

Browser Exercises - I

Alignments and Comparative genomics

1. Navigating to the Genome Browser (JBrowse)

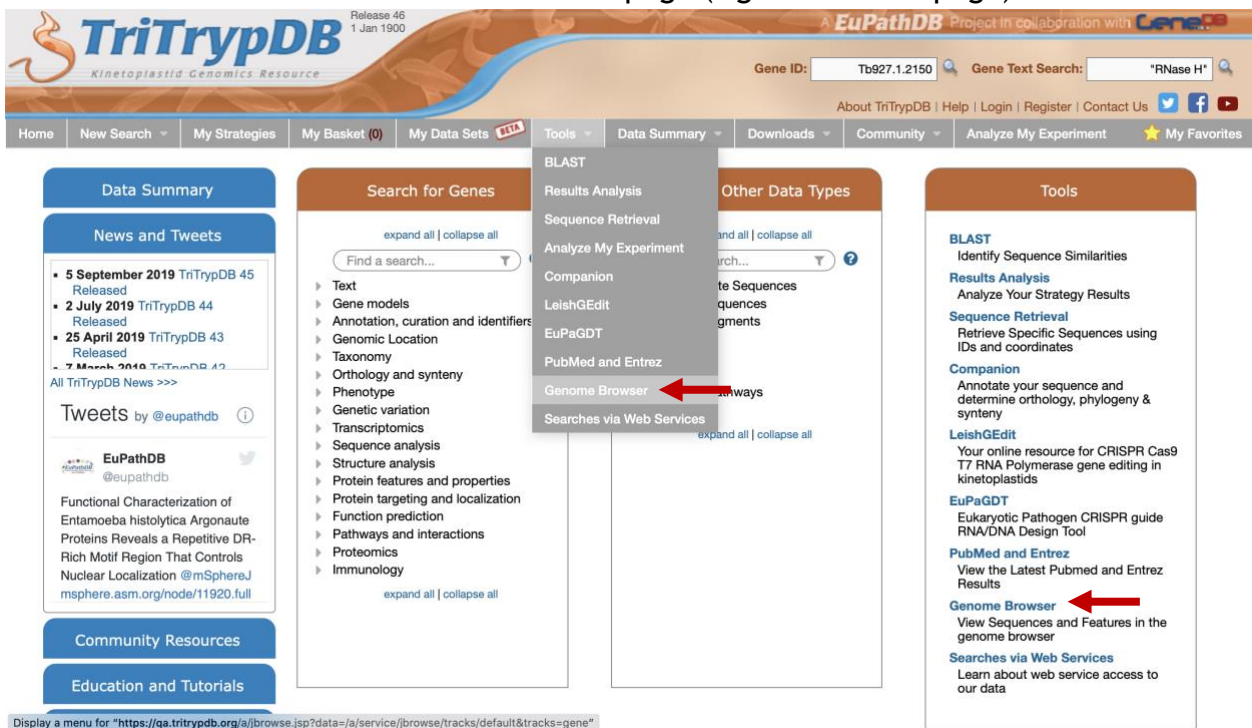
Note: For this exercise use <http://www.tritrypdb.org>

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:

- The tools menu in the grey menu bar on any page.
- The tools section on the home page (right center of page).



- From record pages such as gene, SNP or genomic sequence pages - these links are usually to specific JBrowse configuration that include 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.

View in JBrowse genome browser

2. Getting around JBrowse.

- Use any of the above described JBrowse linking strategies to get to the genome browser.
- Once in JBrowse examine the following features:
 - 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?
 - 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.
 - The **genome view**: this is where the data tracks are displayed.

The screenshot shows the JBrowse genome browser interface for *Trypanosoma brucei brucei* TREU927. The interface is divided into several sections:

- Menu bar**: Located at the top, containing 'Genome', 'Track', 'View', 'Help', and a 'Share' link.
- Navigation bar**: Located below the menu bar, containing zooming (magnifying glass icons), panning (left/right arrows), and highlighting (yellow highlighter) buttons. It also includes a 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 an overview bar which shows the location of the region in view.
- Genome view**: This is where the data tracks are displayed. It shows a genomic track with various features, including transcripts (mRNA, ncRNA) and their coordinates.
- Select Tracks menu**: A pop-up window showing a list of tracks and a table of track details. The table has columns for Name, Category, Subcategory, Dataset, Track Type, RNASeq Alignment, and RNASeq Strand.

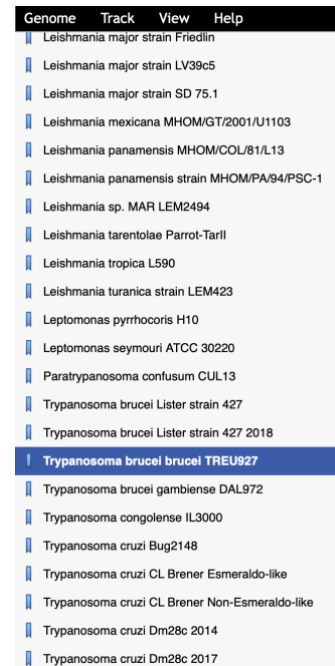
Name	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand
3-end oligo-dT (non-unique) Coverage	Transcriptomics	RNASeq	Procytic form S and 3 end-enriched transcriptomes	Coverage	non-unique	not strand specific
3-end oligo-dT (unique) Coverage	Transcriptomics	RNASeq	Procytic form S and 3 end-enriched transcriptomes	Coverage	unique	not strand specific
3-end random (non-unique) Coverage	Transcriptomics	RNASeq	Procytic form S and 3 end-enriched transcriptomes	Coverage	non-unique	not strand specific
3-end random (unique) Coverage	Transcriptomics	RNASeq	Procytic form S 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	Procytic form S and 3 end-enriched transcriptomes	Coverage	non-unique	not strand specific
5-SL-end (unique) Coverage	Transcriptomics	RNASeq	Procytic form S and 3 end-enriched transcriptomes	Coverage	unique	not strand specific
5-triphosphate-			Procytic form S			

- 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.

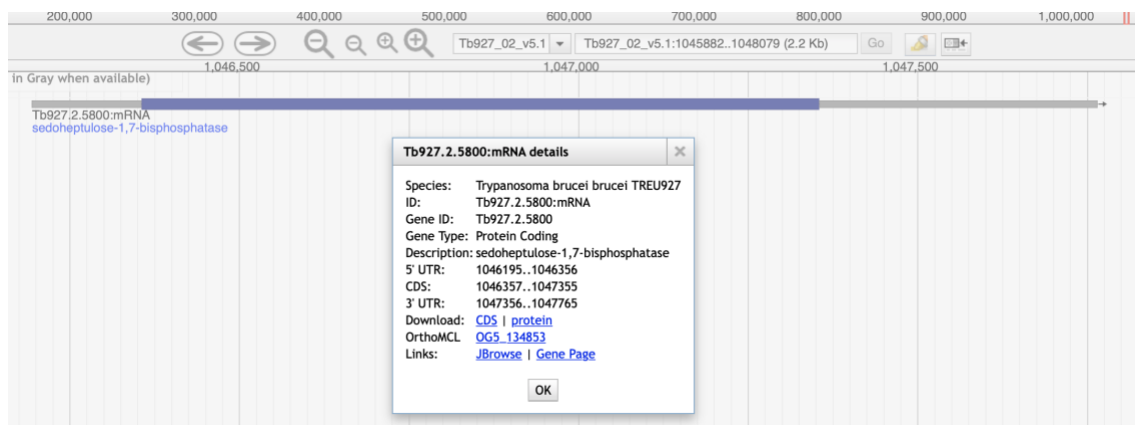
3. Navigating to a specific gene in JBrowse.

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

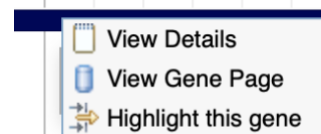
- Make sure the *Trypanosoma brucei* brucei TREU927 genome is selected from the genome menu.
- Start typing the word sedoheptulose in the search box. After a few second you should see the result of the search (do not hit enter). Select the gene from the search dropdown. This will take you to Tb927.2.5800.



- 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? You can

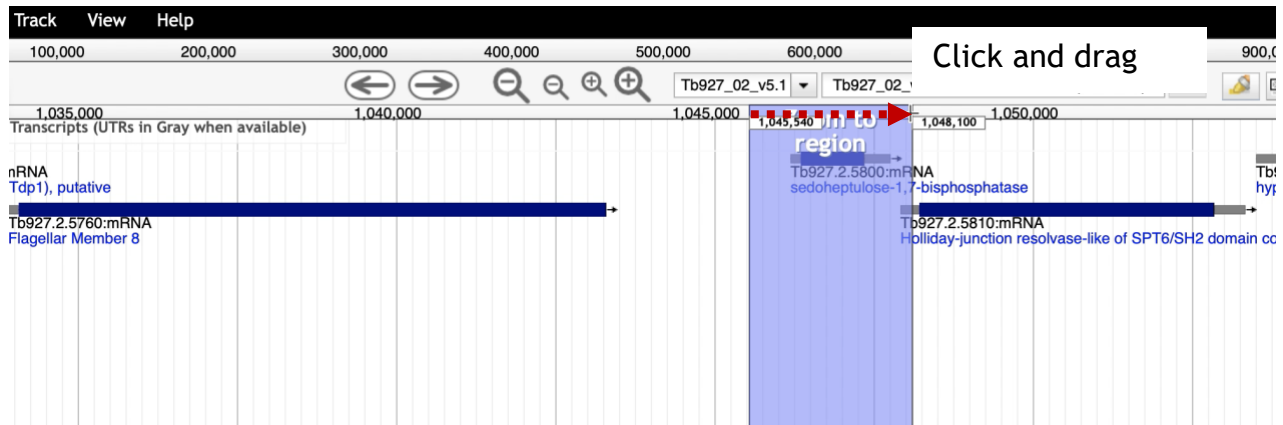


also right click (or control click) on a feature to display the context menu which provides quick links to highlight a feature, go to the feature page (like the gene page) or get the info popup (the same one you get when you click on the feature).



- 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 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.

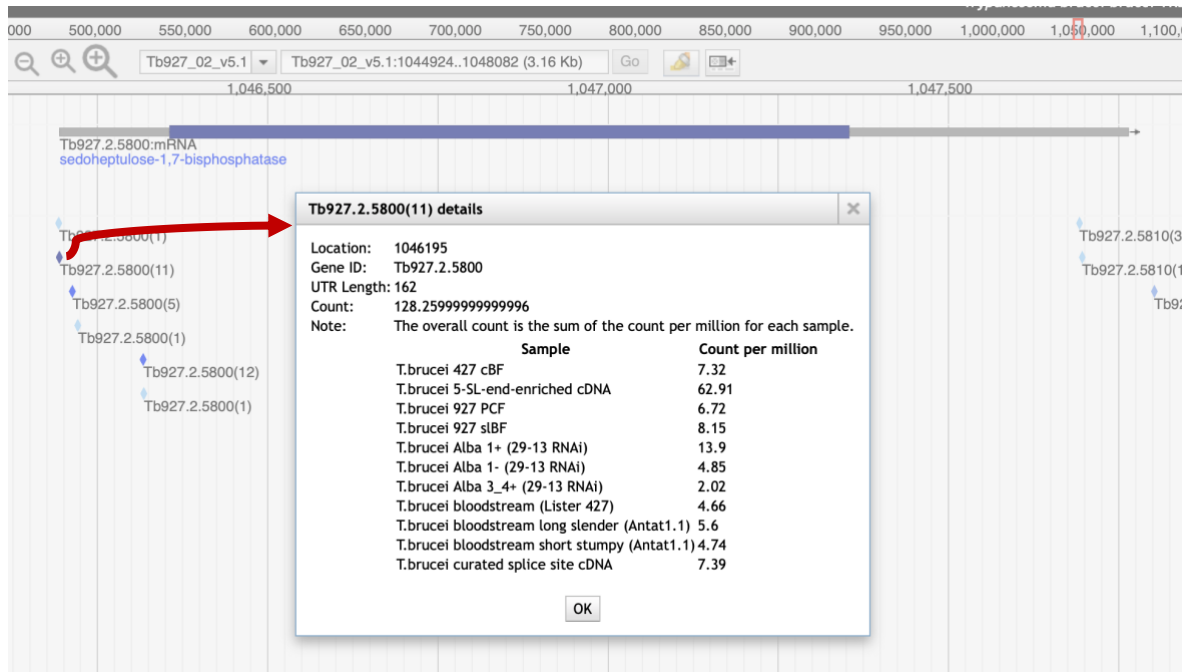
Are you confident about the gene transcription start? (Note: gene features are in blue (left to right) or red (right to left) with untranslated regions (UTRs) in grey). What additional data track would be useful for you to assess this? (hint: Click on the “Select Tracks” button to reveal all available 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).

Select Tracks

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

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

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 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.

Select Tracks

My Tracks: Currently Active, Recently Used

Category: Comparative Genomics

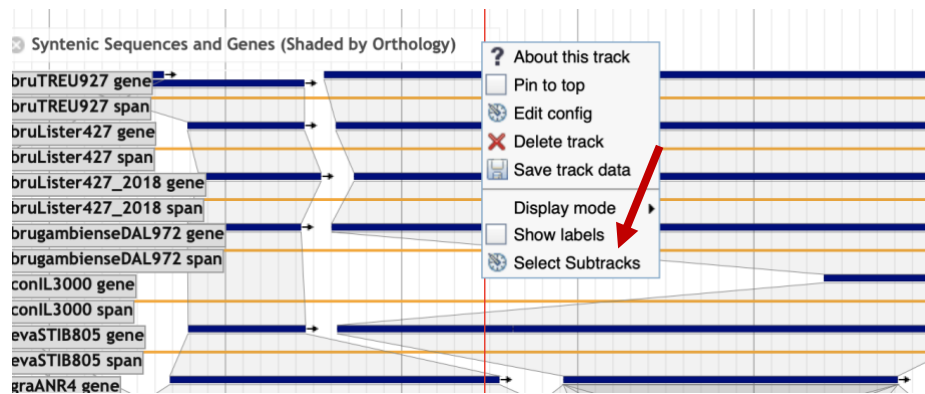
Name	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand
Syntenic Sequences and Genes (Shaded by Orthology)	Comparative Genomics	Orthology and Synteny	...	Segments

- 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*.

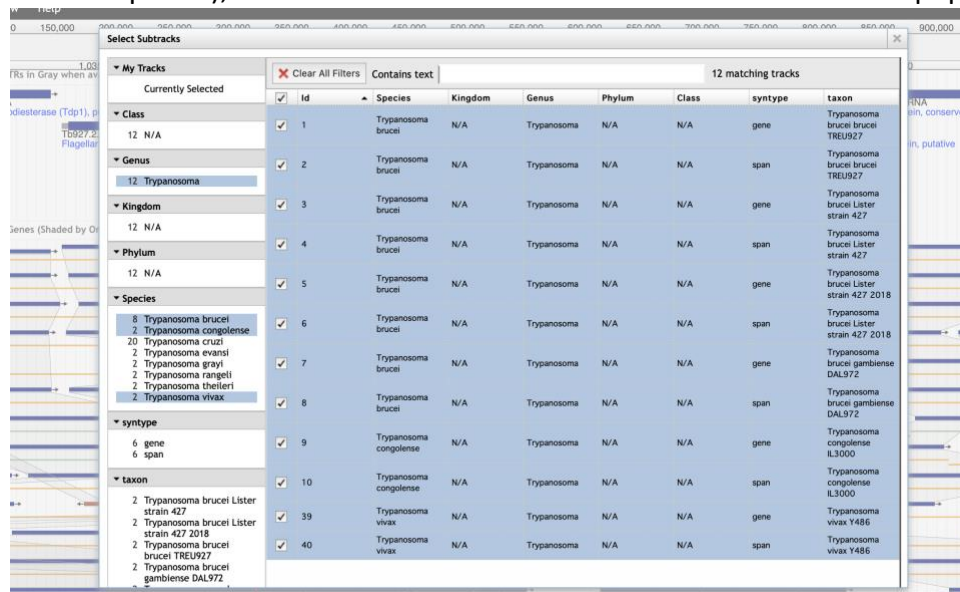
- To configure the subtracks:
 - Click on the down arrow in the track name



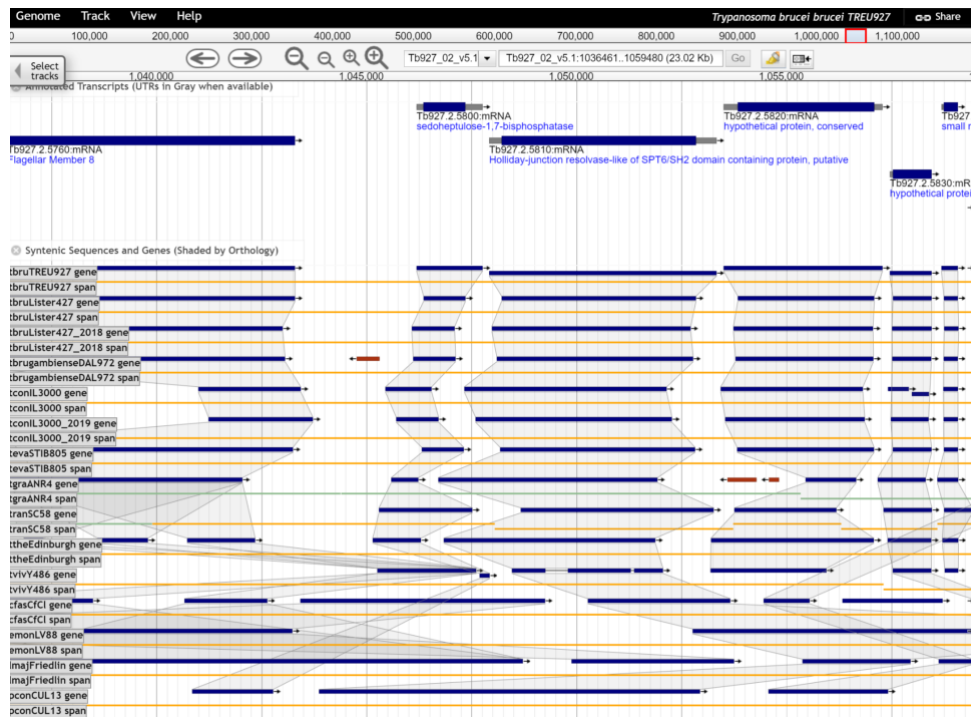
- Select the option called “Select Subtracks” from the menu



- In the next popup first uncheck all organisms, second use the filters on the left 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.

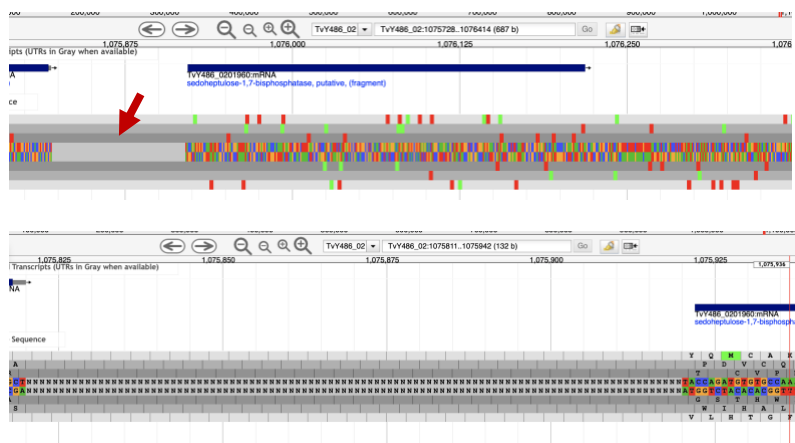


- What does the synteny track in this region look like? Feel free to zoom out some more. Are genes (in general) similarly organized between these species? What does the shading between genes mean?
- What direction is the SBPase gene relative to the chromosome?
- What genes are upstream and downstream of the SBPase? Are these genes syntenic?



- What does synteny 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 synteny track. Hover 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. The TriTrypDB data set record is here:

http://tritrypdb.org/tritrypdb/app/record/dataset/DS_aabc052008 and the publication is here: <http://www.ncbi.nlm.nih.gov/pubmed/19369410>

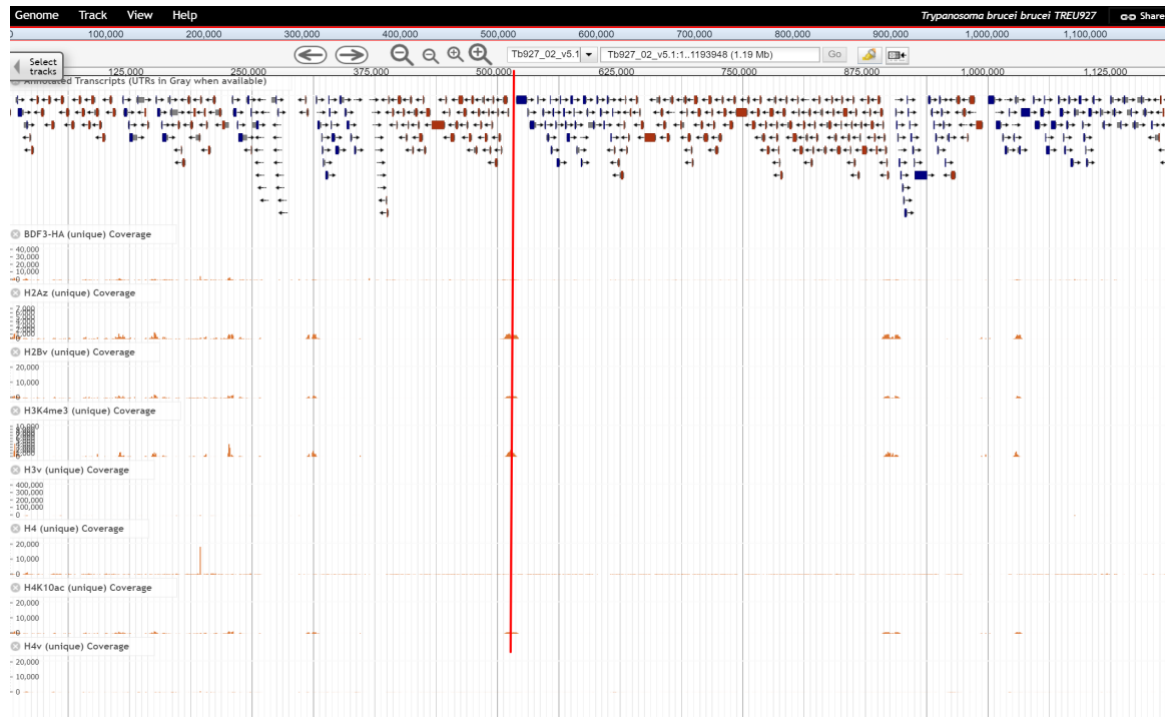


Select Tracks

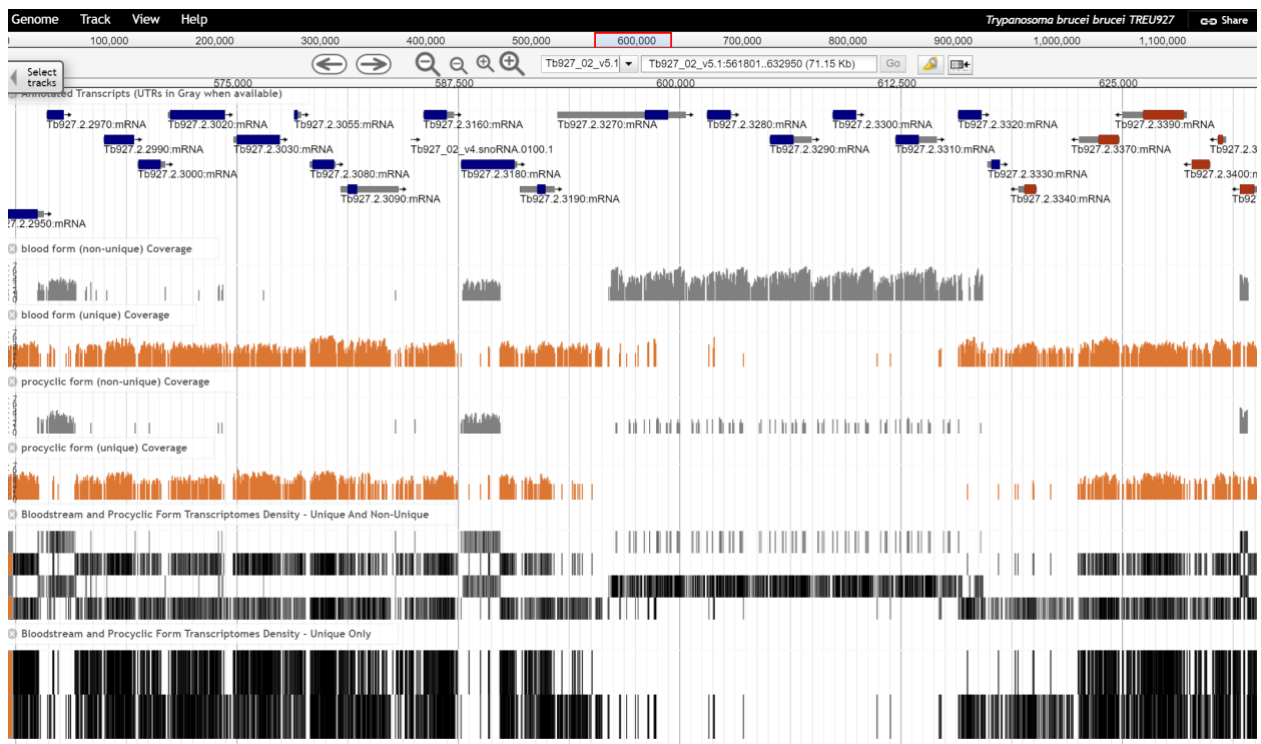
Back to browser Clear All Filters Contains text **Four histone variants** matching tracks

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"/> H2Az (unique) Coverage	Epigenomics	ChIPSeq	ChIP-Seq - Four histone Variants	Coverage
<input checked="" type="checkbox"/> H2Bv (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

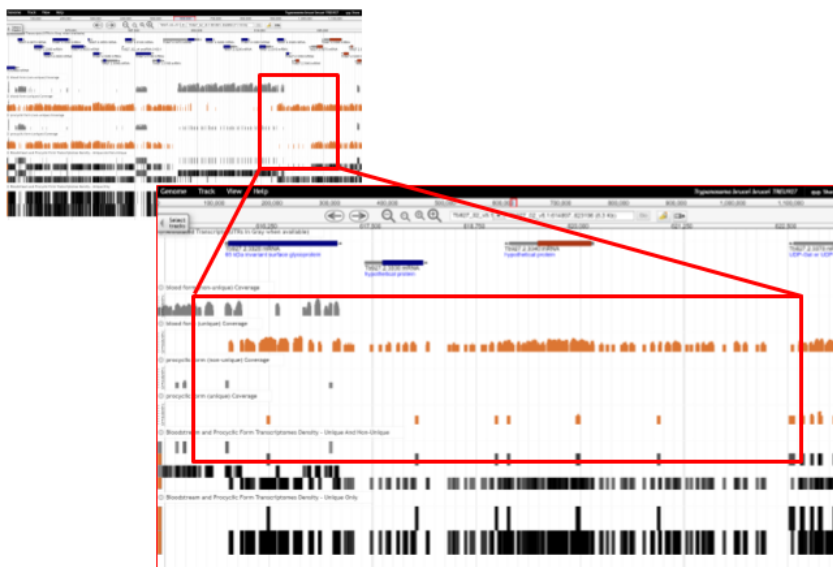
- You may need to adjust the y-axis scaling to bring the tracks into proper view (try setting the score range to "global" by mousing over the track name, clicking the dropdown arrow and selecting "Change Score Range").
- What does this data show you?
- Roughly how many polycistronic units does this chromosome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



- Do the ChIP-seq peaks correlate with the direction of gene transcription (blue vs. red)?
- Now zoom back to around 50Kb. Turn off the ChIP-Seq tracks and turn on the RNASeq Coverage track called: **Bloodstream and Procyclic Form Transcriptomes mRNAseq Coverage aligned to T brucei TREU927**.



- 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 forms and procyclics?



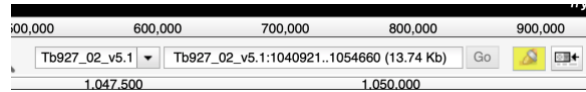
- Zoom in to a gene that looks like it is differentially expressed. What are your conclusions? Are the reads supported by unique or non-unique reads?
- Can you turn on additional tracks that may give some more support to your conclusions?
Hint: turn on the EST and *T. brucei* protein expression evidence 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.



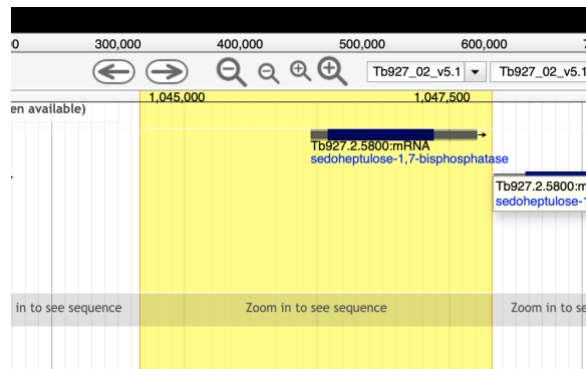
7. Retrieving data from and uploading your own tracks to JBrowse

a. 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”.

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

The image illustrates the process of downloading sequence data from JBrowse. It shows a 'Sequence' track menu with options like 'Pin to top', 'Edit config', 'Delete track', and 'Save track data'. The 'Save track data' option is highlighted with a red arrow. This leads to a 'Save track data' dialog box where the 'Region to save' is set to 'Highlighted region - Tb927_02_v5.1:1044580..1047909 (3.33 Kb)' and the 'Format' is set to 'FASTA'. The 'Filename' field shows 'Reference Sequence-Tb927_02_v5.1-1044580..1047909.fasta'. A red arrow points from the 'Save' button in the dialog to a 'FASTA export' window. This window displays the sequence data in FASTA format, starting with '>Tb927_02_v5.1 Tb927_02_v5.1:1044580..1047909' followed by the sequence text. A 'Close' button is visible at the bottom of the FASTA export window.

b. Uploading data to JBrowse:

JBrowse can accept a number of 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:

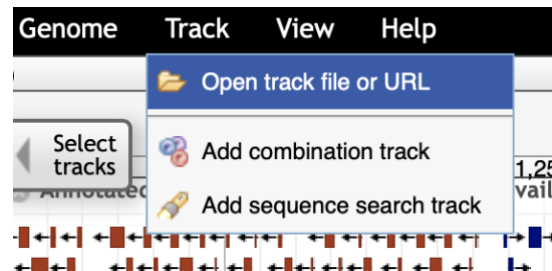
i. Go to this GEO sample record:

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM2407365>

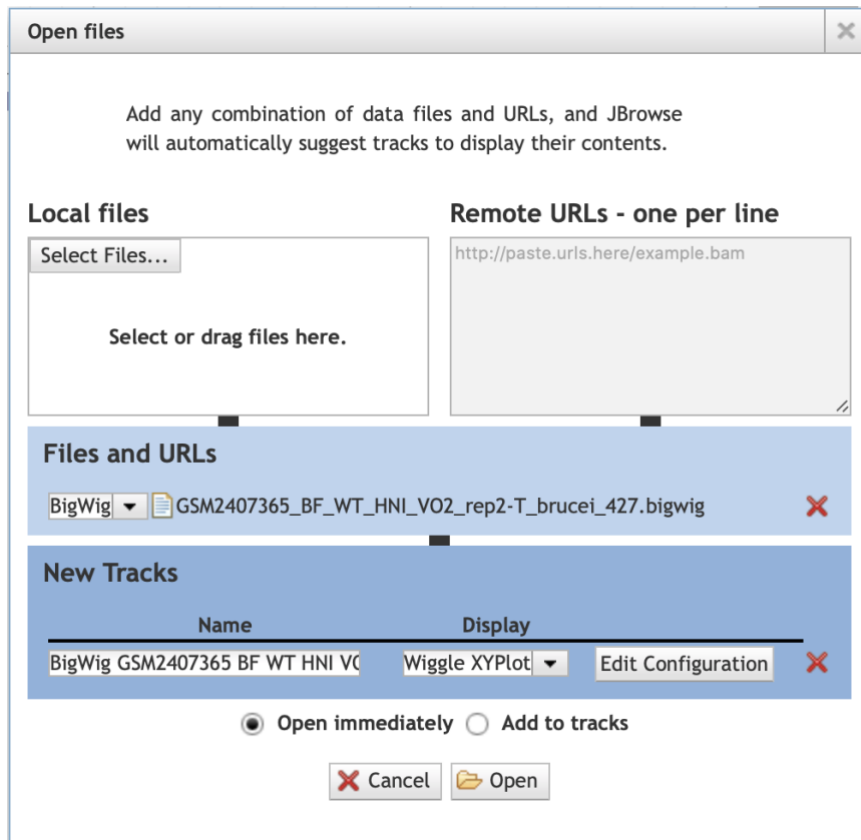
ii. Scroll down to the bottom of the page and download the bigwig file with the http link.

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 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.

