Unannotated Intron Junctions search

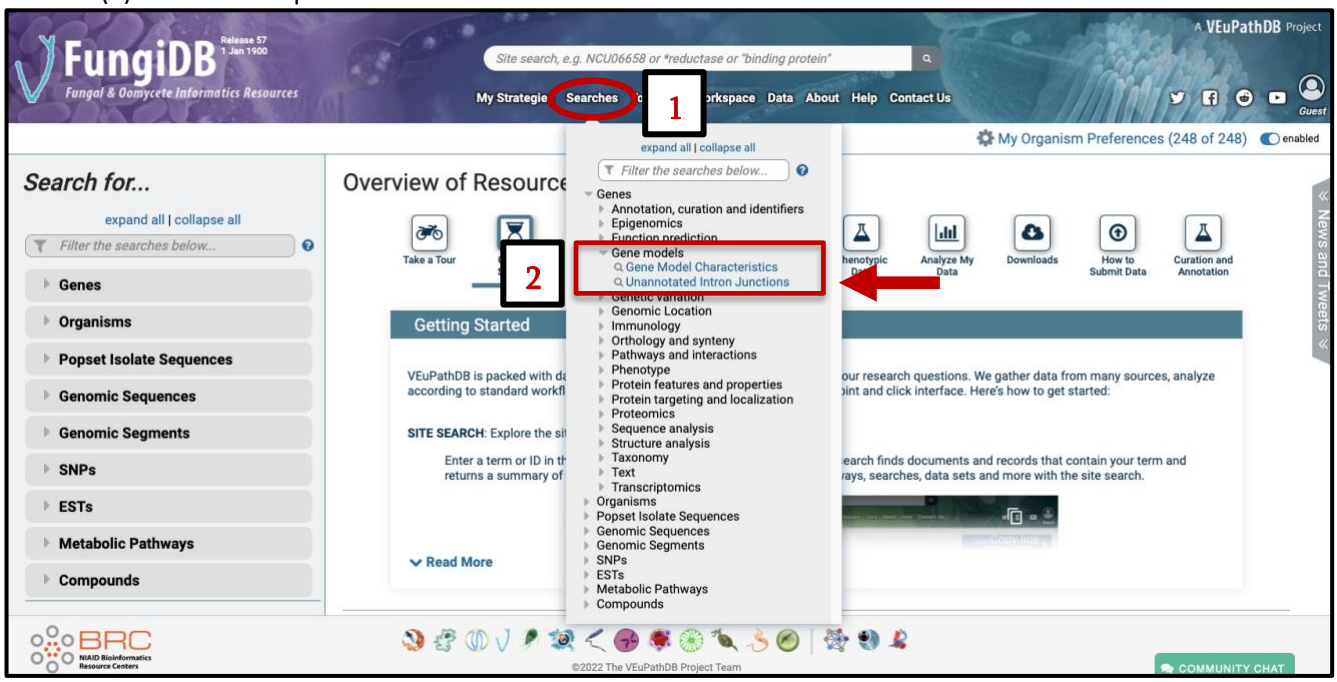
Please note this search is not available on TriTrypDB, TrichDB and HostDB.

The **Unannotated Intron Junctions** search enables users to identify genes that contain, or are flanked by, unannotated high confidence intron junction-spanning reads from RNA-seq data. These genes may be incompletely or inaccurately annotated due to missing introns/exons and/or alternative splice variants. Once you've identified the genes with unannotated introns you can explore them in JBrowse and correct gene structures in Apollo, an open-source software enabling users to inspect, refine and add gene models to the current genome annotations. Apollo is available for reference genomes.

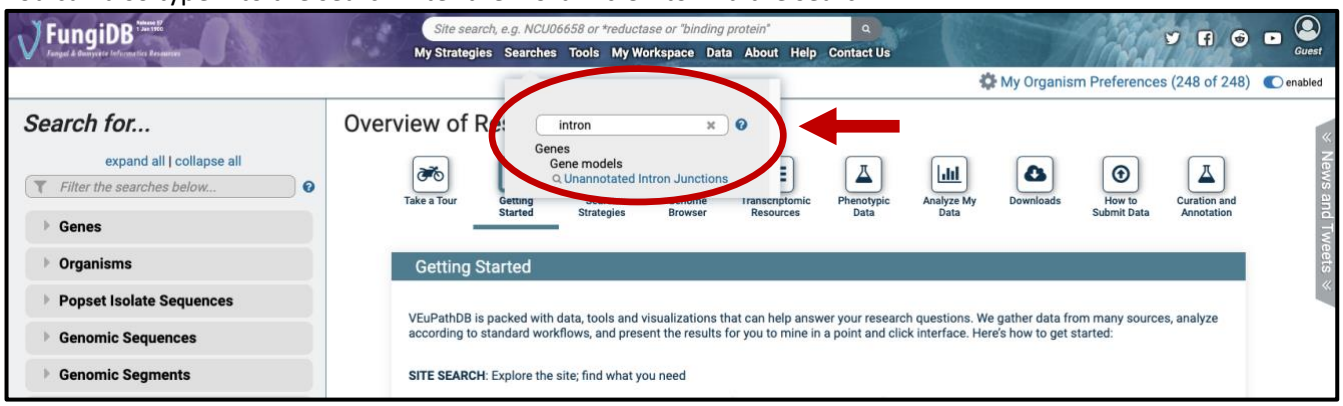
We will show you in this tutorial how to identify possible incorrect gene structures and correct them in Apollo.

1) Accessing the search Unannotated Intron Junctions

To access the search option in any VEuPathDB site go to the **Searches** (**1**) menu and choose from the **Gene models** (**2**) section the option **Unannotated Intron Junctions**.



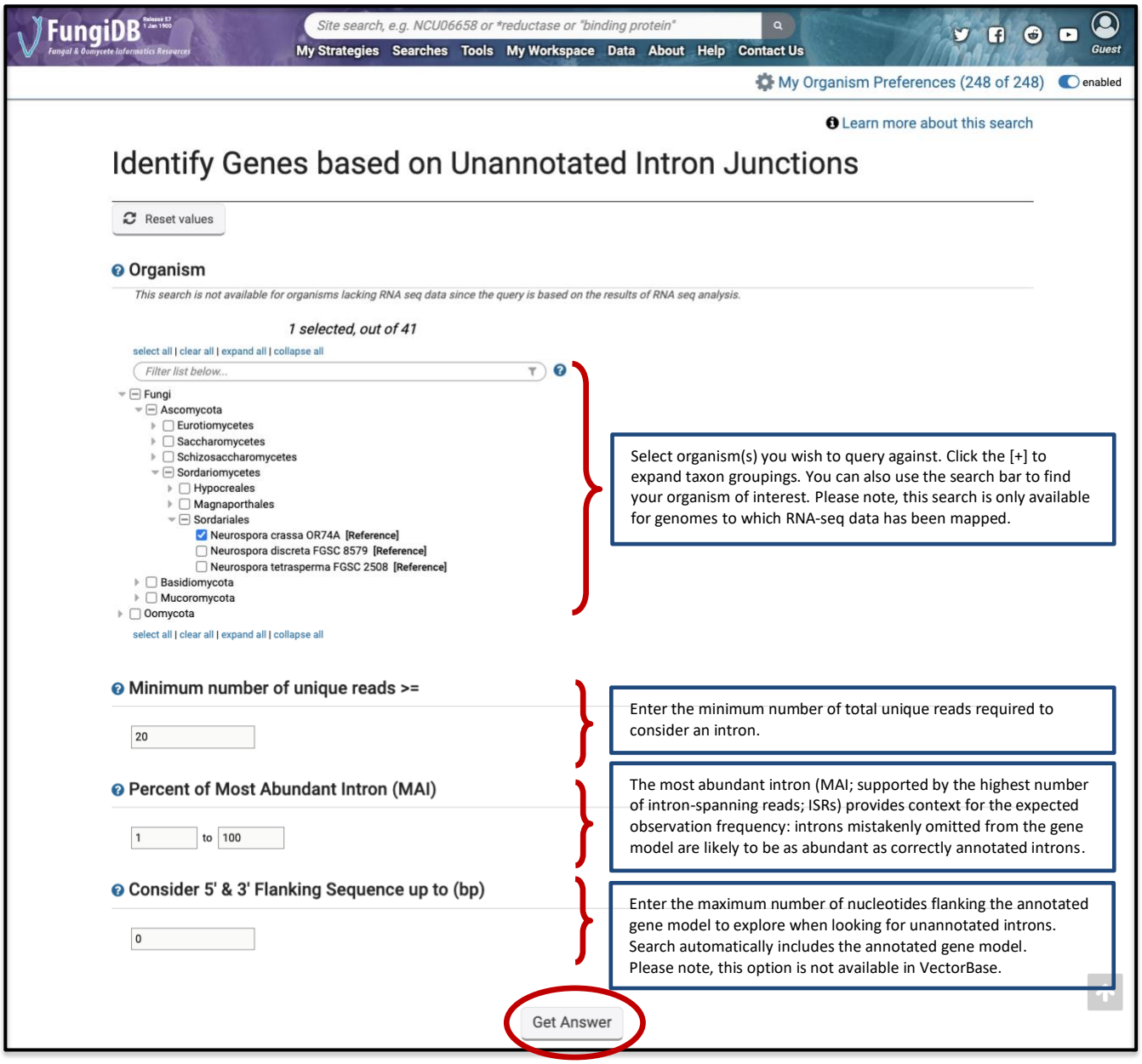
You can also type into the search filter the word **intron** to find the search.



2) Select the search parameters

There are a number of parameters to manipulate in this search:
- Organism
- Minimum number of unique reads
- Percent of most abundant intron (MAI)
- 5'and 3' Flanking sequence (not available in VectorBase)

Using the default parameters on this search you will get a first impression on the number of genes with unannotated introns. If you think this number is too high to explore the data, change the search parameters, the minimum number of unique reads or percent of most abundant intron.

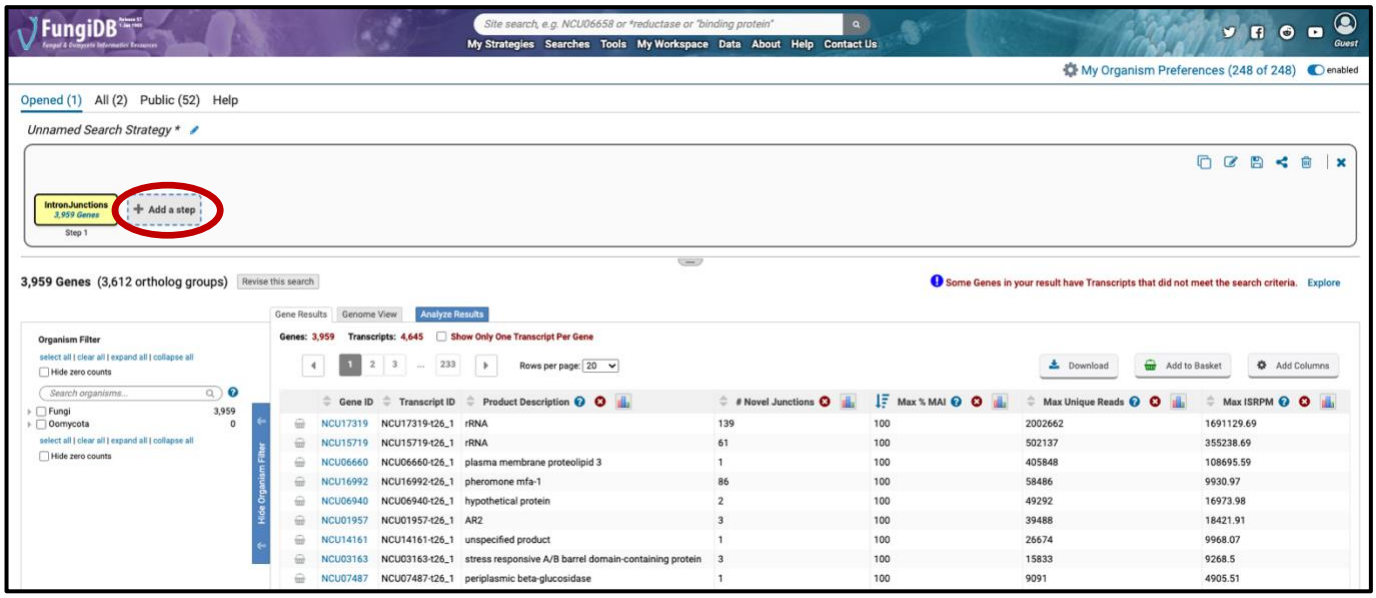


3) Explore the result table

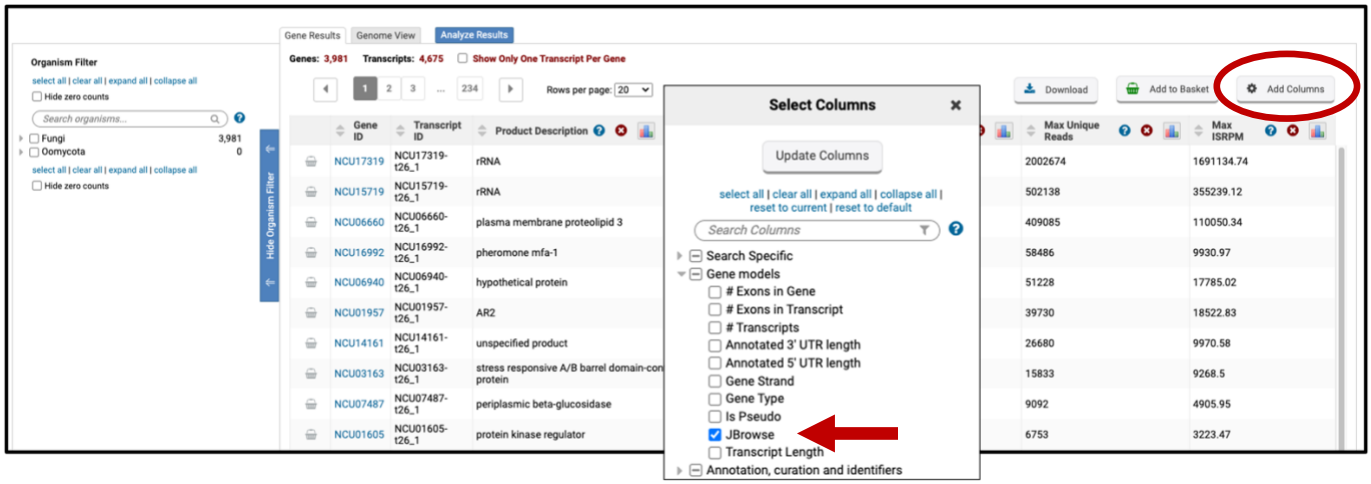
You can order the search result by using the grey arrows on the top of the columns.

- **Number of Novel Splice Junctions**. In case the number is 1, this means your gene has 1 possible unannotated intron. If this number is quite high, i.e. over 50 there is a possibility that your gene of interest is a rRNA, located in a repetitive region or it is part of a gene family. Therefore it is important to explore the results in JBrowse/Apollo with additional evidence.
- **Max % Mai**: Maximum percentage of intron with the maximum total unique reads in this gene for the novel introns that met search criteria
- **Max Unique Reads**: Maximum total unique reads for the novel introns that met search criteria
- **Max ISRPM**: Maximum total ISRPM (Intron Spanning Reads Per Million) for the novel introns

Note: Add an additional step to this search strategy if you would like to combine your search result with an additional search, i.e. a list of gene IDs or number of exons in a gene.



You can explore the results in JBrowse. For easy access use the option Add Columns to include a new column with links to JBrowse.

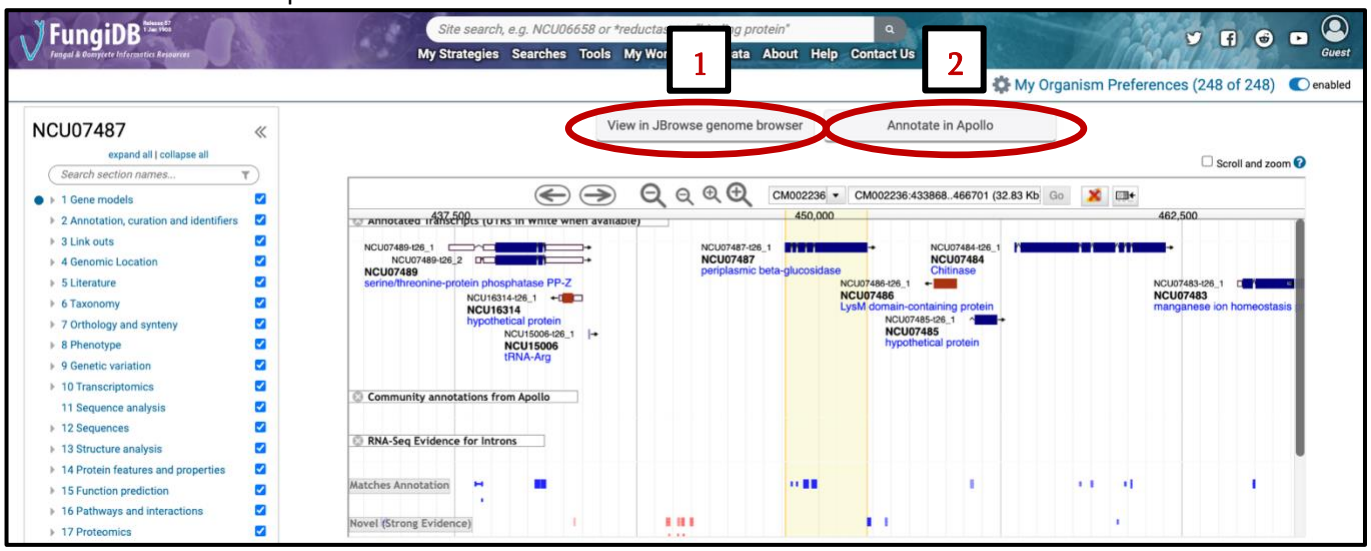


4) Explore the results in JBrowse/Apollo

Click on the JBrowse link in the table. Alternatively, go to the gene record page of your gene of interest by clicking on the Gene ID.

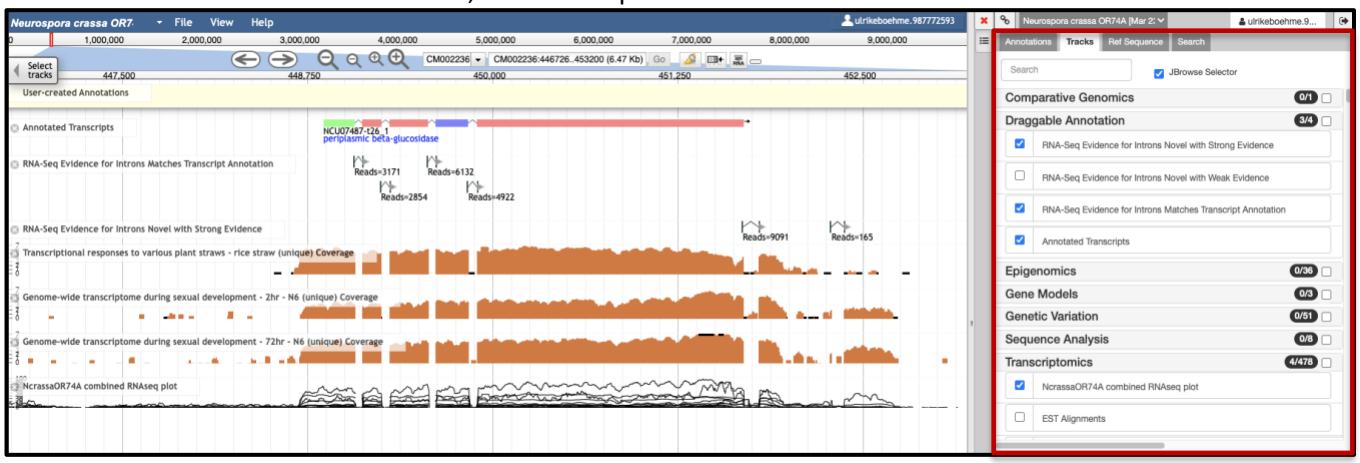
Gene ID	Transcript ID	Product Description	# Novel Junctions	Max % MAI	Max Unique Reads	Max ISRPM	JBrowse
NCU17319	NCU17319-t26_1	rRNA	139	100	2002662	1691129.69	JBrowse
NCU15719	NCU15719-t26_1	rRNA	61	100	502137	355238.69	JBrowse
NCU06660	NCU06660-t26_1	plasma membrane proteolipid 3	1	100	405848	108695.59	JBrowse
NCU16992	NCU16992-t26_1	pheromone mfa-1	86	100	58486	9930.97	JBrowse
NCU06940	NCU06940-t26_1	hypothetical protein	2	100	49292	16973.98	JBrowse
NCU01957	NCU01957-t26_1	AR2	3	100	39488	18421.91	JBrowse
NCU14161	NCU14161-t26_1	unspecified product	1	100	26674	9968.07	JBrowse
NCU03163	NCU03163-t26_1	stress responsive A/B barrel domain-containing protein	3	100	15833	9268.5	JBrowse
NCU07487	NCU07487-t26_1	periplasmic beta-glucosidase	1	100	9091	4905.51	JBrowse

Select **View in JBrowse genome browser (1)** to explore your search result. For reference genomes, you can view and correct the gene structure in Apollo. To open Apollo select **Annotate in Apollo (2)** on the gene record page. To use Apollo you need to be logged into VEuPathDB. If you have not done so yet log now in with your VEuPathDB user ID and password.

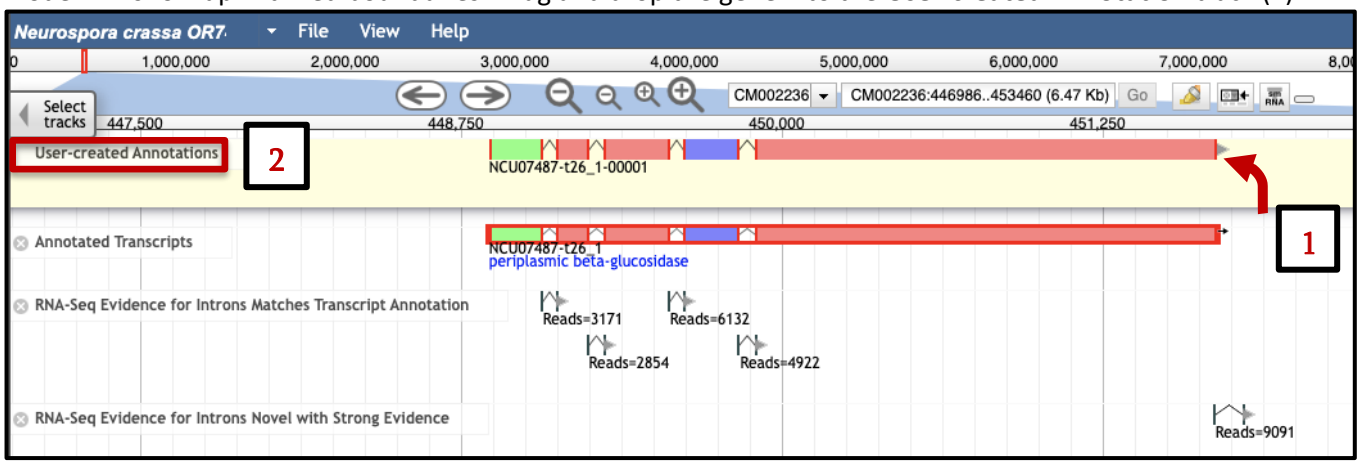


5) Correct the gene structure in Apollo

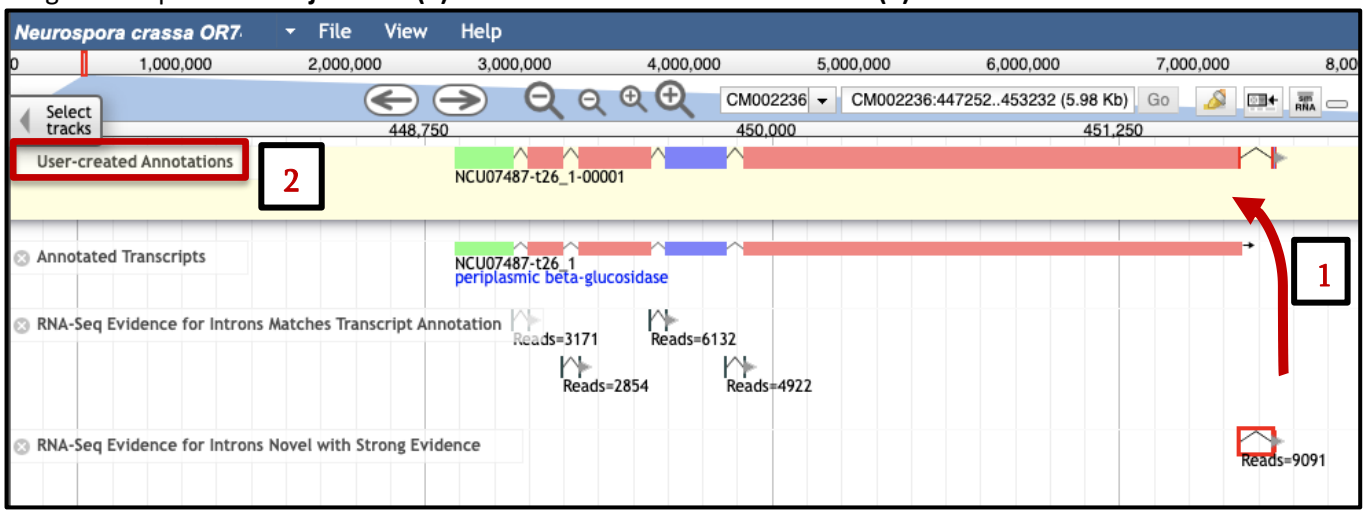
Select on the right-hand side the tab **Tracks.** Click on the menu item **Draggable Annotation** select **RNA-Seq Evidence for Introns Novel with Strong Evidence** and **RNA-Seq Evidence for Introns Matches Transcript Annotation**. Select additional evidence, i.e. Transcriptomics evidence.



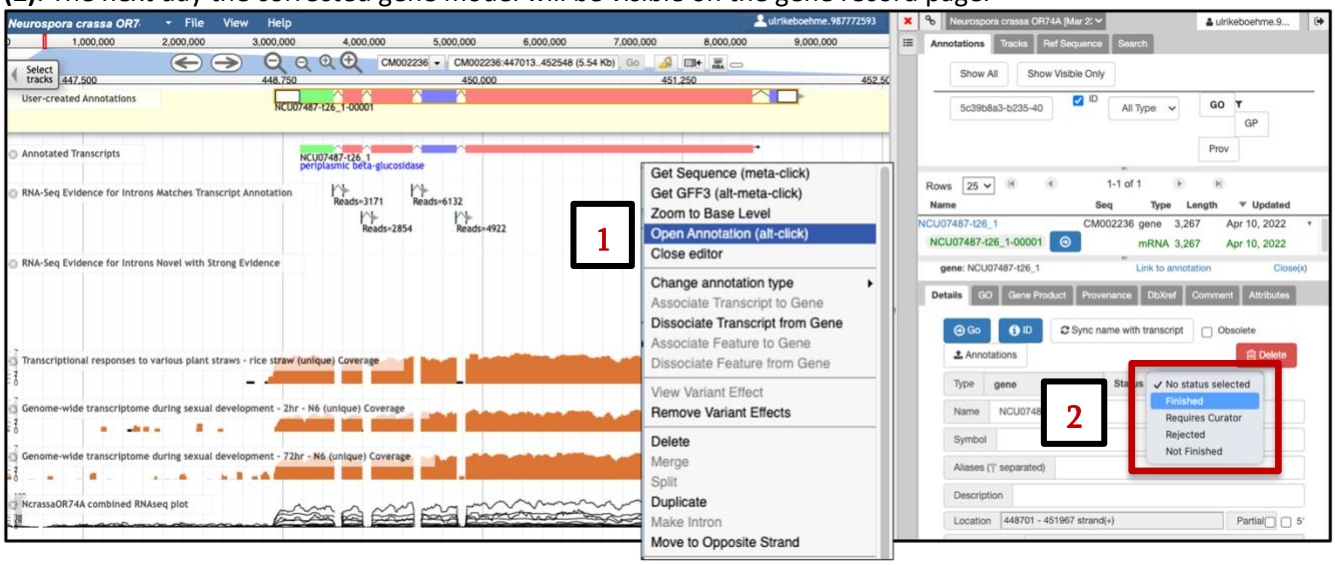
Select the gene model by clicking on one of the introns or by double clicking on the gene model (**1**). The gene model will show up with red boundaries. Drag and drop the gene into the **User-created Annotation** track (**2**).



Drag and drop the **intron junction (1)** into the **User-created Annotations (2)** area.



With a right-click on the gene in the **User-created annotations** area you can access the menu. Once you are happy with the corrected gene model choose **Open Annotation** (1) and select from the status menu **Finished (2)**. The next day the corrected gene model will be visible on the gene record page.



Additional information and tutorials on Apollo can be found in the Help section on VEuPathDB!