Browser Exercises - I

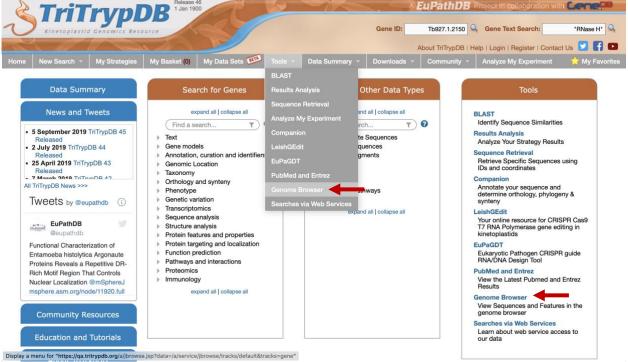
Alignments and Comparative genomics

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:

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

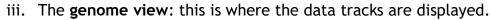


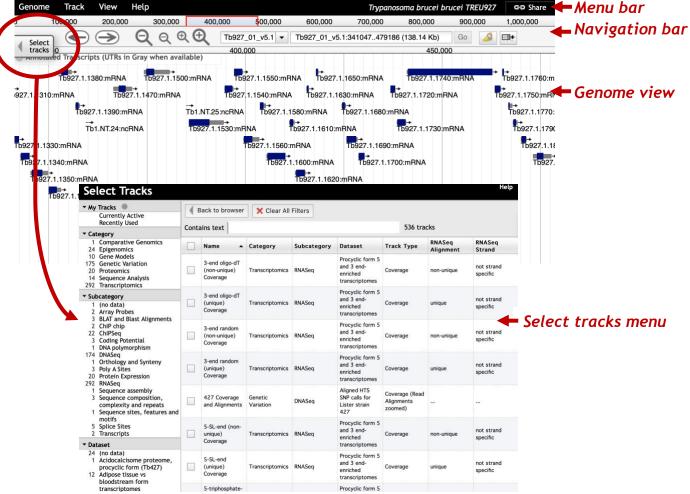
c. From record pages such as gene, SNP or genomic sequence pages - these links are usually to specific JBrowse configuration that include data relevant to the section on that record page. For example, a JBrowse link from an RNAseq dataset on the gene page would display the gene of interest along with 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.

- 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.



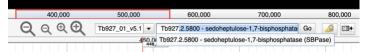


c. Selecting tracks: click on the "select track" button (top left). You can search/filter for tracks and then select them for display by checking the check box next to the track name.

3. Navigating to a specific gene in JBrowse.

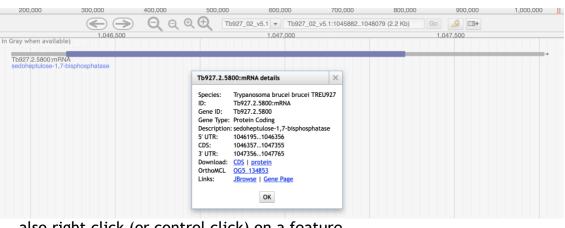
The goal of this step is to navigate to the sedoheptulose-1,7bisphosphatase (SBPase) gene of *T. brucei* 927.

- 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 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.

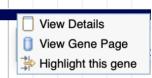




c. You can get information about any feature in the genome view window by clicking on it. Click on the gene feature. What information is available in the popup? You can



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).



d. 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

Track View Help 100,000 300,000 400,000 500,000 900,0 200 000 600.000 Click and drag $\Theta \ominus \oplus \Theta$ Tb927_02_v5.1 - Tb927_02_ <u>s</u> E 1,040,000 1,045,000 1,048,100 1,035,000 Transcripts (UTRs in Gray when available) region Tb927.2.5800:mRNA Tb! nRNA Tdp1), putative hy Tb927.2.5760:mRNA Flagellar Member 8 19927.2.5810:mRNA Iolliday-junction resolvase-like of SPT6/SH2 domain cc

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).

Currently Active			All Filters Cont	ains text splice		X 7 matching	X 7 matching tracks				
Recently Used		Name 🔺	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand			
7 Gene Models	_	Bloodstream and Procyclic for spliced leader transcriptomes	Gene Models	Splice Sites		Segments					
2 Poly A Sites		(927, 427)(2014) - Splice Sites	Gene Models	Splice Sites		Segments					
5 Splice Sites			Bloodstream and procyclic								
Dataset 7 (no data)		form spliced leader transcriptomes (427, Antat) (2010) - Splice Sites	Gene Models	Splice Sites		Segments					
 Track Type 			Curated Poly A Sites from								
7 Segments		bloodstream and procyclic	Gene Models	Poly A Sites		Segments					
RNASeq Alignment		forms									
7 (no data)		Procyclic form spliced leader	Gene Models	Poly A Sites		Segments					
RNASeq Strand		transcriptome - Poly A Sites	Gene Modela	roly A sites		orginanca					
7 (no data)		Procyclic form spliced leader transcriptome - Splice Sites	Gene Models	Splice Sites		Segments		-			
		Spliced Leader and Poly A Sites from bloodstream and procyclic forms - Splice Sites	Gene Models	Splice Sites		Segments					
		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?

00	500,000	550,000 60	0,000 650,00	0 700,000	750,000	800,000	850,000	900,000	950,000	1,000,000	1,0 <mark>50</mark> ,000	1,100
Q	€€	Tb927_02_v5.1 -	Tb927_02_v5	1:10449241048	082 (3.16 Kb)	Go 🧹	S					
		1,046,	500		1,047,	000			1,047	,500		
	Tb927.2.58											→
	sedoheptul	ose-1,7-bisphosphata	ase									
			ТЬ927.2.5	300(11) details				×	1			
	Theorem	00(1)									Tb927.2	2.5810(
	Tb927.2.58	00(11)	Location: Gene ID:	1046195 ТЪ927.2.5800							Tb927.	2 5810
	•	. ,	UTR Length	162							10027.	
	Tb927.2.	5800(5)	Count: Note:	128.2599999999	19996 nt is the sum of t	he count p	or million for	aach cample				Tbs
	Tb927.2	.5800(1)	Note:	The overall cou	Sample	ne count p	Count per					
		Tb927.2.5800(12)		T.brucei 427 cB			7.32	maion				
		•			nd-enriched cDNA	λ.	62.91					
		Tb927.2.5800(1)		T.brucei 927 PC T.brucei 927 sl			6.72 8.15					
				T.brucei Alba 1			13.9					
				T.brucei Alba 1			4.85					
					_4+ (29-13 RNAi) tream (Lister 427	n n	2.02 4.66					
					tream long slende							
					tream short stum							
				I.brucei curate	d splice site cDNA	A	7.39					
					OK							

- 5. Exploring synteny between genomes. Synteny 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									He
✓ My Tracks Currently Active	4	Back to browser	×	Clear All Filters	Contains text			1 matching tra	ack
Recently Used	1	Name	•	Category	Subcategory	Dataset	Track Type	RNASeq Alignment	RNASeq Strand
Category X Comparative Genomics Z4 Epigenomics	<	Syntenic Sequences a Genes (Shaded by Orthology)	and	Comparative Genomics	Orthology and Synteny		Segments		
10 Gene Models 175 Genetic Variation 20 Proteomics 14 Sequence Analysis 292 Transcriptomics									

- 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

😒 Syntenic Sequences and Genes (Shaded by Orthology)

• Select the option called "Select Subtracks" from the menu

Syntenic Sequences and Genes (Shaded by Orthology)	? About this track
bruTREU927 gene	Pin to top
bruTREU927 span	🛞 Edit config
bruLister427 gene	X Delete track
bruLister427 span	Save track data
bruLister427_2018 gene	
bruLister427_2018 span	Display mode
brugambienseDAL972 gene	Show labels
brugambienseDAL972 span	Select Subtracks
conIL3000 gene	
conIL3000 span	
evaSTIB805 gene	
evaSTIB805 span	
graANR4 gene	

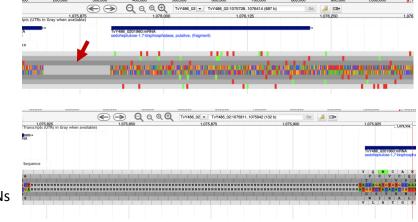
 In the next popup first uncheck all organisms, second use the filters on the left to select Trypanosoma, third select the species of interest (note that you should select both the gene and span subtracks for each species), fourth click on the save button at the bottom of the popup.

Select Subtracks									
▼ My Tracks	×	Clear All Filters	Contains text				12	matching tracks	
Currently Selected	1	ld .	Species	Kingdom	Genus	Phylum	Class	syntype	taxon
			Trypanosoma						Trypanosoma
12 N/A	~	1	brucei	N/A	Trypanosoma	N/A	N/A	gene	brucei brucei TREU927
▼ Genus	•	2	Trypanosoma brucei	N/A	Trypanosoma	N/A	N/A	span	Trypanosoma brucei brucei TREU927
12 Trypanosoma	7	3	Trypanosoma	N/A	Trypanosoma	N/A	N/A	gene	Trypanosoma brucei Lister
✓ Kingdom 12 N/A			brucei	ike.	in ypanosoma	10.0	ie A	Nevile	strain 427
▼ Phylum	~	4	Trypanosoma brucei	N/A	Trypanosoma	N/A	N/A	span	Trypanosoma brucei Lister strain 427
12 N/A	~	5	Trypanosoma brucei	N/A	Trypanosoma	N/A	N/A	gene	Trypanosoma brucei Lister strain 427 201
Species Trypanosoma brucei Trypanosoma congolense		6	Trypanosoma brucei	N/A	Trypanosoma	N/A	N/A	span	Trypanosoma brucei Lister strain 427 201
20 Trypanosoma cruzi 2 Trypanosoma evansi 2 Trypanosoma grayi 2 Trypanosoma rangeli	•	7	Trypanosoma brucei	N/A	Trypanosoma	N/A	N/A	gene	Trypanosoma brucei gambiens DAL972
2 Trypanosoma theileri 2 Trypanosoma vivax	~	8	Trypanosoma brucei	N/A	Trypanosoma	N/A	N/A	span	Trypanosoma brucei gambien DAL972
* syntype 6 gene 6 span	•	9	Trypanosoma congolense	N/A	Trypanosoma	N/A	N/A	gene	Trypanosoma congolense IL3000
▼ taxon	•	10	Trypanosoma congolense	N/A	Trypanosoma	N/A	N/A	span	Trypanosoma congolense IL3000
 Trypanosoma brucei Lister strain 427 Trypanosoma brucei Lister strain 427 2018 	~	39	Trypanosoma vivax	N/A	Trypanosoma	N/A	N/A	gene	Trypanosoma vivax Y486
2 Trypanosoma brucei brucei TREU927 2 Trypanosoma brucei	•	40	Trypanosoma vivax	N/A	Trypanosoma	N/A	N/A	span	Trypanosoma vivax Y486

- 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?

Genome Track	View Help 200.000	300,000		400,000	500,000	600,0	00 700.000	800,000	900,000		1.000.000	1,100,	GD Share
100,000			0					64611059480 (23.02 Ki				1,100,	,00
Select	\leftarrow		्र	~ ~ ~ ~		_02_v5.1 👻			D) GO				
tracks	1.040.000	a available)		1.045.00	0		1,050,000			1.055.0	00		
Minotated franscripts	s (OTRS in Gray whe	ni avaliable)											
					The	27.2.5800:mRN	۵		Tb927	2.5820;r	nRNA	•	Tb9
					sed	oheptulose-1,7-	bisphosphatase		hypoth	etical pro	otein, conserved	d	sma
b927.2.5760:mRNA			•			7.	27.2.5810:mRNA		•				
lagellar Member 8						Ho	liday-junction resolva	se-like of SPT6/SH2 dor	main contai	ning pro	tein, putative		
													•
												Tbs	927.2.5830:n pothetical pro
												1.00	ourouour pro
Syntenic Sequences a	ind Genes (Shaded I	by Orthology)										
bruTREU927 gene			•		_	•				_		•	
bruTREU927 span			-						7				
bruLister427 gene			•		_								
pruLister427 span			/			-/-/						_	
oruLister427_2018 gene			•					+	_				
bruLister427_2018 span												_	
brugambienseDAL972 ge			•	+				•	_				• •
brugambienseDAL972 sp					/				/			/	
conIL3000 gene	_					•		+				• •	·
conIL3000 span													
conIL3000_2019 gene	-			••		÷ +		· ·		_		•• ==	
conIL3000_2019 span													_
evaSTIB805 gene			-		_			•	-	_			
evaSTIB805 span					/ /			/				/ /	
graANR4 gene		+						•	+	•		• —	
graANR4 span					/						/		
ranSC58 gene						•	>		•		•		•
ranSC58 span						-							
heEdinburgh gene	+	-			· ·			+	_				
heEdinburgh span												P	
vivY486 gene								•					
rivY486 span													
asCfCl gene		•	-		UND.		+		-		+	_	•
asCfCl span													
monLV88 gene			•					_					
monLV88 span			1										
najFriedlin gene							•		+				-
najFriedlin span													
conCUL13 gene			•					1					
conCUL13 span													

- 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:

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

✓ My Tracks Currently Active	1	Back to browser	🗙 Clear A	All Filters	Contains text	Four histone variants	×	matching tracks		
Recently Used	1	Name	▲ Category		Subcategory		Dataset	Track Type	RNA Seg Alignment	RNA Seg Strand
Category Section 2 Section 2	•	BDF3-HA (unique)		Epigenomics		ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
Subcategory 9 ChIPSeq		ChIP-Seq - Four his Density - Unique Ar Unique		Epigenomics		ChIPSeq	ChIP-Seq - Four histone Variants	Multi-Density		
Dataset	1	H2Az (unique) Cov	erage	Epigenomics		ChIPSeq	ChIP-Seg - Four histone Variants	Coverage		
9 ChIP-Seq - Four histone Variants	~	H2Bv (unique) Cov	erage	Epigenomics		ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
Track Type	1	H3K4me3 (unique)	Coverage	Epigenomics		ChiPSeq	ChIP-Seq - Four histone Variants	Coverage		
8 Coverage 1 Multi-Density		H3v (unique) Cover		Epigenomics		ChIPSeq	ChIP-Seq - Four histone Variants			
RNASeq Alignment	1	H4 (unique) Covera	3ge	Epigenomics		ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
9 (no data)	1	H4K10ac (unique)	Coverage	Epigenomics		ChIPSeq	ChIP-Seq - Four histone Variants	Coverage		
RNASeq Strand	1	H4v (unique) Cover	rage	Epigenomics		ChiPSeq	ChIP-Seq - Four histone Variants	Coverage		
9 (no data)										

- 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.

	Help									Trypanosoma b			co Sha
100,000	200,000	300,000	400,000	500,000			700,000	800,000	900,000	1,000	,000	1,100,000	
Select			QQ		Tb927_02_v5.1 💌								
tracks 125,000	250,000 Grav when available)	375.	.000	500,000	6	25,000	750.0	00	875,000	1,000.0	000	1,125,00	10
						tastat atata							.
	= = = = = = = =	· 네 네스 (a. 네 · 네 네네 · 네 네네			→ → → → → → ■→ → → → → → ↓→ ■→	+ ++ + + +- ++ + + + + ++ + + +		निनान सम्बन्ध नगरान सम्बन नगरान सम्बन नगरान सम्बन नगरान नगरान नगरान	┫╾╡╺┤╾┨ ╾(╾ ╾┨┝╴ ╼ ╾ ╾ ╾ ┝╾ ╼		- i 	8+ 8+8+8+8+8+	
		- - -	→ + +						4 1 1 4 1				
BDF3-HA (unique) Coverage													
40,000 30,000 20,000 10,000 0													
3 H2Az (unique) Coverage													
H2Bv (unique) Coverage													
20,000													
10,000													
B H3K4me3 (unique) Coverage		· · · · · · · · · · · · · · · · · · ·									-		
\$2000	a serve a here a	- - -		1	•				-184		*		
H3v (unique) Coverage 400,000 300,000 200,000 100,000													
0													
3 H4 (unique) Coverage 20.000													
0,000													
0													
H4K10ac (unique) Coverage													
10,000													
					-						-		
H4v (unique) Coverage													
20,000													
10,000													
U													

- 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**.

enome Track View	Help										ei brucei TREU	
100,000	200,000	300,000	400,000 Q Q	500,000 () () () () () () () () () () () () () (600,000 27_02_v5.1 ▼ Tb927_	700,000	800,000		0,000	1,000,000	J 1,10	0,000
Select tracks	575,000		587,50	1 - 1 - 1	600,00			612,500			625,000	
Annotated Transcripts (UTRs	in Gray when available)								_			
Tb927.2.2970:mRNA	Tb927.2.3020:mRNA	Tb927.2.3055:mRNA	Tb927.2.	3160:mRNA T	b927.2.3270:mRNA	Tb927.2.3280:1	mRNA Tb927.2.3	300:mRNA	Tb927.2.3	320:mRNA	Tb927.2.3	3390:mRNA
Tb927.2.2990		30:mRNA	Tb927_02_	v4.snoRNA.0100.1		Т	b927.2.3290:mRNA	Tb927.2.3			b927.2.3370:mRI	NA Tb927.2
Tb927	*.2.3000:mRNA	Tb927.2.3080:mR	NA 1	rb927.2.3180:mRNA					Tb9	27.2.3330:mRM		Tb927.2.340
		Tb927.2.30	90:mRNA	Tb927.2.3	190:mRNA					Tb927.2.3340	mRNA	Ť
2.2950:mRNA												
blood form (non-unique) Cov	erage											
and to a				és sléde.	الانتخاب ومريالا	الربيس الأرباقات بحس	sin'i arriv'i	ير التواصلية وي	atal cult			
	_ 1 li _ 1		1				بججارة جب					
blood form (unique) Coverage	2											
adie in is some stability statistic	Advantages, and a feature	en illi dahten.	a ideal to the late.	Mer at the d	and a state of the				dilit.co	different etc.	والمتشرف والأشارات	States and the second
procyclic form (non-unique)								1 1				
procyclic form (non-unique)	coverage											
				hil fahr	- DETERMENT	MILLION ALL	hala bi U h a h	n dan t	0.1			
procyclic form (unique) Cove												
	and the designed	4.4484.5.115										ما معد المعالم
				r i l 🆬 Maniferr					11	111		
Bloodstream and Procyclic Fo	rm Transcriptomes Densi	ity - Unique And Non-	Unique									
								1 I				
Bloodstream and Procyclic Fo	orm Transcriptomes Dens	ity - Unique Only										
												

- 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

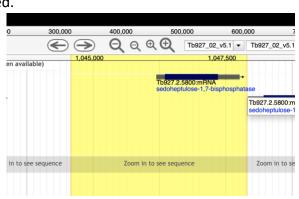
- a. Downloading sequence in FASTA format from a region of interest:
 - i. Make sure the "annotated transcripts" and the "reference sequence" tracks are turned on.

1.047.500

600,000

00,000

- ii. Click on the "highlight a region" button in the navigation bar. It should turn yellow when activated.
- iii. Click and drag in the genome view region and select the area you would like to highlight.
- iv. Click on the down arrow on the reference sequence track and select "Save track data".



700,000

Tb927_02_v5.1 - Tb927_02_v5.1:1040921..1054660 (13.74 Kb) Go 🔗

800,000

1.050.000

900,000

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

quence	? About this track	Save track data						
2	Pin to top	Region to save						
	S Edit config	Highlighted region - Tb927_02_v5.1:10445801047909 (3.33 Kb) Visible region - Tb927_02_v5.1:10409211054660 (13.74 Kb)						
	X Delete track	OWhole reference sequence - Tb927_02_v5.1:11193948 (1.19 Mb, too large)						
	🔛 Save track data	Format Image: Second						
	Show forward strand	Filename						
	Show reverse strand	Reference Sequence-Tb927_02_v5.1-10445801047909.fasta						
	Show translation	X Cancel View Save						
	Show color							
	TAAATGACAGAATAAAGGGAAGGTAATGTGTGAGAGAC AGTGTATGACAATAAGGGAAGGTAATGTGTGTAGAGAC TGTATGACAATATATATATATATATATATATATATATATA	AGAAGCGACCTITCICTITCATITITITITTACCTICIATATT TITGIGACCACTGCGGGTTGCACAGGGTTGACAGGGTGA GITTATCICACAAACTGCGGAAAGAAAGGGTGTCACAGGTGA GGTTTATCICACAAACTGCGGAAAGAAGGTGTTGCG GGGGCGAAGGCCGTTTITITICCICGTCGTGG GGGGTGACGCCGTTTITITICCICGTCGTCG GGGGTGACGCGCGTCATGCGGCTGATGGCGCTG GTTTGTCTCACACTGCGAAGGCGGTGATGCGCGTG GGGGTGACGGCGCGGCGGGGGCGGCGGCGGGCGGGTGCG GGGGTGACGGCGGGCGGGGGGGGGGGGGGGGGGGGGGGG						
	TACGCCGCGAACTGCACCGTGACGGCACCGGCACC	Torgetant account of the construction of the c						

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

Open files	>
Add any combination of data file will automatically suggest tracks	,
Local files	Remote URLs - one per line
Select Files	http://paste.urls.here/example.bam
Select or drag files here.	
Files and URLs	_
BigWig ▼ GSM2407365_BF_WT_HNI_V	02_rep2-T_brucei_427.bigwig
New Tracks	
Name	Display
BigWig GSM2407365 BF WT HNI VC Wig	gle XYPlot 👻 Edit Configuration 🗙
Open immediate	ly 🔵 Add to tracks
🗙 Cancel	🗁 Open

vii. Click on the Open button. The bigwig output should appear very quickly in your browser.

Genome Track View Help						Trypanos	oma brucei Lister strain 4	27 GD Share
500,000	1,000,000	1	,500,000	2,000,000	2,500,000	3,000,000	3,500,000	4,000,000
Select		$\Rightarrow Q$	Q @ 🕀	Tb427_10_v5 👻	Tb427_10_v5:11480011614000 (466 Kb)	Go 🔬 🎫		
tracks	1,250,000 n available)			1,375,000	1,423,394	1,500,000		
પાનન પાનાનાનનન્ત્ર પાનનાનાન બાહ્ય સમયાના વાનનાના ન નેવ તેને તેને સૌથી નાનને ન ા ને ને નેને સૈથી તેને સૈથી ચા		• • • • • • • •	• • • • • • • • • • 		는 말라 말라 다 나 나 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다 다	ગામનાન સમાના તેના તેના તે તે તે તેને તો તે તે તે તે	નેની તેની તેની તેની સંસ્થાની તેની તેની સંસ્થા તેની તેની તેની	ીની નીનીની નેન નેનેનીન∰ની ની નીને નીન નો .
 BigWig GSM2407365 BF WT HNI VO2 rep2 	-T brucei 427	a data mattanta a	and the second secon	1 -	directions in the state of the		a la phile a subscription of the	
				.1			1	