My Tracks Back to browser Clear All Filters Contains text: **splice** 7 matching tracks

Category: 7 Gene Models  
Subcategory: 2 Poly A Sites, 5 Splice Sites  
Dataset: 7 (no data)  
Track Type: 7 Segments  
RNASeq Alignment: 7 (no data)  
RNASeq Strand: 7 (no data)

Name	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand
<input type="checkbox"/> Bloodstream and Procyclic for spliced leader transcripts (927, 427)(2014) - Splice Sites	Gene Models	Splice Sites	...	Segments	...	...
<input type="checkbox"/> Bloodstream and procyclic form spliced leader transcripts (427, Antat) (2010) - Splice Sites	Gene Models	Splice Sites	...	Segments	...	...
<input type="checkbox"/> Curated Poly A Sites from bloodstream and procyclic forms	Gene Models	Poly A Sites	...	Segments	...	...
<input type="checkbox"/> Procyclic form spliced leader transcriptome - Poly A Sites	Gene Models	Poly A Sites	...	Segments	...	...
<input type="checkbox"/> Procyclic form spliced leader transcriptome - Splice Sites	Gene Models	Splice Sites	...	Segments	...	...
<input type="checkbox"/> Spliced Leader and Poly A Sites from bloodstream and procyclic forms - Splice Sites	Gene Models	Splice Sites	...	Segments	...	...
<input checked="" type="checkbox"/> Unified Spliced Leader Addition Sites	Gene Models	Splice Sites	...	Segments	...	...

Tb927\_02\_v5.1 | Tb927\_02\_v5.1:1044924..1048082 (3.16 Kb) | Go

1,046,500 1,047,000 1,047,500

Tb927.2.5800:mRNA  
sedoheptulose-1,7-bisphosphatase

Tb927.2.5800(11) details

Location: 1046195  
Gene ID: Tb927.2.5800  
UTR Length: 162  
Count: 128.25999999999996  
Note: The overall count is the sum of the count per million for each sample.

Sample	Count per million
T.brucei 427 cBF	7.32
T.brucei 5-SL-end-enriched cDNA	62.91
T.brucei 927 PCF	6.72
T.brucei 927 sIBF	8.15
T.brucei Alba 1+ (29-13 RNAi)	13.9
T.brucei Alba 1- (29-13 RNAi)	4.85
T.brucei Alba 3_4+ (29-13 RNAi)	2.02
T.brucei bloodstream (Lister 427)	4.66
T.brucei bloodstream long slender (Antat1.1)	5.6
T.brucei bloodstream short stumpy (Antat1.1)	4.74
T.brucei curated splice site cDNA	7.39

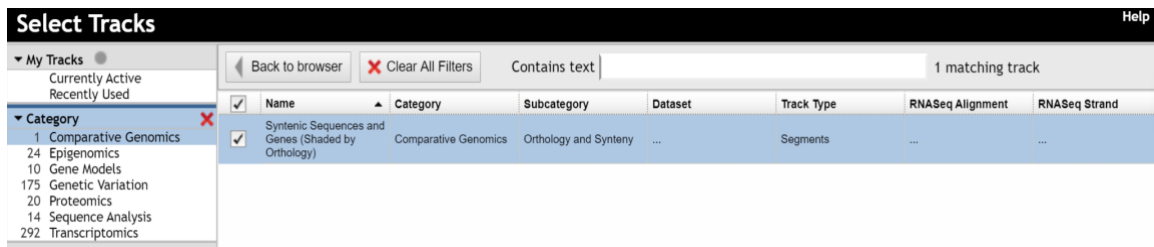
OK

for a splice junction?

## 5. Exploring synteny between genomes.

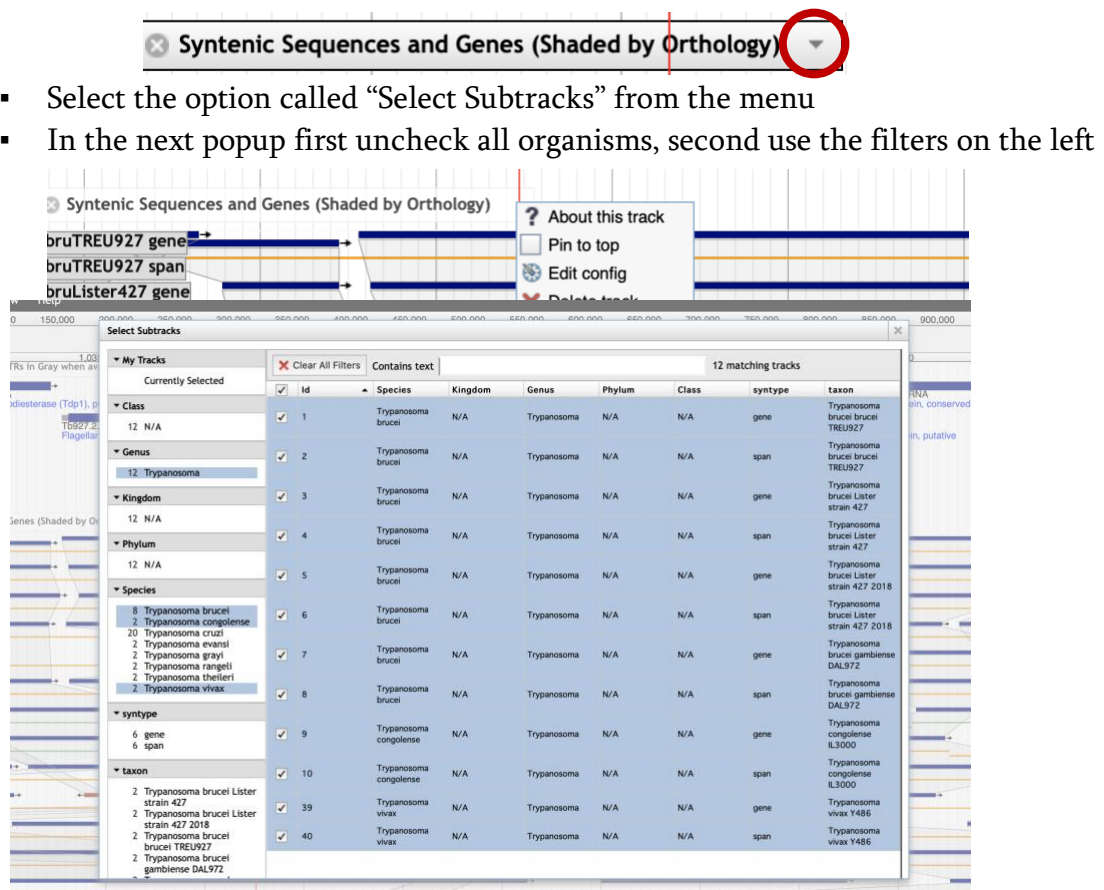
Synten helps define conservation of homologous genes and gene order between genomes.

- 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.

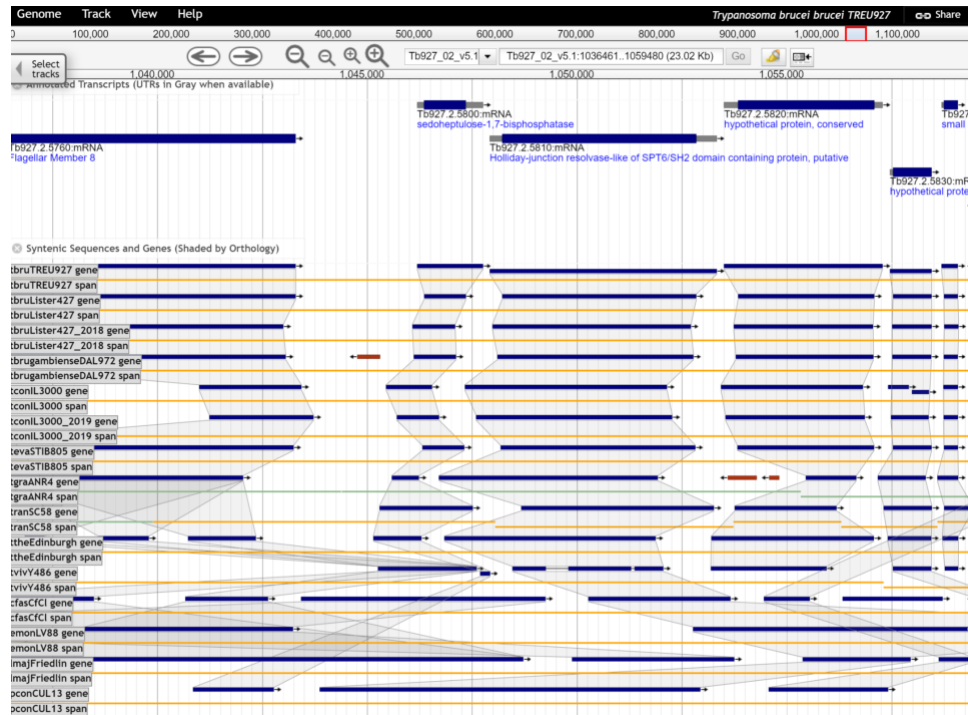
- Return to the browser by clicking “Back to Browser” and zoom out so you can see a couple of genes on either side of SBP (does not have to be exact)
- Configure the synteny track to include the following species subtracks: *Trypanosoma brucei* 927, *T. brucei* 427, *T. brucei gambiense*, *T. congolense*, *T. evansi*, *T. grayi*, *T. theileri* and *T. vivax*.
  - To configure the subtracks:
    - Click on the down arrow in the track name



to select *Trypanosoma*, third select the species of interest (note that you should select both the gene and span subtracks for each species), fourth click on the save button at the bottom of the popup.

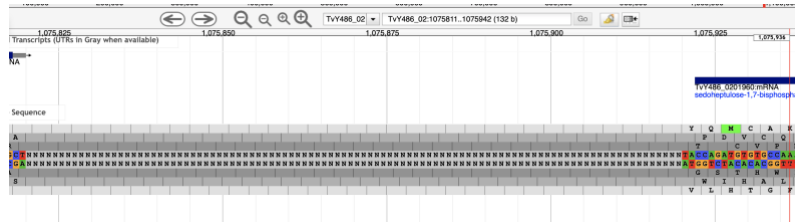


- What does the syntenic track in this region look like? Feel free to zoom out some more. Are genes (in general) similarly organized between these species? What does the shading between genes mean?
- What direction is the SBPase gene relative to the chromosome?
- What genes are upstream and downstream of the SBPase? Are these genes syntenic?
- What does syntenic look like if you add more distantly related species? Does SBPase appear to have orthologs in *Leishmania*? *Endotrypanum*? *Crithidia*?



- Examine the gene corresponding to the *T. vivax* SBPase in the syntenic track. Hover over the gene image to find the gene name in the popup. Does this gene appear to be a fragment? What could be some possible reasons for this?
- Do you think all the genomes in the database are fully sequenced? Is it possible that gaps in sequence exist in the available genomes? Let's find out if there is a gap next to the SBPase gene in *T. vivax*:
  - Select *T. vivax* from the list of genomes in the menu bar.
  - Turn on the **annotated transcripts** and the **Reference sequence** tracks.
  - Search for the SBPase gene by typing "sedoheptulose" in the search box then select the gene.

- Zoom to about 600bps. Do you see something missing on the left side of the gene?
- Zoom in to this area (click and drag). What do you see? What do all of these Ns mean?





## 6. Exploring other data tracks in JBrowse.

For this example, we will view *T. brucei* data, so the data tracks you turn on will display data only if the data is aligned to the *T. brucei* genome. Return to the SBPase gene in *T. brucei* by searching for the gene ID in the (Tb927.2.5800) in 'Landmark or Region' to redirect the browser. Then zoom to the area between 0.7M and the end of the chromosome.

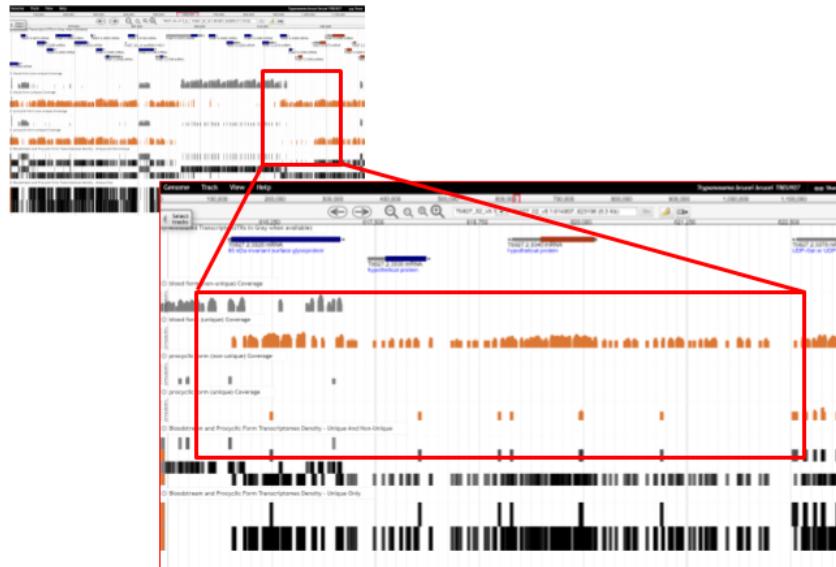
Turn on the ChIP-seq coverage plots and turn off the syntenic gene and region tracks. The data tracks are from an experiment called: **ChIP-Seq - Four histone Variants ChIP-Seq Coverage aligned to T brucei TREU927 (Cross) (linear plot)**. For this experiment, chromatin was immunoprecipitated using several different histone antibodies. The DNA that precipitated with the histone was sequenced and aligned to the *T. brucei* TREU927 genome. Peaks in the sequence coverage plots represent areas of histone binding. Different histone variants can be associated with start and termination sites for transcription ( <http://www.ncbi.nlm.nih.gov/pubmed/19369410>)

Name	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand
<input checked="" type="checkbox"/> BDF3-HA (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> ChIP-Seq - Four histone Variants Density - Unique And Non-Unique	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Multi-Density		
<input checked="" type="checkbox"/> H2A (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> H2B (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> H3K4me3 (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> H3v (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> H4 (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> H4K10ac (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
<input checked="" type="checkbox"/> H4v (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		

- You may need to adjust the y-axis scaling to bring the tracks into proper view (try setting the score range to “global” by mousing over the track name, clicking the dropdown arrow and selecting “Change Score Range”).
- What does this data show you?
- Roughly how many polycistronic units does this chromosome have? Zoom out to the entire chromosome.



- Move to the **region around 0.6Mbs of the chromosome** (you should be on chromosome 2) and turn on all four subtracks. Take note of the orange and grey bars in the coverage plots. What do you think the grey bars indicate?
- Now zoom out to 100Kb – do you see a difference between the blood and procyclic forms?

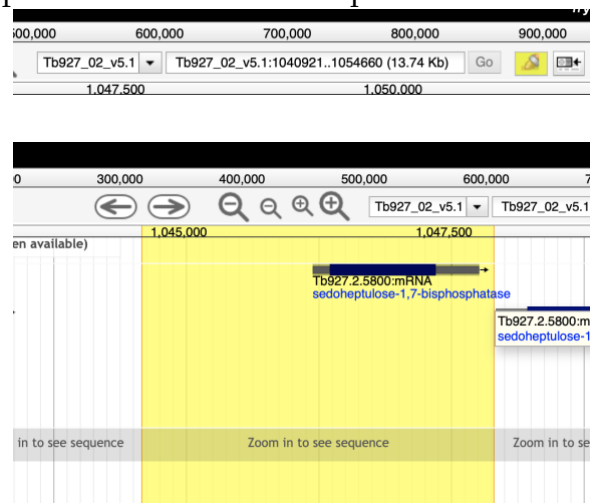


- Zoom in to a gene that looks like it is differentially expressed. What are your conclusions? Are the reads supported by unique or non-unique reads?
- Can you turn on additional tracks that may give some more support to your conclusions?  
Hint: turn on the EST and *T. brucei* protein expression evidence tracks.
  - Is there any proteomics evidence for this region?
  - How about EST evidence? Click on an EST graphic (glyph) to get additional information.
- Turn off the RNA-seq graphs and make sure the *T. brucei* protein expression evidence tracks are on. **Zoom out to 500Kb**. Explore the evidence for gene expression based on mapped peptides from proteomics experiments – which gene in this view has the highest number of peptide hits? Try looking at the “All MS/MS peptides (feature density)” track for an overview.



## 7. Retrieving data from and uploading your own tracks to JBrowse

- Downloading sequence in FASTA format from a region of interest:
  - Make sure the “annotated transcripts” and the “reference sequence” tracks are turned on.
  - Click on the “highlight a region” button in the navigation bar. It should turn yellow when activated.
  - Click and drag in the genome view region and select the area you would like to highlight.
  - 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.



- b. Uploading data to JBrowse:

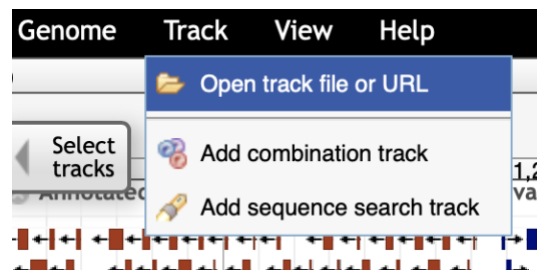
JBrowse can accept several standard-format data files by direct upload or through a URL if the data is stored remotely. Some file formats like BAM and VCF require indexing before uploading. In this exercise we will download a bigwig file from GEO and then upload it to JBrowse:

- i. Go to this GEO sample record:  
<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2407365>
- ii. Scroll down to the bottom of the page and download the bigwig file with the http link.

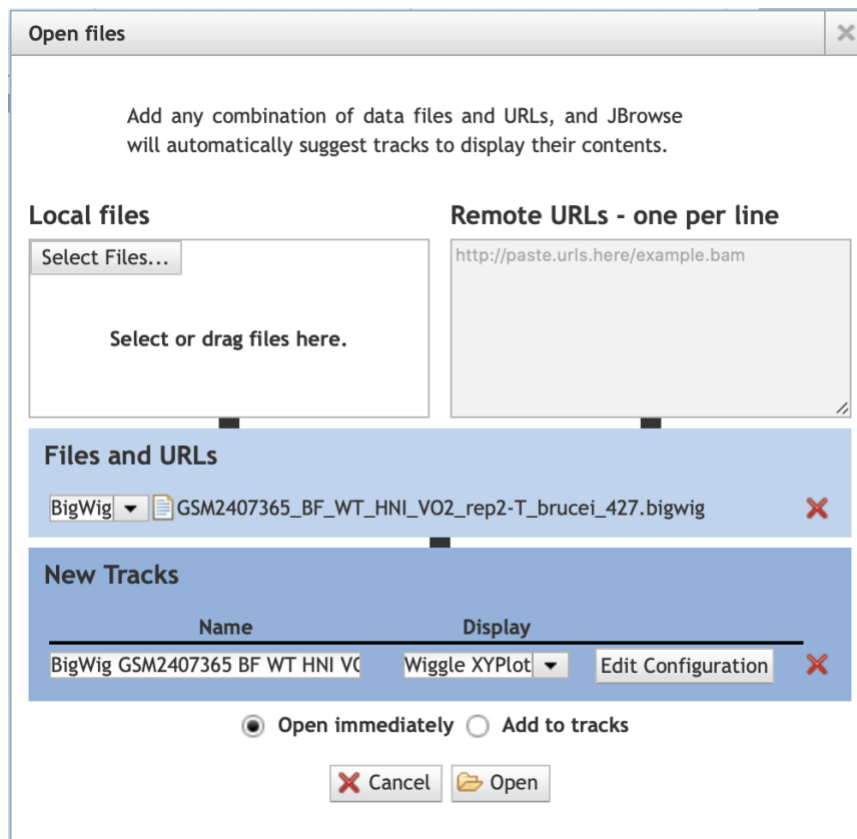
Supplementary file	Size	Download	File type/resource
GSM2407365_BF_WT_HNI_VO2_rep2-T_brucei_427.bigwig	12.4 Mb	(ftp)(http)	BIGWIG

- iii. Once the file is downloaded go to JBrowse and select *Trypanosoma brucei* Lister 427 as the reference genome (hint: use the Genome link in the menu panel, top left).
- iv. Turn on the track for annotated transcripts if it is not on already.

- v. Click on the Tracks menu item and select “Open track file or URL”.
- vi. In the popup click on select file then select the file you just downloaded. JBrowse should automatically recognize that the file is in bigwig format.



- vii. Click on the Open button. The bigWig output should appear very quickly in



your browser.

