

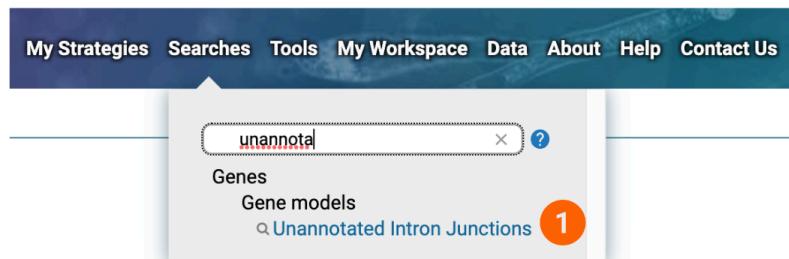
Assessing (& editing) gene annotation (JBrowse/Apollo) (optional)

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: *Mucor lusitanicus* CBS 277.49

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.

Identify Genes based on Unannotated Intron Junctions

Reset values to default

Organism
Note that this search is only available for genomes to which RNA sequencing reads have been mapped.
1 selected, out of 30
 select only these | add these | clear these

muco
 Fungi
 Mucoromycota
 Mucor lusitanicus CBS 277.49 [Reference] 2

Minimum number of unique reads >=

Percent of Most Abundant Intron (MAI)
 to

Consider 5' & 3' Flanking Sequence up to (bp)

↓

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.

The screenshot shows a search results table with the following columns:

	# Unannotated Junctions	Max % Mai	Max Unique Reads	Max ISRPM
1	100	1906	801.98	
1	62.1	118	68.54	
1	22.5	170	82.25	
1	42	50	16.58	
1	47.2	351	181.81	

A red circle highlights the column header "# Unannotated Junctions". A red arrow points to the value "100" in the first row.

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 multi-exon genes only:

The sidebar on the right shows the following options:

- ① Choose how to combine with other Genes:
 - 1 INTERSECT 2 (selected)
 - 1 UNION 2
 - 1 MINUS 2
 - 2 MINUS 1
- ② Choose which Genes to combine. From...
 - A new search (selected)
 - An existing strategy
 - My basket

A red arrow points to the search input field "model" in the sidebar. Below the sidebar is a histogram showing the distribution of Gene/Transcripts versus Gene Exon Count.

Gene Exon Count	Gene/Transcripts
5	~1000
6	~1000
7	~1000
8	~1000
9	~1000
10	~1000

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.

The screenshot shows a results table with several columns. At the top right, there is a "Add Columns" button, which is circled in red. An arrow points from this button to a "Select Columns" dialog box. The dialog box has a title "Select Columns" and a message "6 columns selected, out of 80 columns allowed". It contains a list of selected columns under the heading "JBrow" and "Gene models". The "JBrowse" checkbox is checked. At the bottom right of the dialog box is a "Update Columns" button, which is also circled in red.

	Gene ID	Transcript ID	Product Description	# Unannotated Junctions	Max % MAI	Max Unique Reads	Max ISRPM	JBrowse
QYA_157425	QYA_157425T0		1-Acyl dihydroxyacetone phosphate reductase and related dehydrogenases	1	100	1906	801.98	JBrowse
QYA_148011	QYA_148011T0		2-oxoisovalerate dehydrogenase subunit alpha [Source:UniProtKB/TREMBL;Acc:A0A168I9R9]	1	62.1	118	68.54	JBrowse
QYA_116745	QYA_116745T0		3-hydroxy-3-methylglutaryl coenzyme A reductase [Source:UniProtKB/TREMBL;Acc:A0A168MWS8]	2	22.5	170	82.25	JBrowse
QYA_72770	QYA_72770T0		3-hydroxyacyl[acyl-carrier protein] dehydratase [Source:UniProtKB/TREMBL;Acc:A0A162MQ1]	1	42	50	16.58	JBrowse
QYA_153891	QYA_153891T0		3-hydroxyacyl-CoA dehydrogenase	1	47.2	351	181.81	JBrowse
QYA_151919	QYA_151919T0		3-hydroxyisobutyryl-CoA hydrolase, mitochondrial [Source:UniProtKB/TREMBL;Acc:A0A168ND88]	1	23.5	28	13.63	JBrowse
QYA_157539	QYA_157539T0		40S ribosomal protein S1 [Source:UniProtKB/TREMBL;Acc:A0A168MX07]	1	100	2155	1162.14	JBrowse

Strategy URL:

<https://fungidb.org/fungidb/app/workspace/strategies/import/d5cddb62413777fa>

Correcting gene structure in Apollo.

Apollo can be accessed from gene record pages:

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

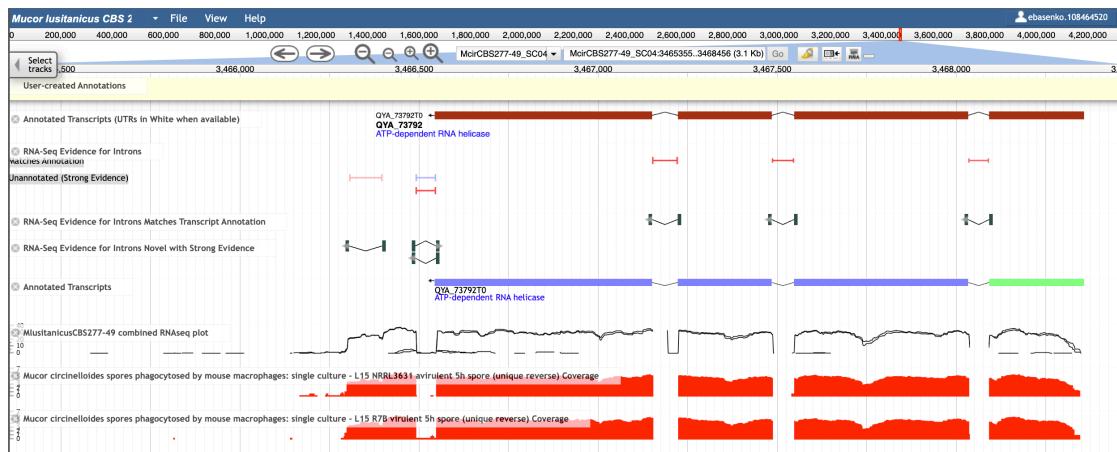
Once in Apollo, use the right panel to select the “Tracks” tab to bring up the following tracks:

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.

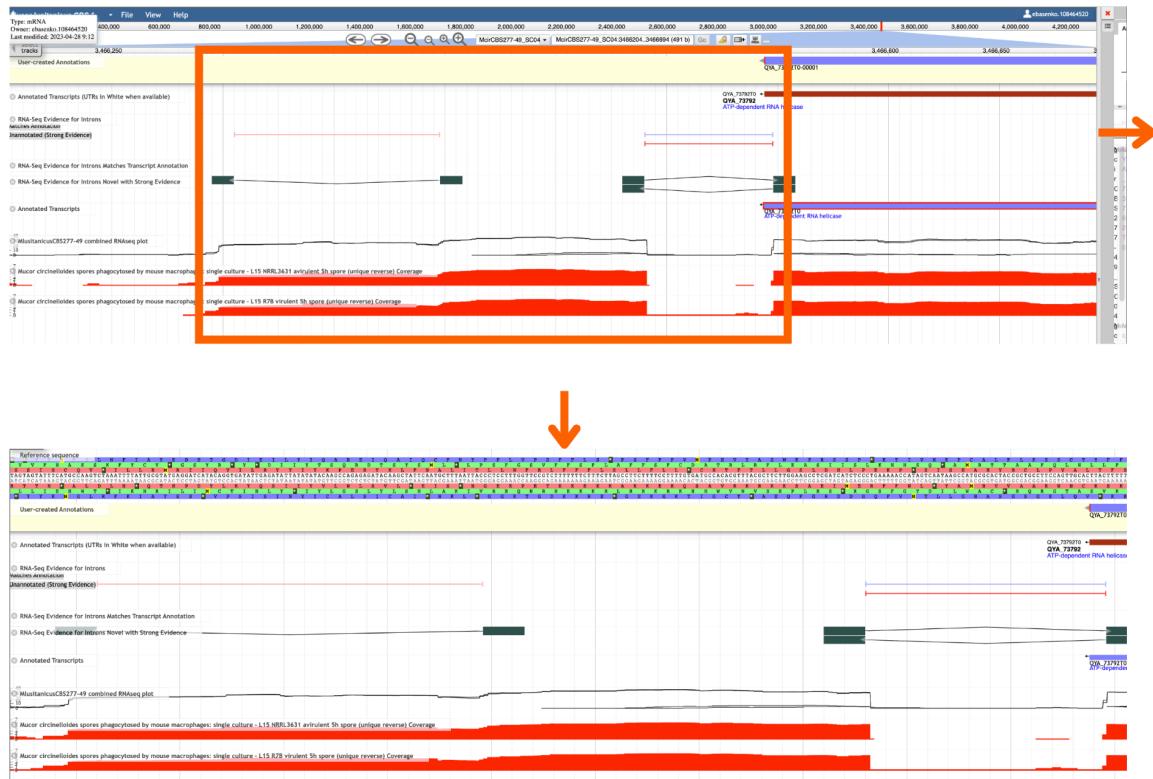
To do this, double click on the gene in the “Annotated Transcripts” tracks which was selected from the “Draggable Annotation” section on the right. Double-clicking will highlight the whole gene rather than an individual component. Note: The “Annotated Transcripts (UTRs in White when available” track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



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

Activate the “Reference sequence” track from the Track menu on the left. 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:



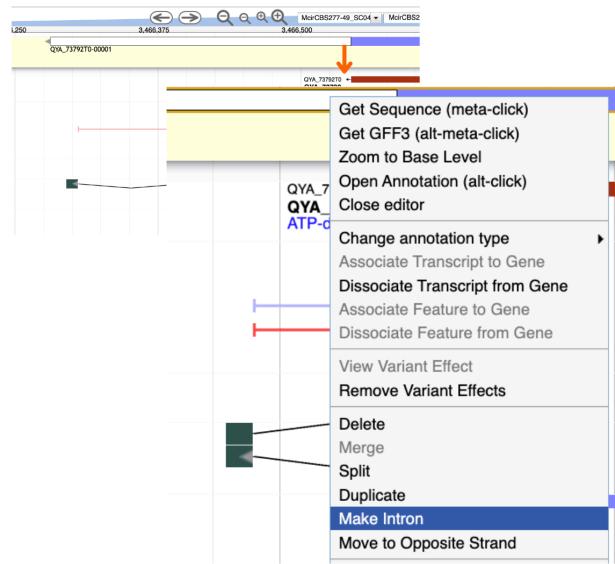
3. Extend gene model using evidence tracks for guidance.

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:



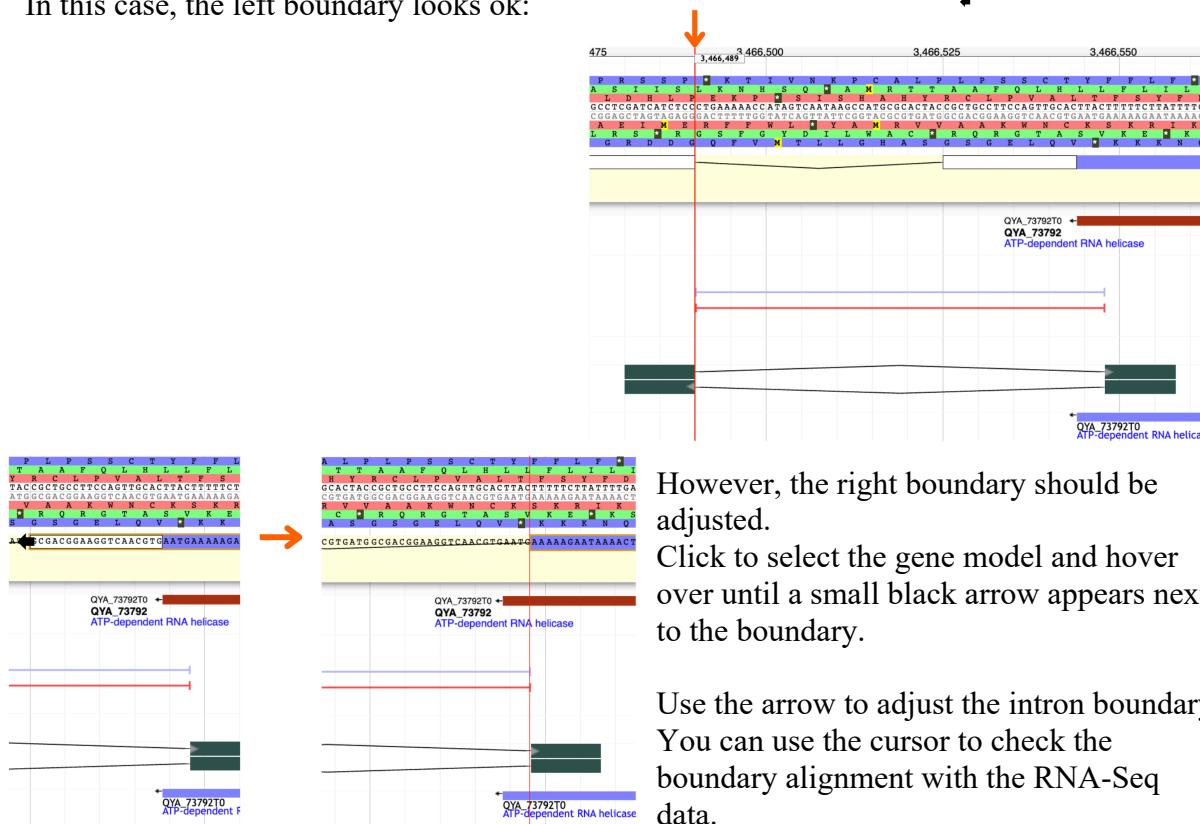
4. Create intron.

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:

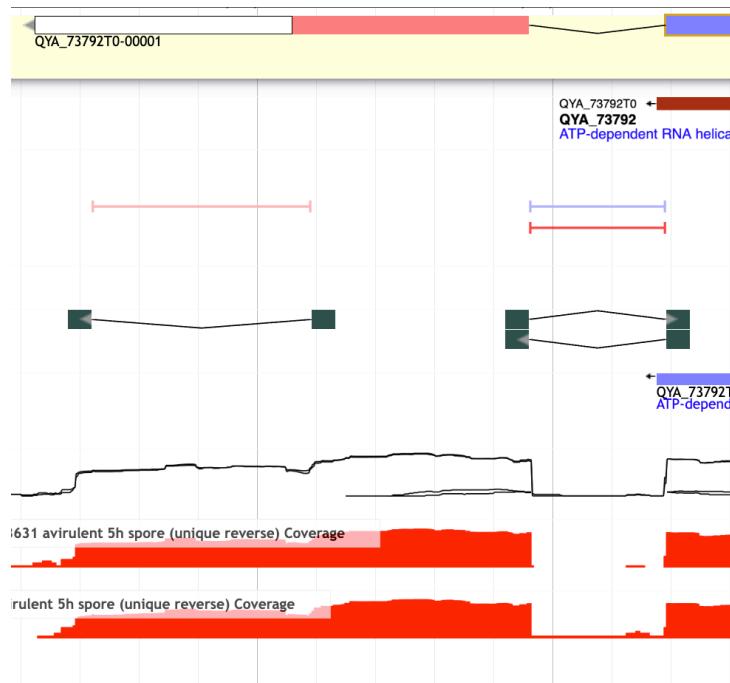


However, the right boundary should be adjusted.

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. Once the status is changed to “Finished” the new gene model will become visible for other users.

Annotations Tracks Ref Sequence Search Organism Users Groups Admin

Show All Show Visible Only

Annotation Name ID All Types GO GP Prov

Reference Sequence All Users All Statuses

Rows 25 1-50 of 1,334

Name	Seq	Type	Length	Updated
QYA_73792T0	MciCBS277-49_SC04	gene	2,076	Apr 28, 2023
MciCBS277-49_SC01aaaaag	MciCBS277-49_SC01	gene	2,435	Jan 03, 2023
QYA_105303T0	MciCBS277-49_SC01	gene	523	Sep 08, 2022
QYA_157431T0	MciCBS277-49_SC10	gene	1,629	Sep 08, 2022
QYA_149982T0	MciCBS277-49_SC10	gene	873	Sep 08, 2022
QYA_149802T0	MciCBS277-49_SC10	gene	3,209	Sep 08, 2022
QYA_14515T0	MciCBS277-49_SC06	gene	4,234	Sep 08, 2022
MciCBS277-49_SC06c	MciCBS277-49_SC06	gene	2,193	Sep 08, 2022
MciCBS277-49_SC06a	MciCBS277-49_SC06	gene	1,879	Sep 08, 2022
QYA_75161T0	MciCBS277-49_SC06	gene	1,011	Sep 08, 2022
MciCBS277-49_SC03aaad	MciCBS277-49_SC03	gene	1,502	Sep 07, 2022
QYA_152667T0	MciCBS277-49_SC03	gene	2,578	Sep 07, 2022
MciCBS277-49_SC03aac	MciCBS277-49_SC03	gene	745	Sep 07, 2022
QYA_140074T0	MciCBS277-49_SC03	gene	809	Sep 07, 2022
QYA_93110T0	MciCBS277-49_SC09	gene	5,230	Sep 07, 2022
QYA_75293T0	MciCBS277-49_SC09	gene	2,461	Sep 07, 2022
QYA_15569T0-00001	MciCBS277-49_SC04	gene	1,338	Sep 07, 2022
QYA_15679RT0	MciRS277-49_SC06	gene	1,912	Sep 07, 2022

Link to annotation Close(x)

Details GO Gene Product Provenance DoXref Comment Attributes

Go ID Sync name with transcript Obsolete Annotations Delete

Type: gene Status: Finished

Name: QYA_73792T0

Symbol:

Aliases ("|" separated):

Description:

Location: 3466281 - 3468356 strand()

Ref Sequence: MciCBS277-49_SC04

Owner: ebasenko.108464520

Created: Apr 28, 2023 09:14 AM

Updated: Apr 28, 2023 09:14 AM