Browser Exercises - I
Alignments and Comparative genomics

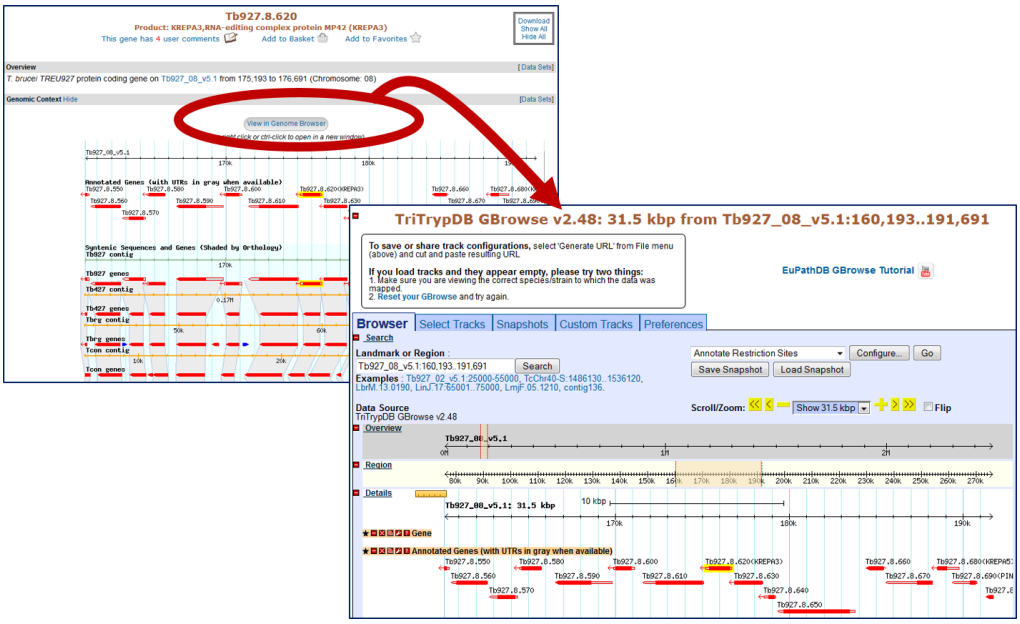
1. Navigating to the Genome Browser (GBrowse)

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

1 a. Navigate to GBrowse from TriTrypDB.

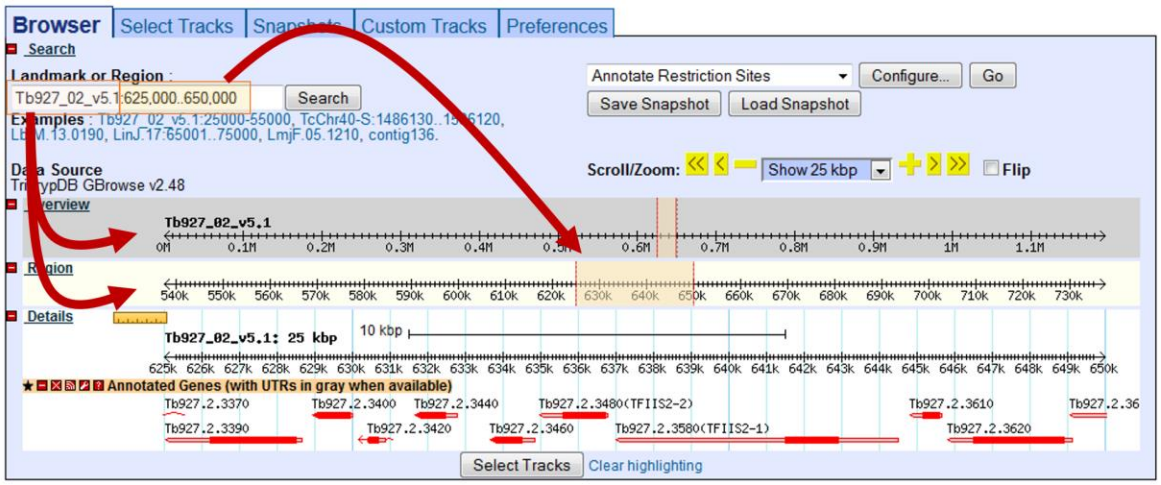
From record pages, like a gene page, genomic sequence or EST page, click on the "View in Genome Browser" link. You can also use the Tools section on the homepage or the grey toolbar in the header section

- **Go to GBrowse from the TriTrypDB home page.** Explore this page – take note of the different sections: Instructions, Search, Overview, Region, Details, Tracks, etc…

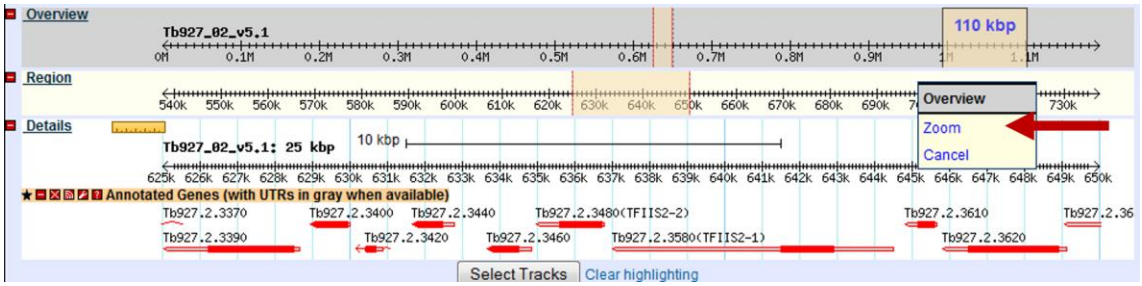


1 b. Explore a genomic sequence and the annotation track in GBrowse.

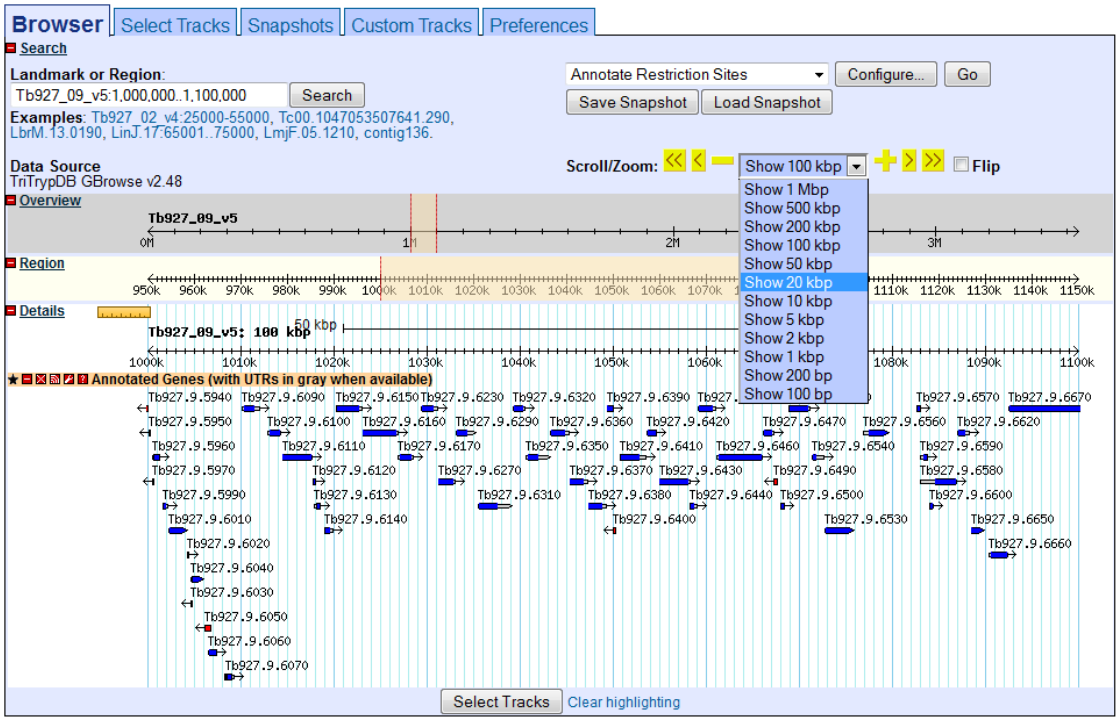
- Look at the "Landmark or Region" box. What information does the "Landmark or Region" box contain? The Landmark or Region box should read – Tb927_02_v5.1:625,000..650,000.
- What chromosome is displayed?
- What location of the chromosome is displayed?



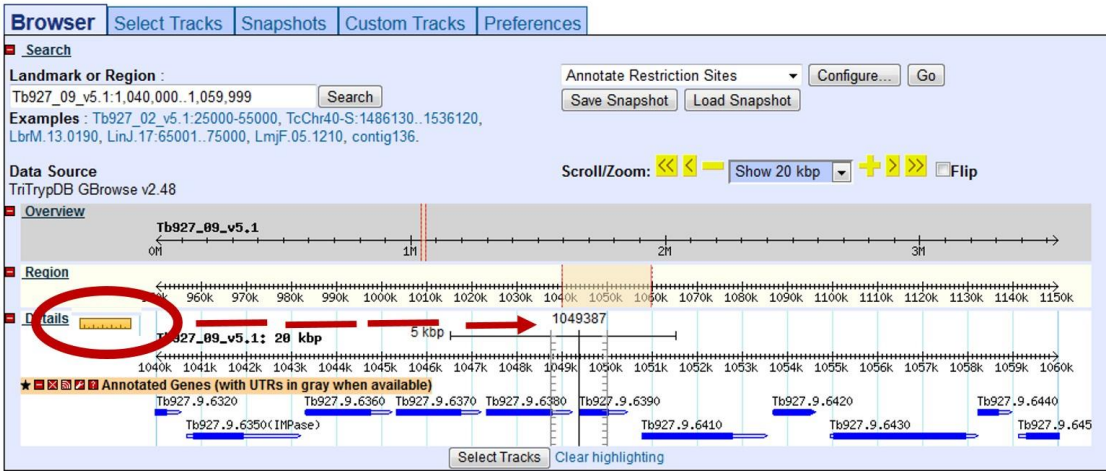
- Move to a different genomic region on this chromosome. For example, visit the right arm of this chromosome.
  - Hint: change the coordinate numbers in the "landmark or region" box to correspond to an area in that region. Look at the overview to give you an indication of the total size of this chromosome, ie. 1000000..1100000).
  - **OR** highlight the area representing approx. 1000000-1100000 on the scale in the Overview section and then choose zoom from the popup.



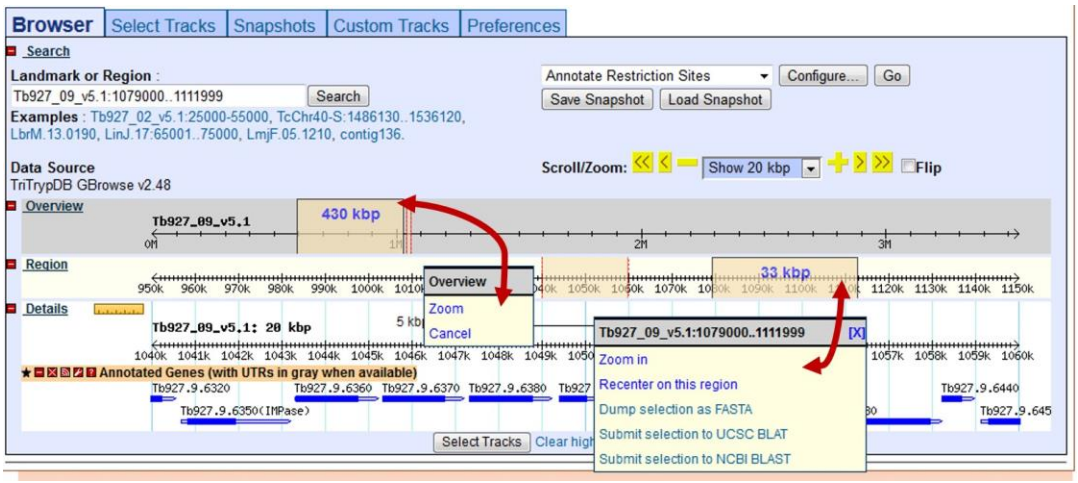
- Move to chromosome 9. How did you do this?
  - Hint: Change the chromosome number in the "landmark or region" box. It should look like this: Tb927_**09**_v5.1:1,000,000..1,100,000. The Tb927_09_v5.1 portion is the genomic sequence ID for chromosome 9.
- Zoom in to a 20Kb region. Select 20Kb from the Scroll/zoom drop down menu.



- What genes are in this region? Mouse over the gene graphics and look at the popups.
- Explore the ruler tool. Click on the ruler to engage then drag it across the window. The ruler tool displays the nucleotide coordinates of the ruler's solid center line. This is very useful for comparing between the annotation data track and others that we will add later.



- There are other ways to move and zoom. Try highlighting an area along the scale in the overview, region or details sections of GBrowse.

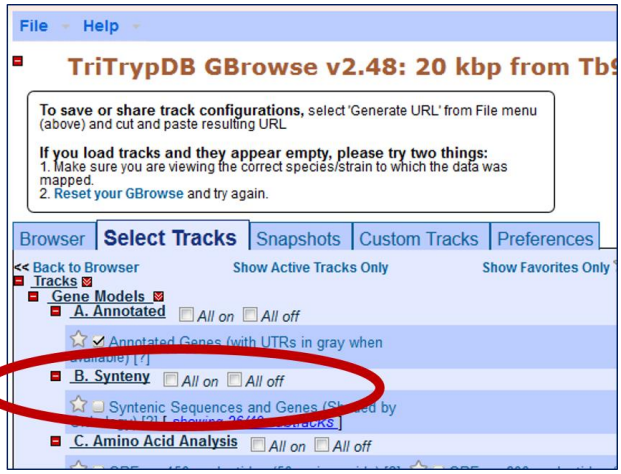


- What if you want to go to a specific gene in Gbrowse? Try to figure out how to go to this gene: Tb927.2.5800
- Type the ID in the "landmark or region" box. The landmark box has a search function that supports gene IDs. What else does it support?
- What is this gene?

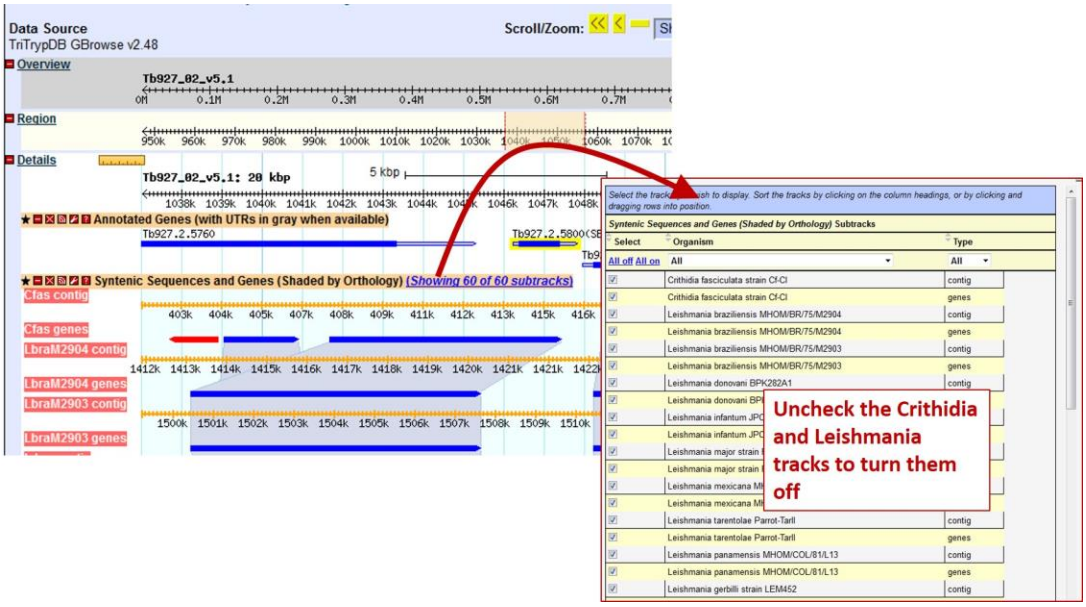
2. Exploring data tracks in GBrowse

2 a. Is the region containing the sedoheptulose-1,7-bisphosphatase (SBPase) gene syntenic in all kinetoplastids?

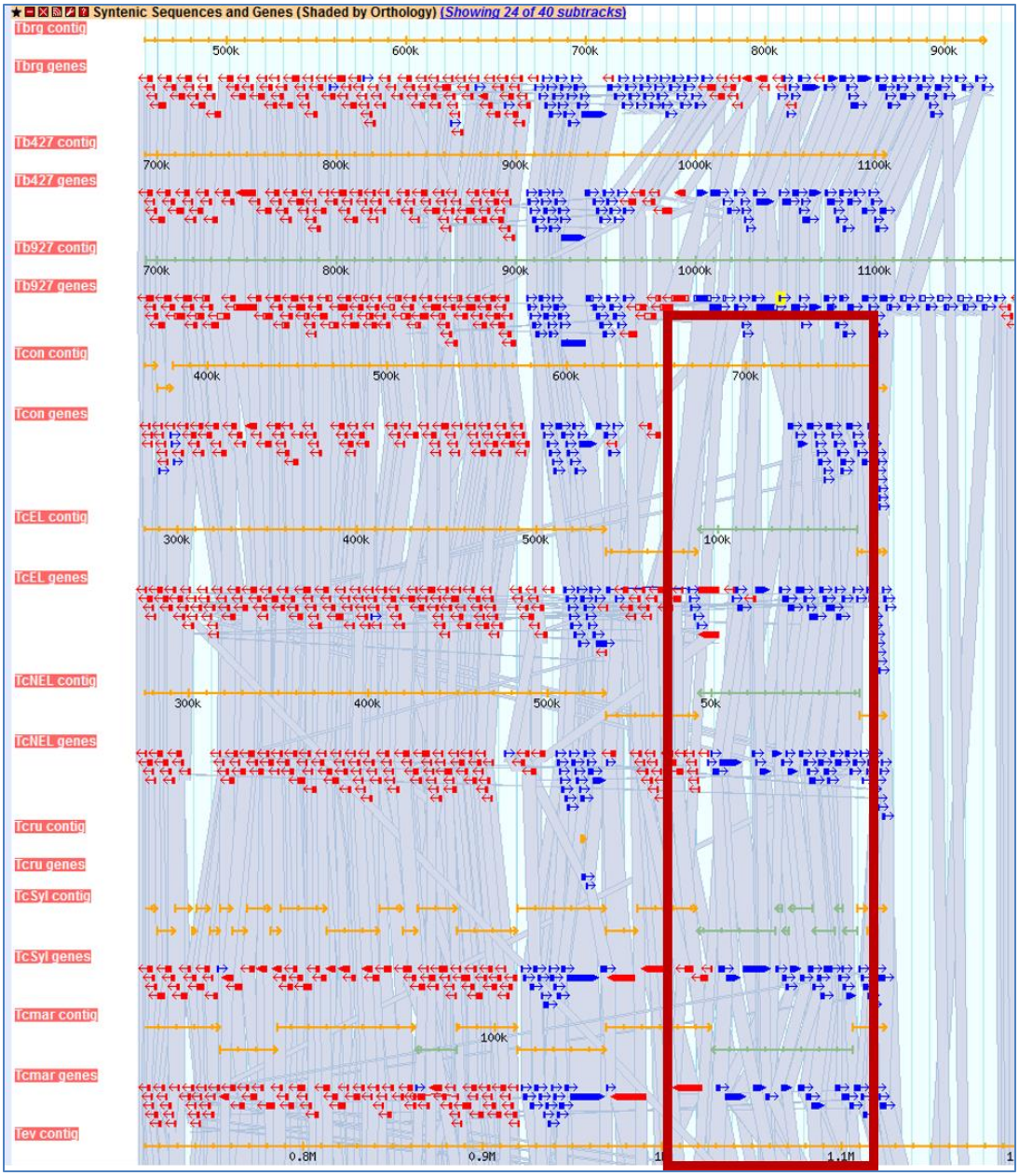
- Go to the "Select Tracks" section and turn on the track called "Syntenic Sequences and Genes". The browser is automatically updated with tracks you select. Note that this track contains multiple subtracks.



- Return to the browser by clicking the "Browser" tab and **zoom out to 20Kb**.
- What does this region look like?
- What direction is the SBPase gene relative to the chromosome?
- What genes are upstream and downstream of the SBPase?
- Modify the subtracks to remove *Leishmania* and *Crithidia* species from the view. Click on the link 'showing 60 of 60 subtracks', wait for the popup and uncheck all *Leishmania* and *Crithidia* species. **Then click Change**.



- Examine the gene corresponding to the *T. vivax* SBPase in the synteny track. Hover the image to find the gene name in the popup. Why is it a fragment? What could be some possible reasons for this?
- Zoom out to 50KB. Look at the genomic sequence for *T. congolense* – why does the synteny look like it does?
- Zoom out to 500KB – what could you conclude about this region in *T. congolense*? (See image on next page if needed).
- You will also notice that some of the genomes have contigs that are not contiguous. Why is that?
- Mouse over the two contigs and look at the information in the popups – do these pieces belong to the same chromosome? What does this mean?



2 b. Exploring other data tracks in Gbrowse.

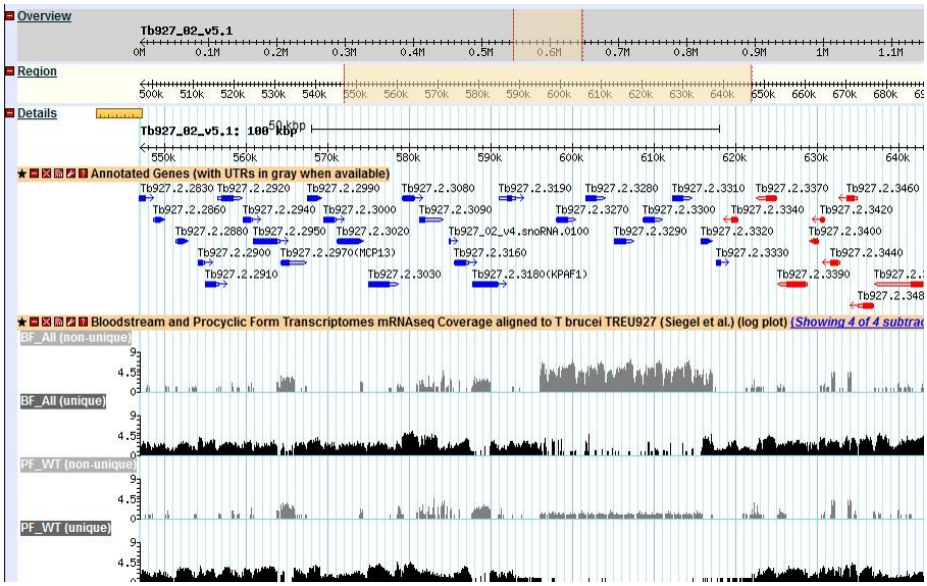
In this example we are viewing *T. brucei* SBPase, so the data tracks you turn on will display data only if the data is aligned to the *T. brucei* genome.

Turn on the ChIP-seq coverage plots and turn off the syntenic gene and region tracks. The data track is called: **ChIP-Seq - Four histone Variants ChIP-Seq Coverage aligned to T brucei TREU927 (Cross) (log 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

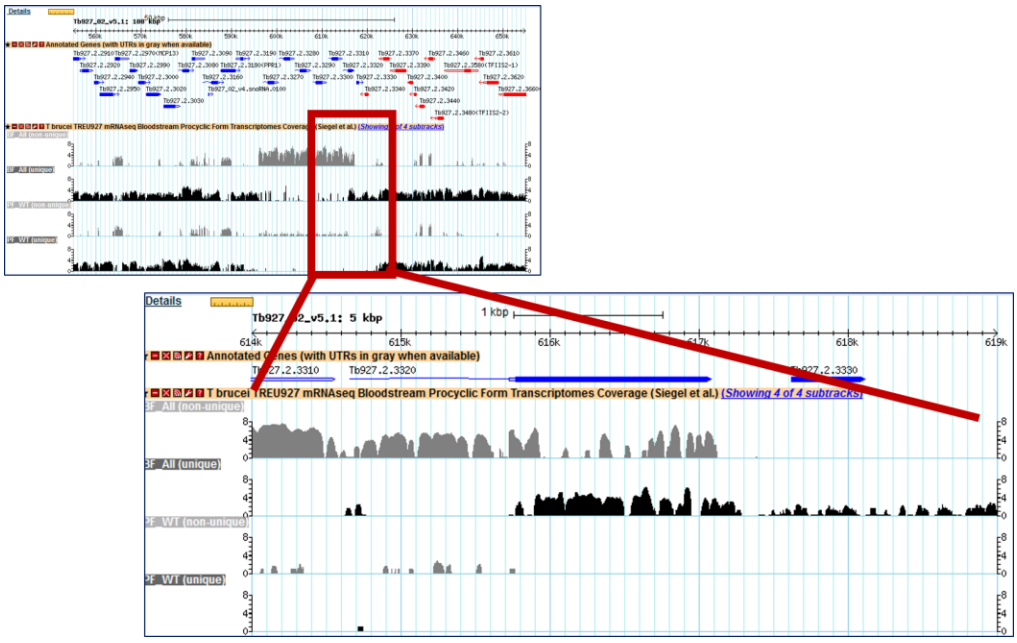
genome. Peaks in the sequence coverage plots represent areas of histone binding and transcription start sites. http://www.ncbi.nlm.nih.gov/pubmed/19369410

- 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 50Kb. Turn off the ChIP-Seq tracks and turn on the track called: **Bloodstream and Procyclic Form Transcriptomes mRNAseq Coverage aligned to T brucei TREU927 (Siegel et al.) (log plot)**(4 subtracks).
- 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 black and grey bars in the coverage plots. What do you think the grey bars indicate?

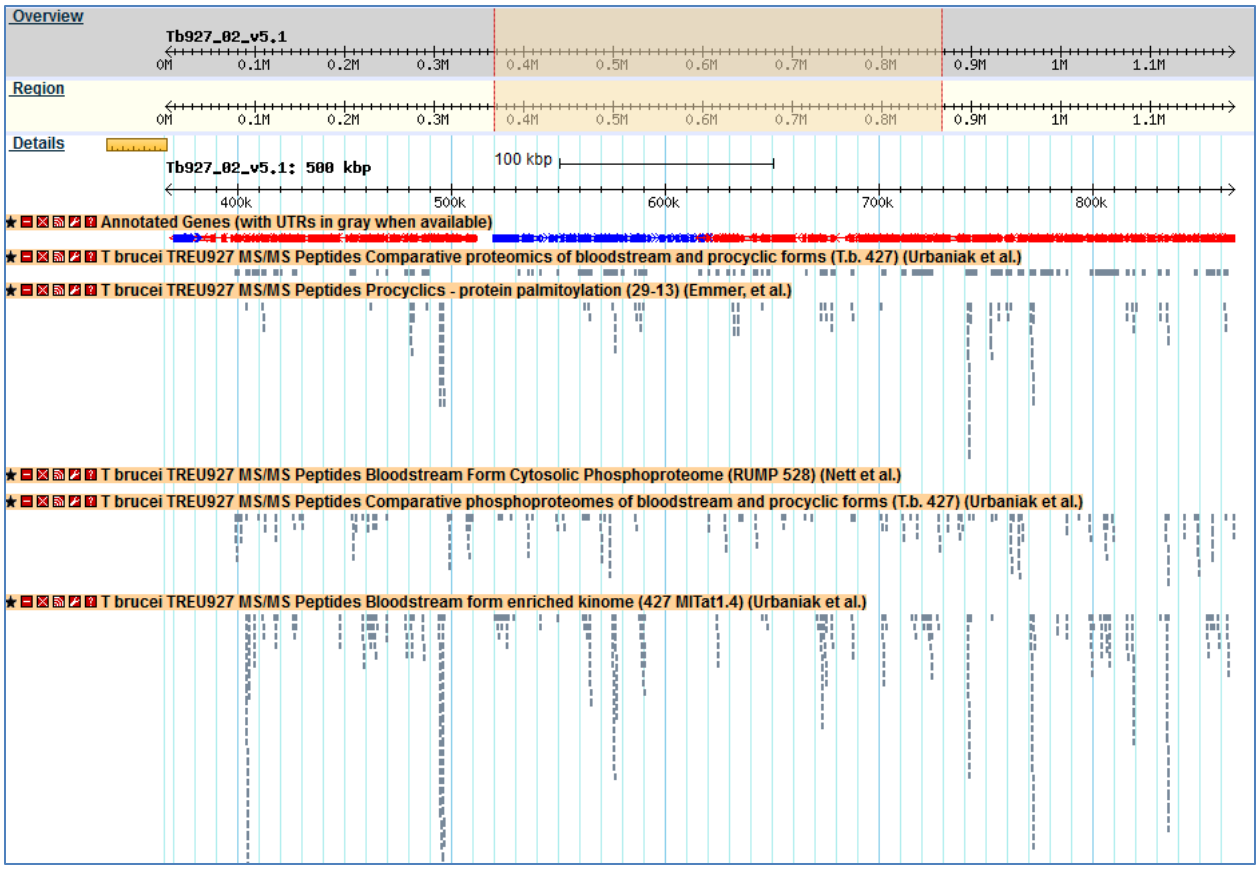


- Now zoom out to 100Kb – do you see a difference between the blood forms and procyclics?
- Zoom in to a gene that looks like it is differentially expressed. What are your conclusions? Are the reads supported by unique or non-unique reads?



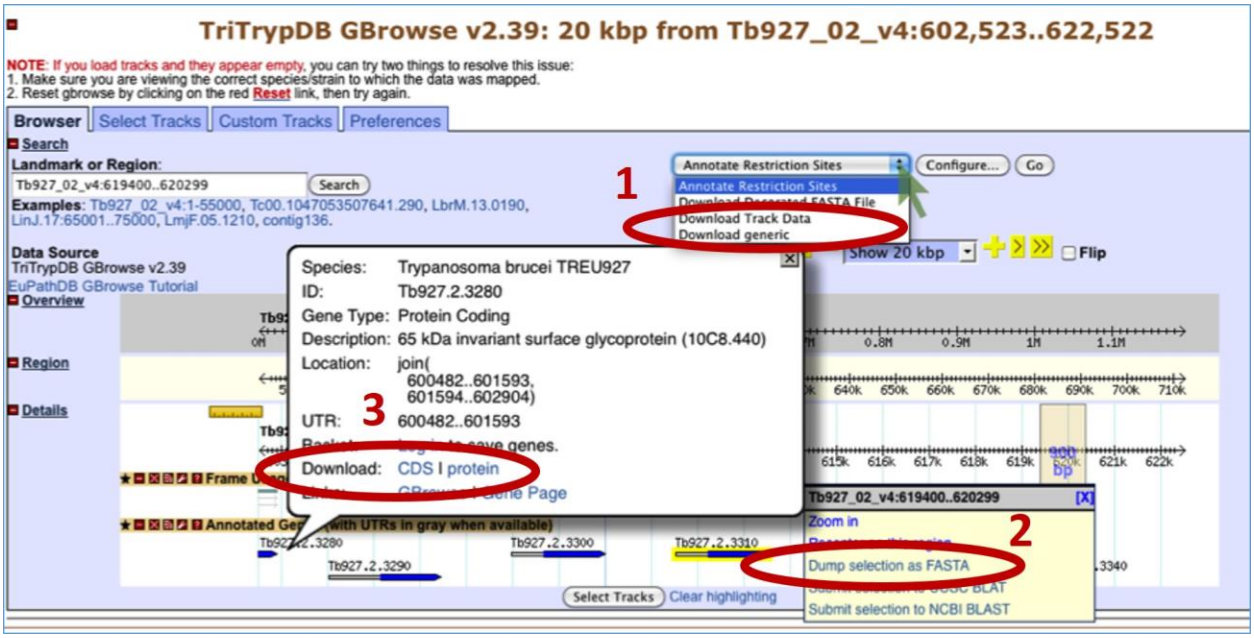
- Can you turn on additional tracks that may give some more support to your conclusions?
  Hint: turn on the EST and *T. brucei* protein expression evidence tracks.
  - Is there any proteomics evidence for this region?
  - How about EST evidence? Click on an EST graphic (glyph) to get additional information.

- 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?



3. Downloading data from GBrowse and uploading your own tracks to GBrowse

You can download data from GBrowse in multiple ways and formats.

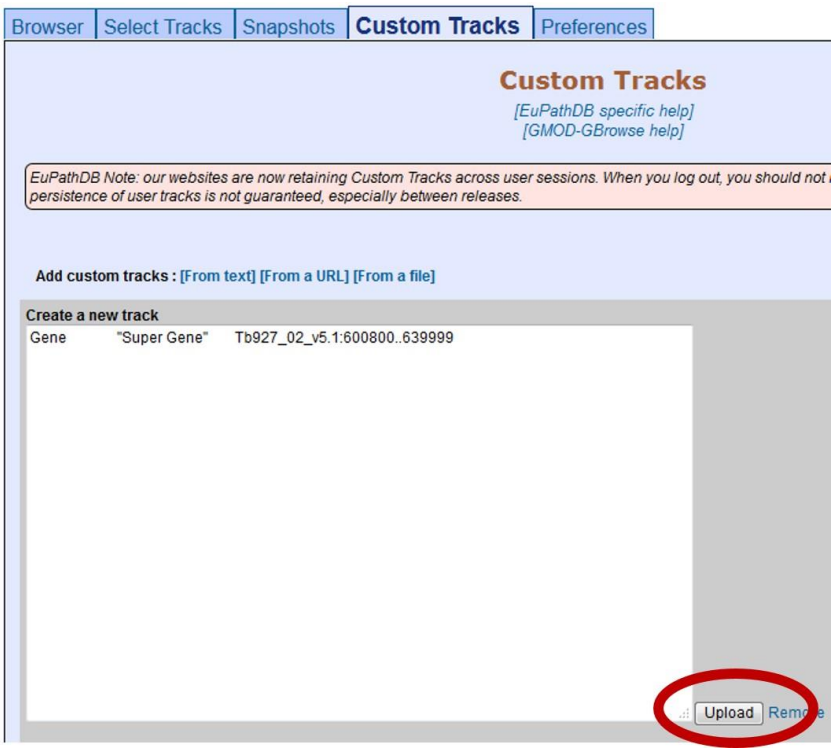


1. The Report and Analysis drop down menu allows you to select a format for the download file that will contain the all the features that you have displayed in the region you are viewing.
2. Highlighting a section of the Details scale allows you to retrieve a FASTA dump of the nucleotide sequence from this region. You can also use this same tool to submit a sequence to NCBI Blast.
3. Mousing over a gene will reveal a popup window with the option to get the coding (CDS) or amino acid sequence of that gene.

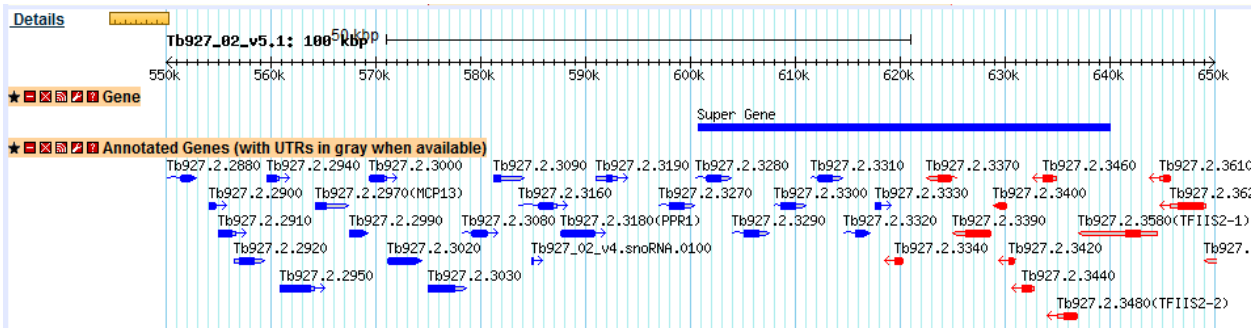
- **Uploading your own tracks is also possible.** One reason to upload your own tracks is to display your own data on a chromosome or genomic segment and view it in the context of gene models and other data. To do this you have to follow some rules to ensure that the file you are uploading can be understood by GBrowse. In this exercise we will only go through a couple of simple examples to give you an idea of the possibilities. There are many online resources if you wish.

- Imagine that you have cloned a new gene and you would like to display it in GBrowse. Click on the "Custom Tracks" GBrowse tab and add a custom track "From text".

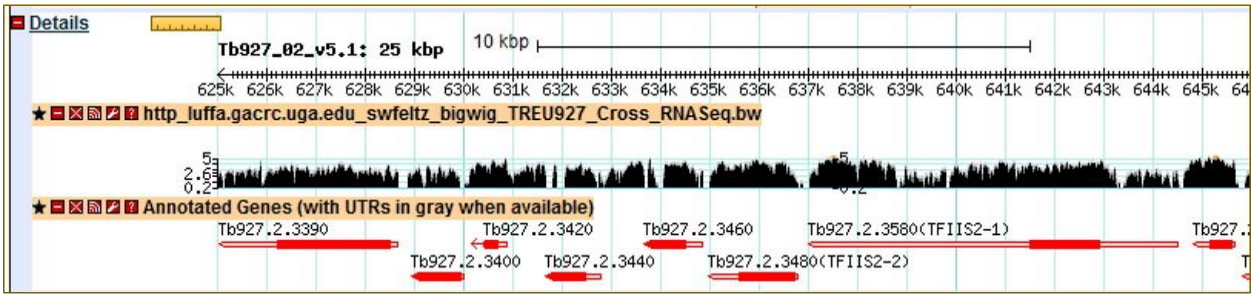
There are many types of formats that can be used. For this example we are going to tell GBrowse that we have a gene and a few things about the gene, like its location. Paste the following into the editor (the next window after you click on Add Custom Tracks From Text), and then click on upload (hint: sometimes you have to zoom in or out a little to see your new glyph): **Gene "Super Gene" Tb927_02_v5.1:600800..639999**



You should see a new track with your gene displayed.



- Now let us load a more complex graphic, a bigwig file of some RNA Sequencing data. For this we posted the file to a public site and are using the URL to direct GBrowse to the file location. In the field "Fetch track file from this URL", enter the following and click Import:
  http://luffa.gacrc.uga.edu/swfeltz/bigwig/TREU927_Cross_RNASeq.bw



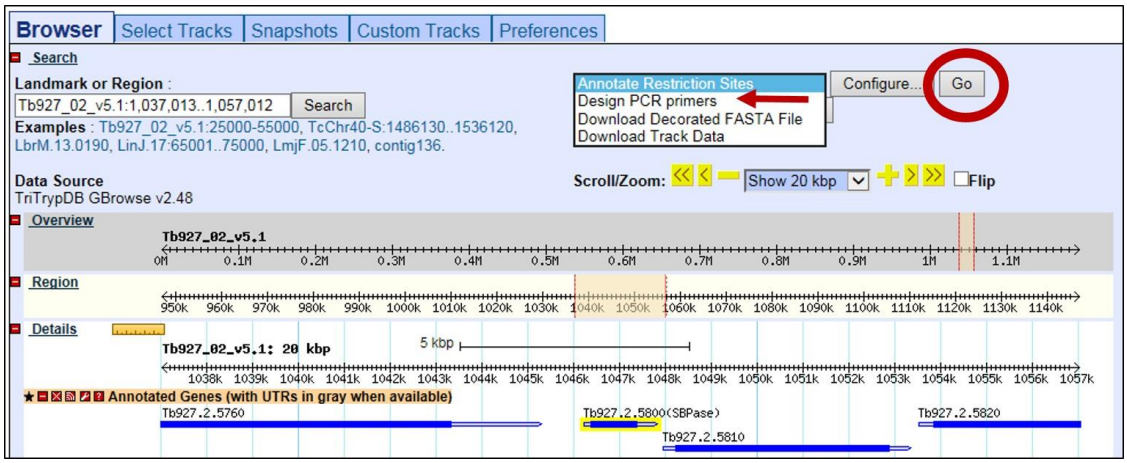
4. Designing PCR Primers with GBrowse

Open GBrowse at the genomic location for your new primers.

- Go to gene page of a gene you want to design primers for and use the 'View in GBrowse' button. This example is using SBPase, Tb927.2.5800 and we have zoomed out to 20K.
- OR open GBrowse from the home page and then enter genomic coordinates of your region in the landmark region.

Choose "Design PCR Primers" from the drop down menu and then **click GO**.

- This opens the Design Primers application.



Choose a target:

- The graphic is interactive. To choose a target, highlight an area on the scale. You can zoom in with the controls in the upper left corner. The PCR primers that you design with this application will flank the shaded region.
- Once you choose a target, the Product size range is automatically updated in the parameter table at the bottom of the page.
- You can choose to customize the primer design using other parameters.

Click DESIGN PRIMERS to run the application.