Exploring Gene Models in JBrowse

Learning objectives:

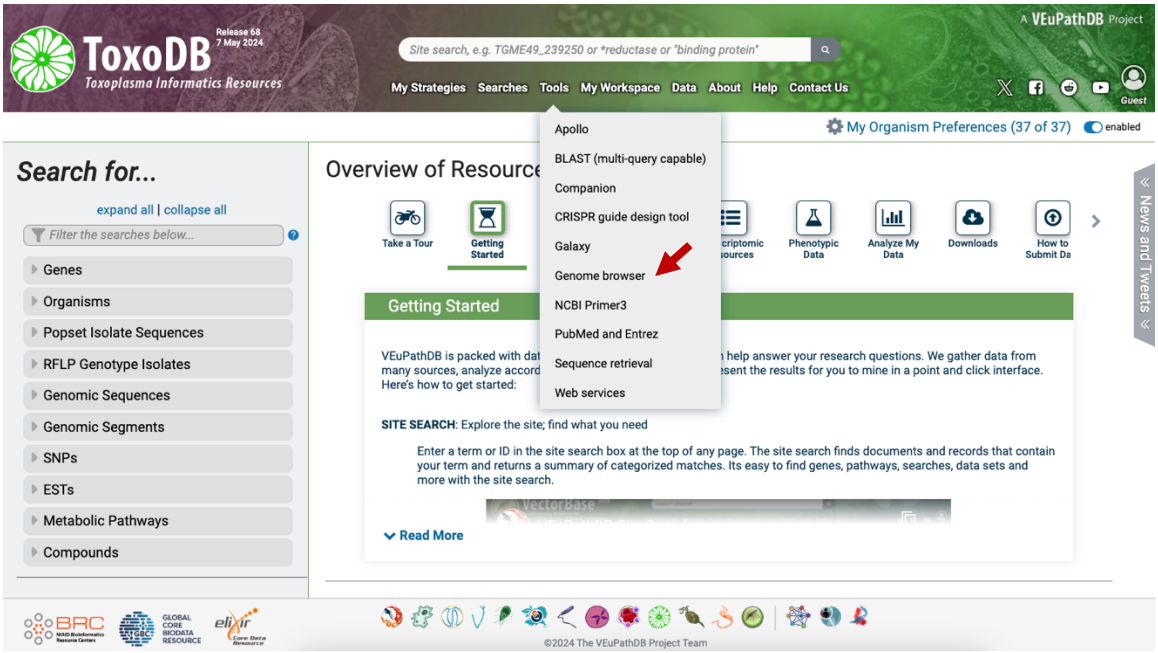
- Navigate to the genome browser
- Use the menu and navigation bars
- Examine gene models in JBrowse
- Assess gene models based on RNAseq data
- Assess gene models based on ChIP-chip and ChIP-Seq data
- Determine if a gene model is accurate or if alternate models are possible
- Explore transcription start site data

In the first part of this exercise, you will learn how to access the Genome browser and navigate to a specific gene. In the second part you will interpret RNA-seq datasets, and use this information to examine gene models.

1. **Navigate 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:

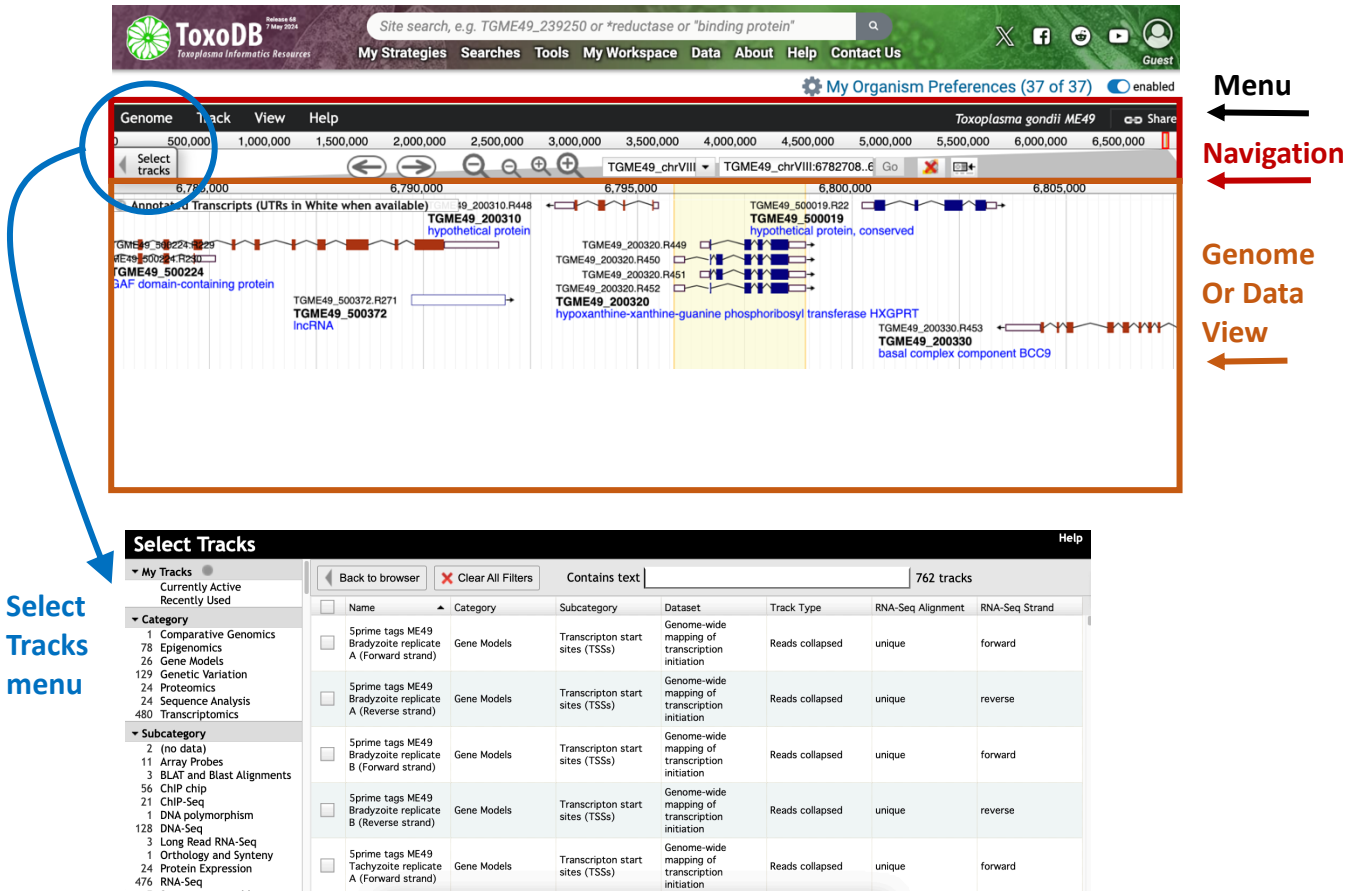
a. **The tools menu** in the banner of any page.



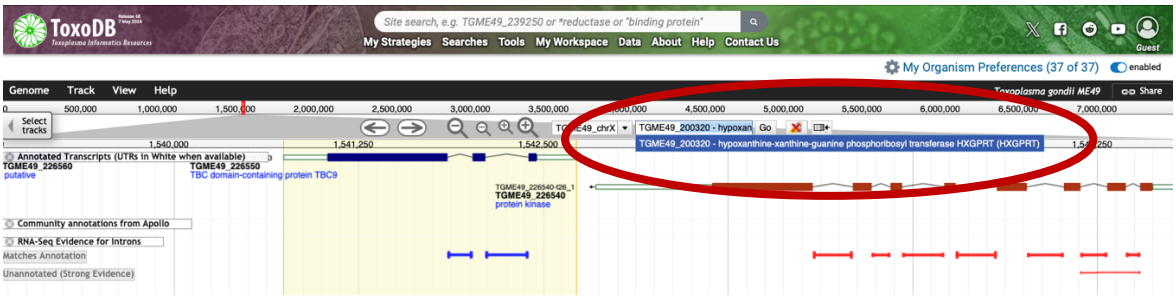
b. **Record pages** such as gene, SNP or genomic sequence pages – these links are usually to a specific JBrowse configuration that includes data relevant to the section on that record page. For example, a JBrowse link from an RNAseq dataset on the gene page would display the gene of interest along with the RNAseq data as density or coverage plots. These links are usually indicated by "View in JBrowse genome browser" button.

2. **Getting around in JBrowse.** Use the tools menu or the View in JBrowse button to open JBrowse. Once in JBrowse examine the following features:

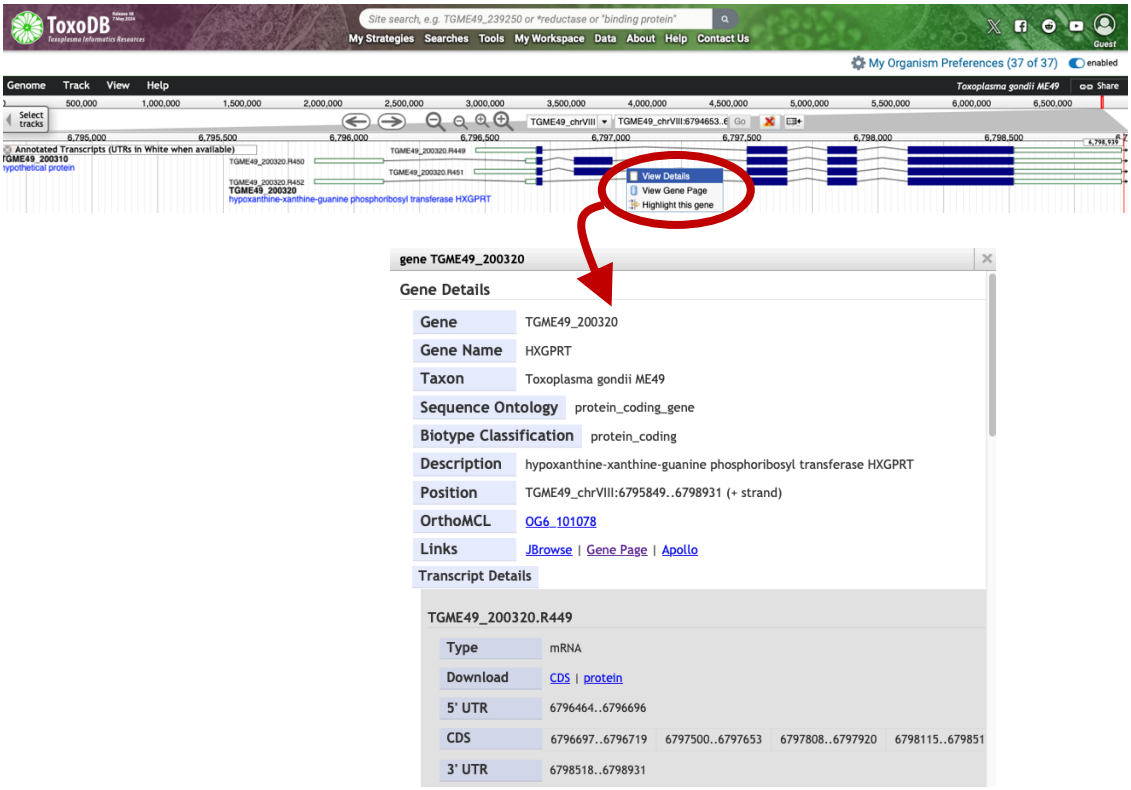
   a. The **menu bar**: located at the top of the JBrowse frame. This includes the Genome menu, Track menu, View menu, Help menu and the Sharing link. What do each of these do/provide?
   b. The **navigation bar**: located below the menu bar. This contains zooming (magnifying glass icons), panning (left/right 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. Zoom features are also built into the scale on the top of the navigation panel. Select an area to zoom to that location
   c. The **genome view**: this is where the data tracks are displayed. When viewing the annotation track (top most track), you can move upstream and downstream by dragging the track features left or right.
   d. **Select tracks**: Click on the "select track" button (top left). This menu contains all the data tracks that are aligned to the genome that you are viewing. The list of tracks can be filtered using the 'clickable' left panel categories, or with the search/filter function above the detailed right panel.



3. **Navigate to a specific gene in JBrowse**. The goal of this step is to navigate to the *HXGPRT gene of T. gondii* ME49.
   a. Make sure the *Toxoplasma gondii* ME49 genome is selected from the genome menu.
   b. You can either type the gene ID (TGME49_200320) or start typing the word *hypoxanthine* in the search box. After a few seconds you should see the result in the search dropdown. Click GO.



c. Details about most features are available in pop-up panels. Click (or control click) on the gene feature to view the details panel. You can also right click to choose the same details panel, a link to the gene page, or highlight the gene in yellow. What information is available in the popup?



In the previous exercises, you spent some time learning about gene pages and examining genes in the context of the JBrowse genome browser. It is important to recognize that gene models (structural annotation) are often open to interpretation, however, especially with respect to:

- transcript initiation and termination sites (5' and 3' untranslated regions, or UTRs)
- alternative processing events … if you sequence deep enough, virtually *all* genes (in organisms that process transcripts) display alternative splicing, even for single exon genes.
- the potential significance of non-coding RNAs

Even heavily curated genomes (*Plasmodium falciparum, Trypanosoma brucei, Saccharomyces cerevisiae*) do not fully reflect all available knowledge about stage-specific splicing, as new information is emerging all the time! In addition, many gene models were computationally derived using methods that may have not relied on experimental evidence supporting intron/exon boundaries (e.g. RNAseq data).

In this part of the exercise, we will explore genome browser track configuration options in greater detail, focusing on the interpretation of RNA-seq datasets, and using this information to examine the differentially spliced HXGPRT gene of T. gondii. You will then apply your newfound skills to examine other genes that may be alternatively spliced … and report your findings back to the group as a whole.

The screen shot below (Fig. 1) shows a sample of data tracks that can be turned on and configured in JBrowse. There are a few tracks that are worth examining which help in determining the accuracy of annotated gene models and that help in defining possible alternative splice variants of a gene. The link below will display the JBrowse view from figure 1, except for any special configurations which are not stored in the URL. For example, tracks 1c and 1d are collapsed in Fig 1 but will appear expanded in the JBrowse view after clicking on the link:

https://tinyurl.com/2p6f5738

- What evidence do each of the tracks provide?
- Are the ChIP-ChIP and Chip-seq tracks similar in what they show?
- Do you agree with the current annotated alternative splice forms of HXGPRT? Would you include any more?
- Are there other data tracks that might be useful to examine?

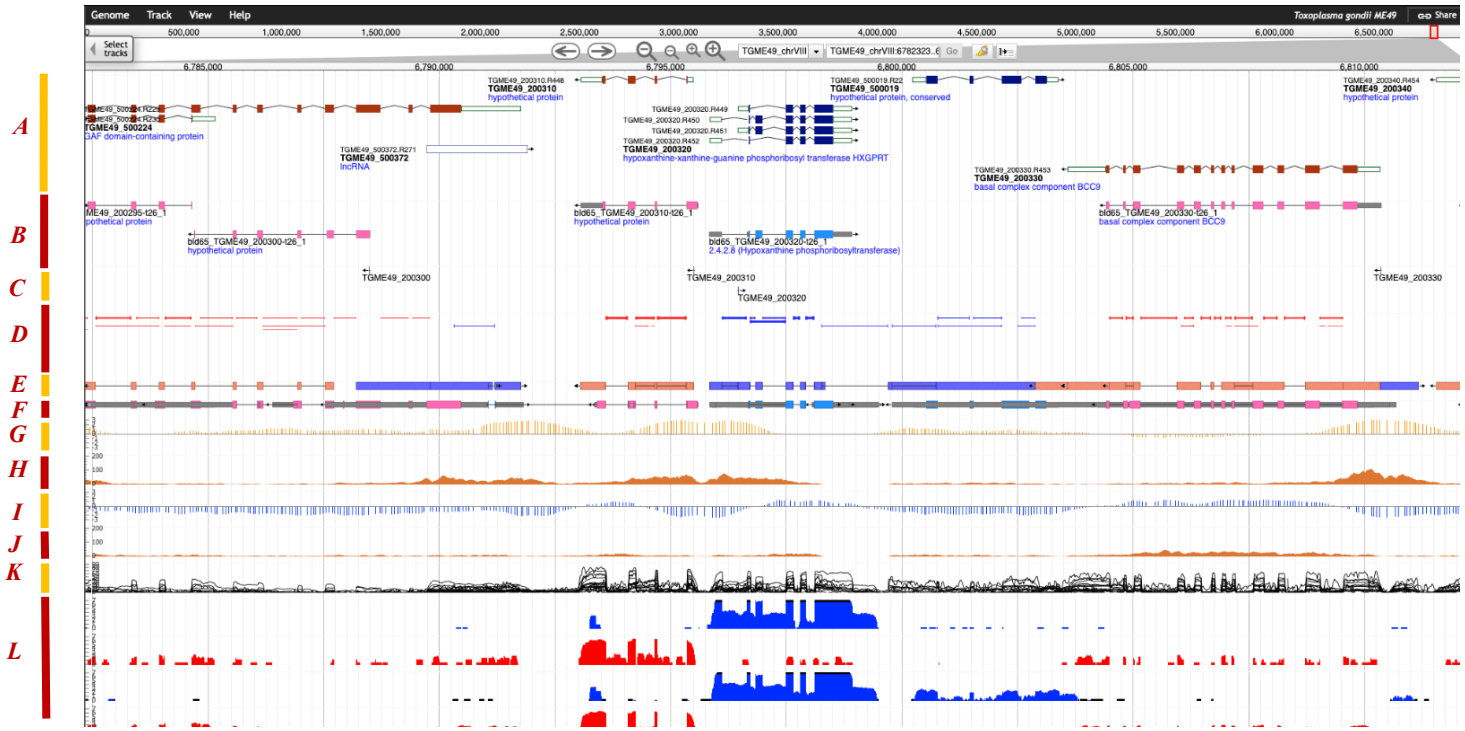


*Figure : Screen shot from ToxoDB JBrowse. **A**. Official gene models. **B**. Release 65 gene models. **C**. Predicted transcription start sites. **D**. Splice junction evidence based on available RNAseq data. **E**. Nanopore long-read transcriptomic data (collapse view). **F**. Alternative gene models using RNAseq evidence from 12 experiments (collapsed view). **G**. Chip-ChIP H3K9ac. **H**. Chip-Seq H3K4me3. **I**. Chip-ChIP H3K4me1. **J**. Chip-Seq H3K4me1. **K**. Combined RNA-seq. **L**. RNAseq coverage from Toxoplasma gondii strain CZ clone H3 in feline enteroepithelial stage (strand specific).*

Working in groups, please examine the genes in your list, to evaluate their official gene models based on RNA-seq data and any other available evidence. See if you can discover which exon(s) were represented … and determine whether ***Working in groups, please examine the genes in your list, to evaluate their official gene models based on RNA-seq data and any other available evidence. See if you can discover which exon(s) were represented … and determine whether*** ***these genes are actually alternatively spliced (constitutively or stage-specifically). We will then reconvene to hear a brief report from each group.***

Group 1:	**Group 4:**	**Group 7:**
TGME49_278510	TGME49_201270	TGME49_281440
TGME49_256650	TGME49_214440*	TGME49_208718
TGME49_283540	TGME49_250115	TGME49_222930
Group 2:	**Group 5:**	**Group 8:**
TGME49_265390	TGME49_261720	TGME49_217490
TGME49_225730	TGME49_268610	TGME49_292150
TGME49_288000	TGME49_266310	TGME49_276170
Group 3:	**Group 6:**	**Group 9:**
TGME49_213660	TGME49_280380	TGME49_297850
TGME49_297160	TGME49_293720	TGME49_299010
TGME49_256025	TGME49_248445	TGME49_240470

***** It is not possible to search in JBrowse for genes with prime in the product description, i.e. *4'-phosphopantetheinyl*. To examine TGME49_214440 please search for the gene next to it, TGME49_214430 and then navigate to TGME49_214440.