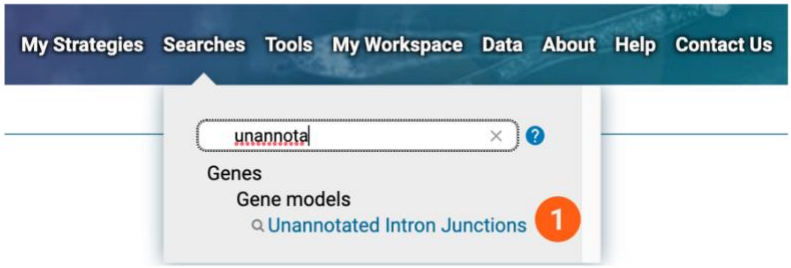
Assessing (& editing) gene annotation (JBrowse/Apollo) (optional)

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

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.
Note: This search is only available for genomes with mapped RNA-Seq datasets.

- **Identifying possible incorrect gene structures via the "Unannotated Intron Junctions "search.**
    1. **Deploy the "Unannotated Intron Junctions" search.**



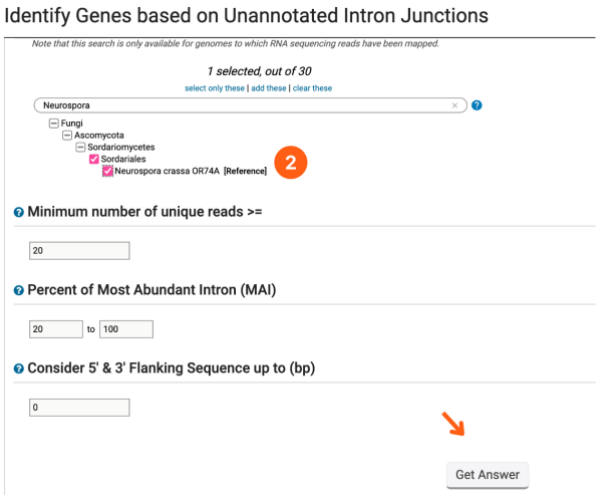
    2. **Set search parameters.**

Organism: *Neurospora crassa*
Minimum number of unique reads: keep at default.
Percent of most abundant intron (MAI): keep at default.

Note: The most abundant intron (MAI; supported by the highest number of intron-spanning reads; ISRs) provides context for the expected observation frequency: introns mistakenly omitted from the gene model are likely to be as abundant as correctly annotated introns.
- Consider 5' and 3' Flanking sequence up to (bp): keep at default.

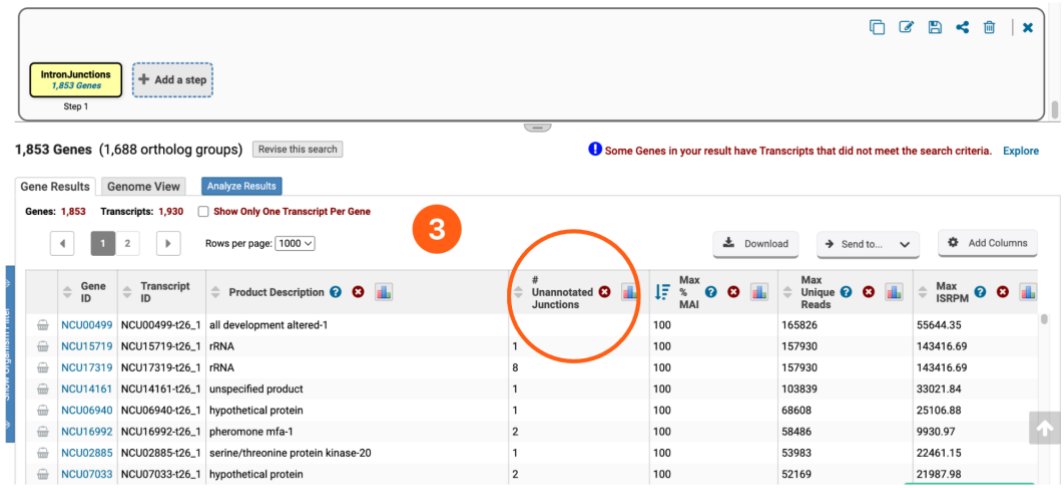
Note: Here you can enter the maximum number of nucleotides flanking the annotated gene model to explore when looking for unannotated introns. Search automatically includes the annotated gene model.

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 results table.

Note: Search results can be ordered by using the "Novel junctions" filter.

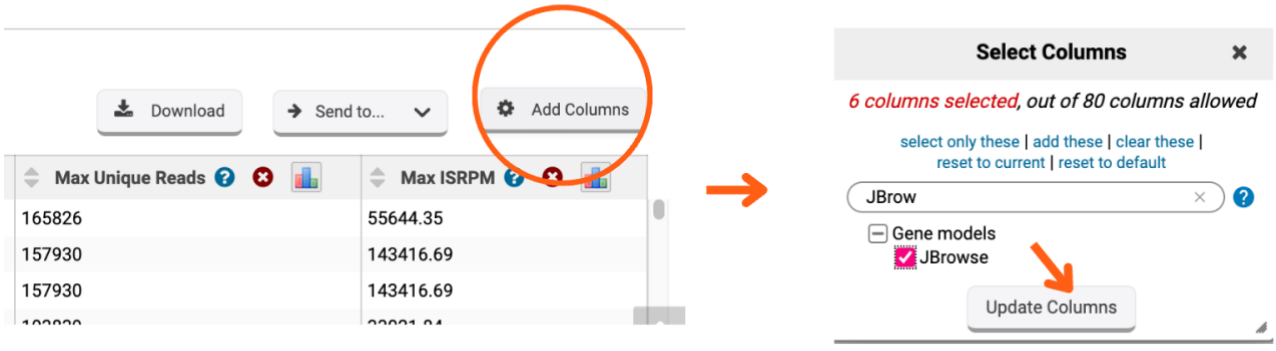


Number of Novel Splice Junctions. In case the number is 1, this means your gene has 1 possible unannotated intron. If this number is quite high, i.e. over 50 there is a possibility that your gene of interest is a rRNA, located in a repetitive region or it is part of a gene family. Therefore, it is important 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.

This search can be combined with the "Gene Model Characteristics" search to limit the results on the number of exons in the gene. This may be useful if you want to look possible structural annotation errors in muti-exon genes only:

Exploring evidence in JBrowse.

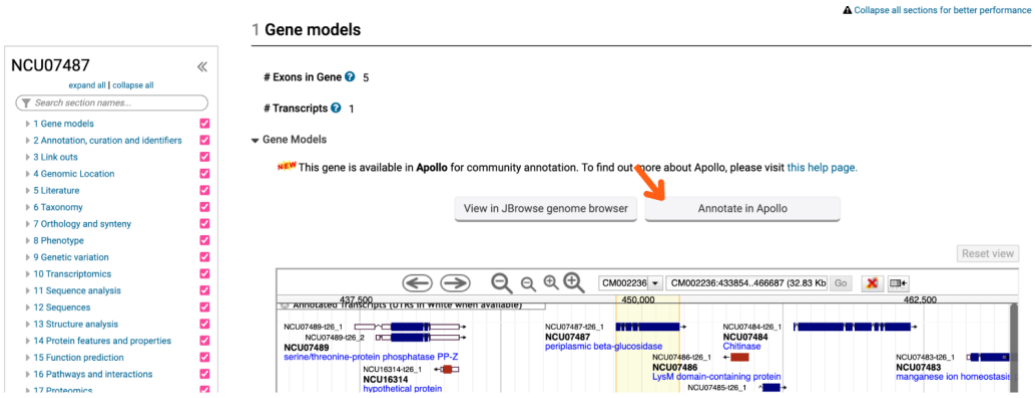
Clicking on the Gene ID linked in blue will re-direct you to the gene record page where you can click on JBrowse button. However, you can also modify the results table to include direct JBrowse links for easy navigation. To do this, click on the "Add Columns" button and select JBrowse from the menu.



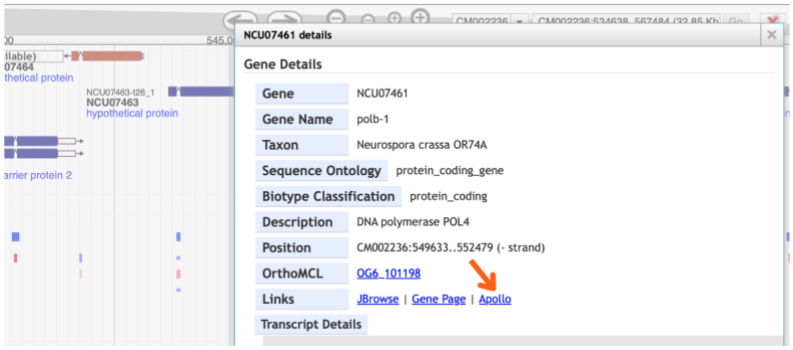
	Gene ID	Transcript ID	Product Description	# Unannotated Junctions	Max % MAI	Max Unique Reads	Max ISRPM	JBrowse
	NCU00499	NCU00499-t26_1	all development altered-1	1	100	165826	55644.35	JBrowse
	NCU15719	NCU15719-t26_1	rRNA	1	100	157930	143416.69	JBrowse
	NCU17319	NCU17319-t26_1	rRNA	8	100	157930	143416.69	JBrowse
	NCU14161	NCU14161-t26_1	unspecified product	1	100	103839	33021.84	JBrowse

Correcting gene structure in Apollo.

Apollo can be accessed from gene record pages:



and also in JBrowse (left click on the gene to brin up the pop-up window):

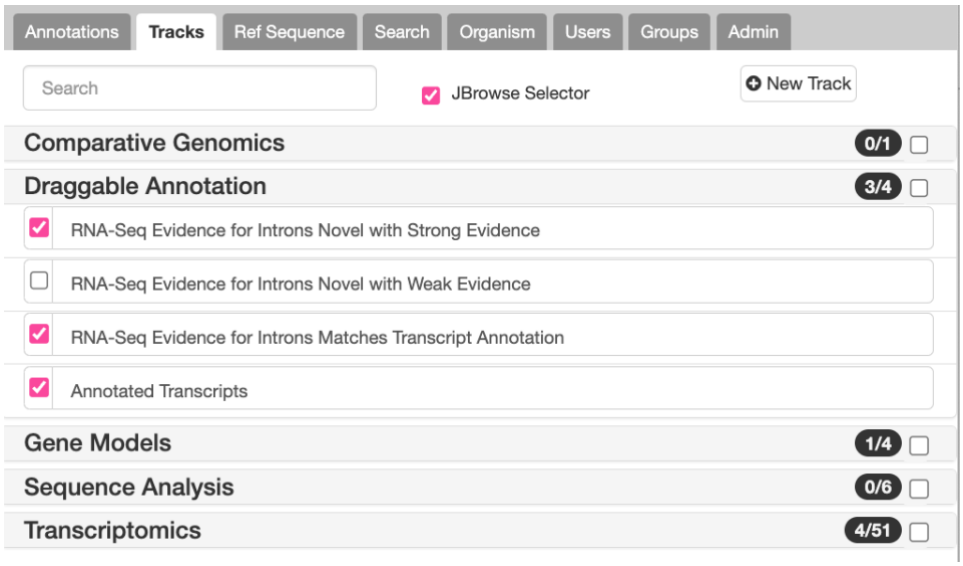


Once in Apollo, use the right panel to select the “Tracks” tab to bring up the following tracks:

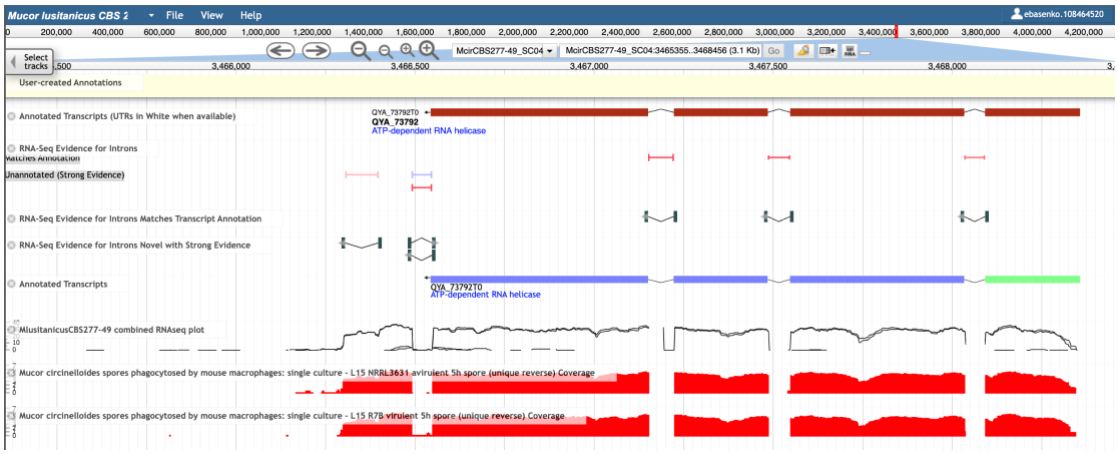
Draggable Annotation

Check the box to select the following tracks:

- RNA-Seq Evidence for Introns Novel with Strong Evidence
- RNA-Seq Evidence for Introns Matches Transcript Annotation
- Annotated transcripts

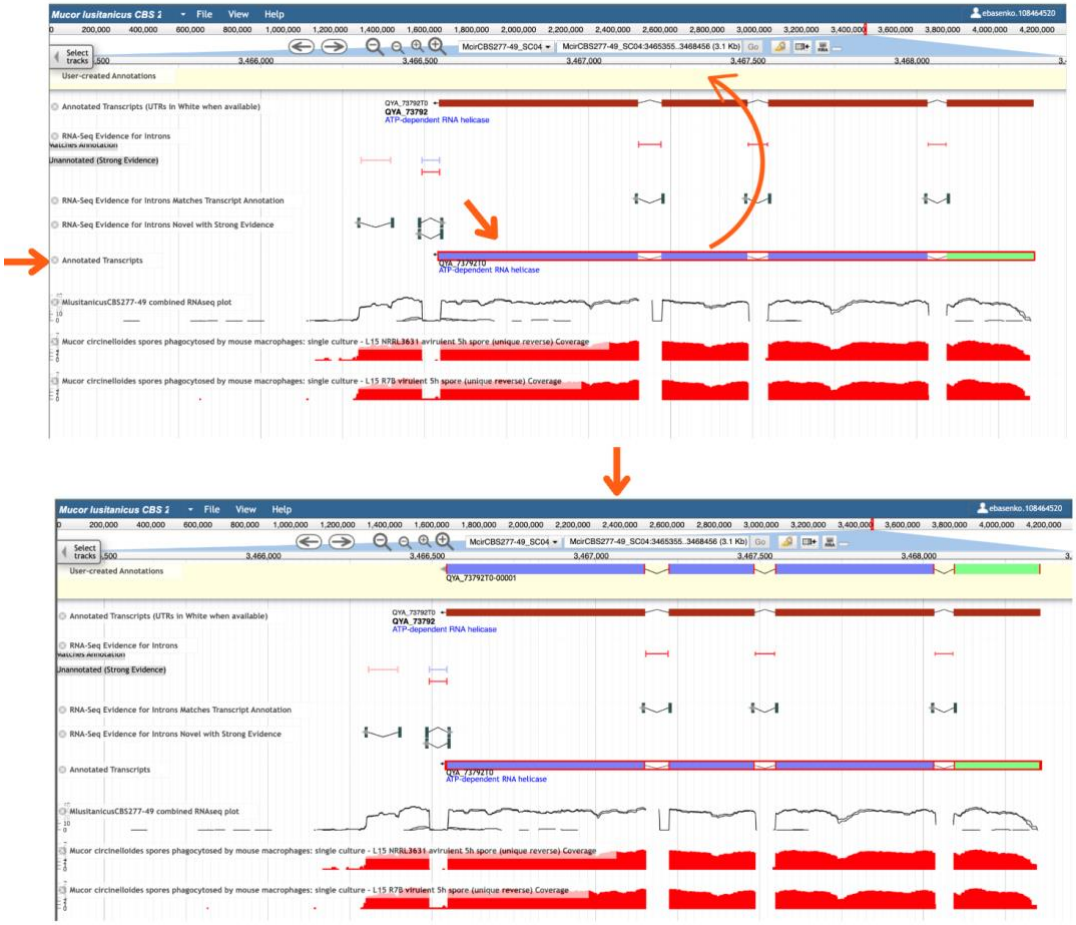


Note: You can also deploy several unique reverse RNA-Seq tracks as a guide when making changes to the structural gene annotation. The RNA-Seq tracks are available under the “transcriptomics” menu.



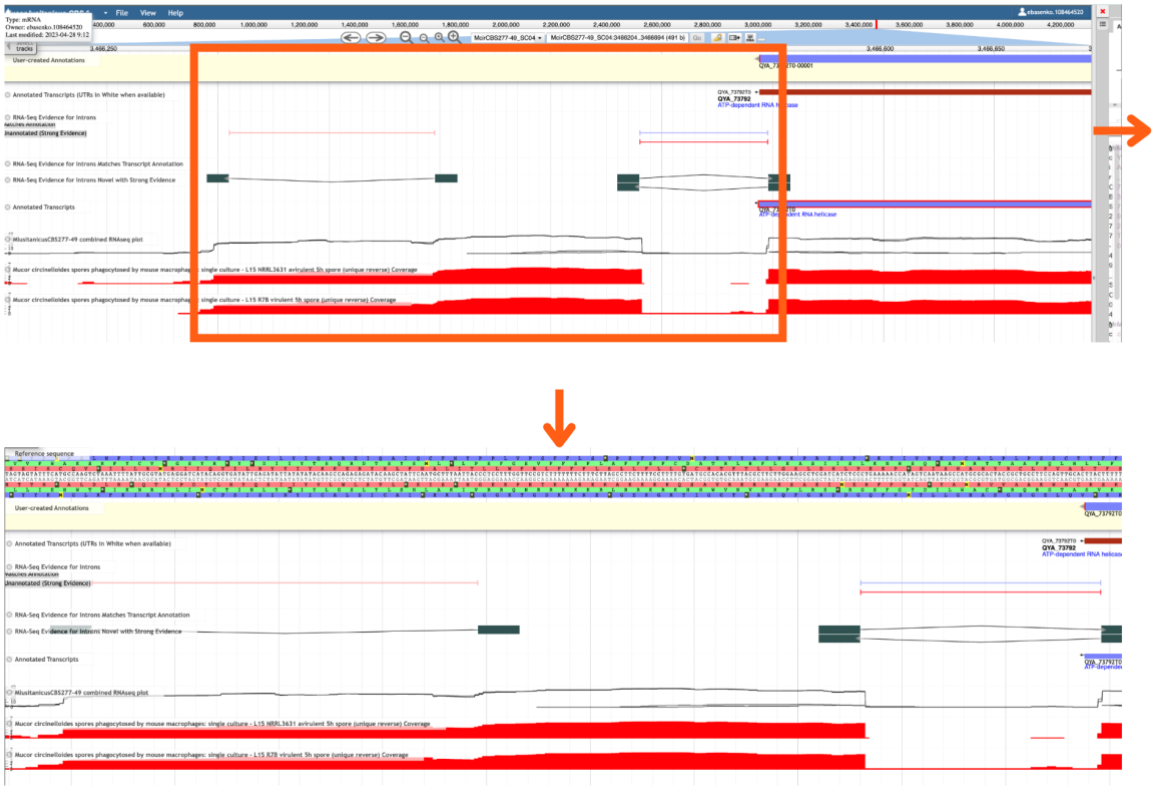
1. **Drag a gene model into the User-created Annotation workspace.**

To do this, double click on the gene in the “Annotated Transcripts” tracks which was selected from the “Draggable Annotation” section on the right. Double-clicking will highlight the whole gene rather an individual component. Note: The “Annotated Transcripts (UTRs in White when available” track cannot be used for this purpose.



2. Activate the “Reference sequence” track to guide gene correction.

To be able to view the Reference sequence track, you must be zoomed in a good bit.
You may want to drag the right window to the right to create more working space within the Apollo editor and then use the cursor to zoom in to the section highlighted by the orange box:

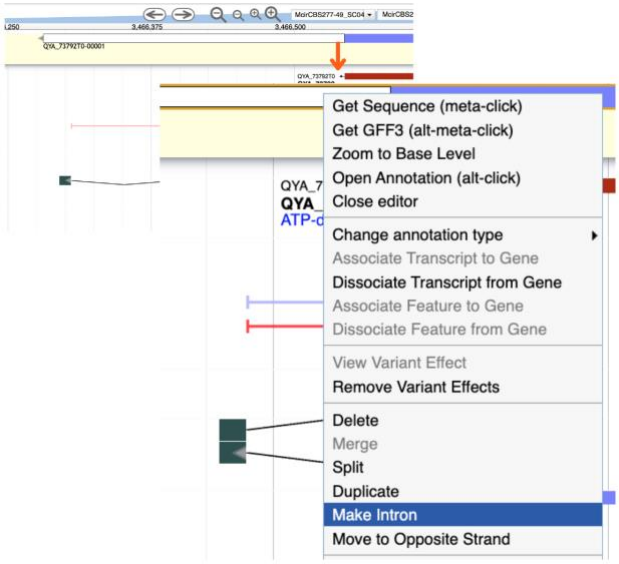


3. Extend gene model using evidence tracks for guidance.

Hover over the track in the “User-created Annotation” until a small black arrow appears at the end of the track. Left click on the arrow and extend gene model:

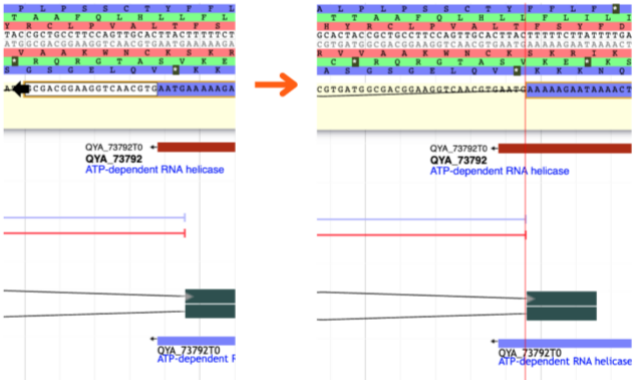
4. Create intron.

Zoom out, right-click on the white box gene feature created by Apollo as a result of the gene model extension, and select "Make intron" option.

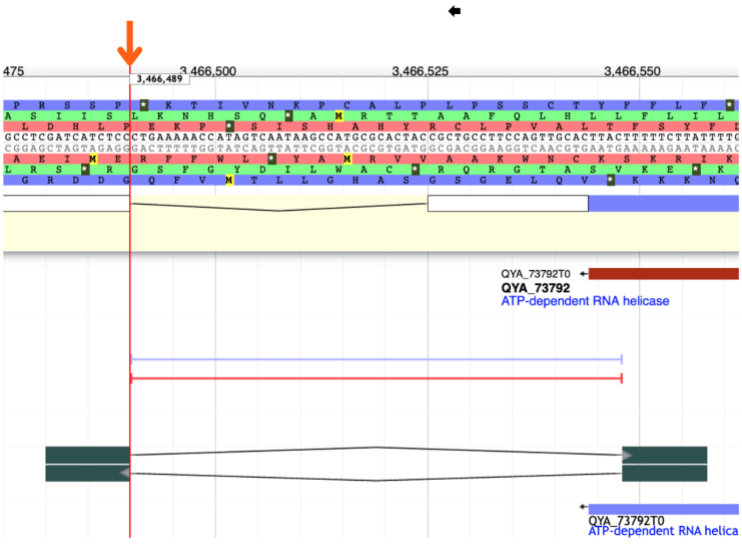


5. Modify intron boundaries.

Apollo will automatically create an intron feature. Now, zoom in to adjust the boundaries and use the "RNA-Seq Evidence for introns Novel with Strong Evidence" track for guidance. In this case, the left boundary looks ok:



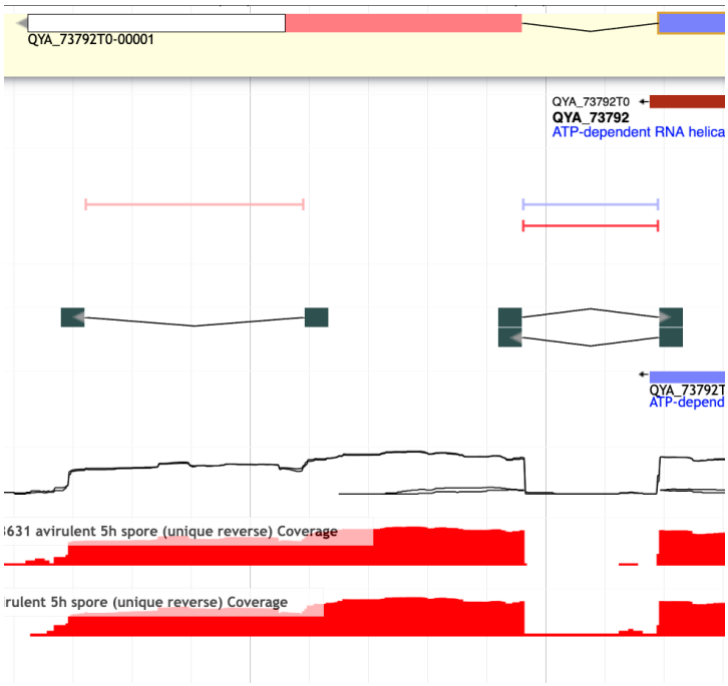
However, the right boundary should be adjusted.



Click to select the gene model and hover over until a small black arrow appears next to the boundary.

Use the arrow to adjust the intron boundary. You can use the cursor to check the boundary alignment with the RNA-Seq data.

Notice that this adjustment automatically annotated an extra exon. Apollo also automatically predicted a UTR.



Once the new gene model is complete, navigate to the Annotations > Details, etc. tabs to provide evidence and comments. One the status is changed to “Finished” the new gene model will become visible for other users.