

Genome annotation in Apollo

Apollo is a real-time collaborative genome annotation and curation platform. More information can be found on this publication:

<https://pubmed.ncbi.nlm.nih.gov/30726205/>

Learning objectives:

- Accessing Apollo
- Use the menu and navigation bar of Apollo
- Add pre-loaded data tracks
- Changing gene structures in Apollo
- Adding functional annotation with metadata in Apollo

1) Accessing Apollo

Links to Apollo are available from multiple locations:

- 1) The gene record page from the link **View and update community annotations in Apollo**
- 2) The gene record page by clicking on the button **Annotate in Apollo**.
- 3) The tools menu in the banner of any page.
- 4) Go directly to Apollo with the following link: <https://apollo.veupathdb.org>.

The screenshot shows the VectorBase website interface. At the top, there is a search bar and navigation tabs: My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. The main content area displays the gene record for AEL010827, which is a programmed cell death protein. The gene details section includes fields for Gene Type, Biotype Classification, Chromosome, Location, Species, Strain, and Status. A red circle highlights the link "View and update community annotations in Apollo" under the Status field. A red circle and the number "1" highlight this link. A red circle and the number "2" highlight the "Annotate in Apollo" button at the bottom right. A red circle and the number "3" highlight the "Apollo" link in the top navigation bar. The sidebar on the left contains a list of sections: 1 Gene models, 2 Annotation, curation and identifiers, 3 Link outs, 4 Genomic Location, 5 Literature, and 6 Taxonomy. The "Annotate in Apollo" button is also highlighted with a red circle and the number "2".

For this exercise we will use the VEuPathDB Sandbox Apollo instance. This Apollo instance is available to get familiar with all menus and tools. The changes will not affect any of the organism's official gene set, neither will they be preserved.

Access the Sandbox with the following link:

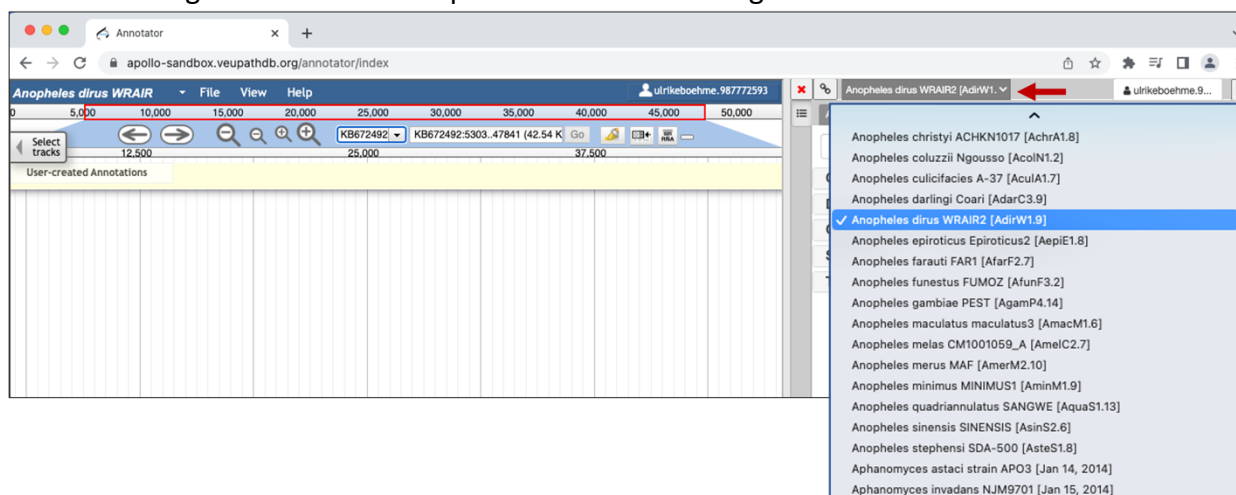
<https://apollo-sandbox.veupathdb.org/annotator/index>

To use Apollo you need to be logged into VEuPathDB. If you have not done so yet log now into VEuPathDB with your user ID and password. If you do not have an account with VEuPathDB, register for one here:

<https://vectorbase.org/vectorbase/app/user/registration>

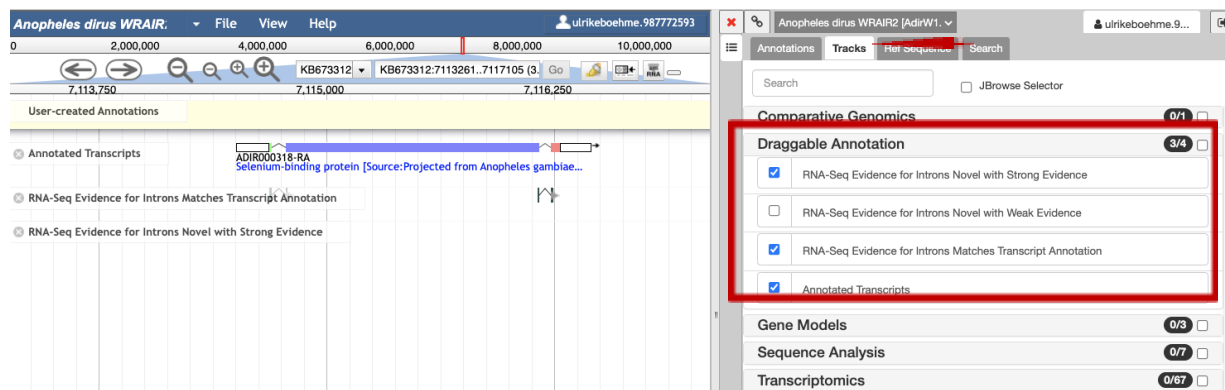
2) Choosing the organism

Choose the organism from the drop-down menu on the right-hand side.

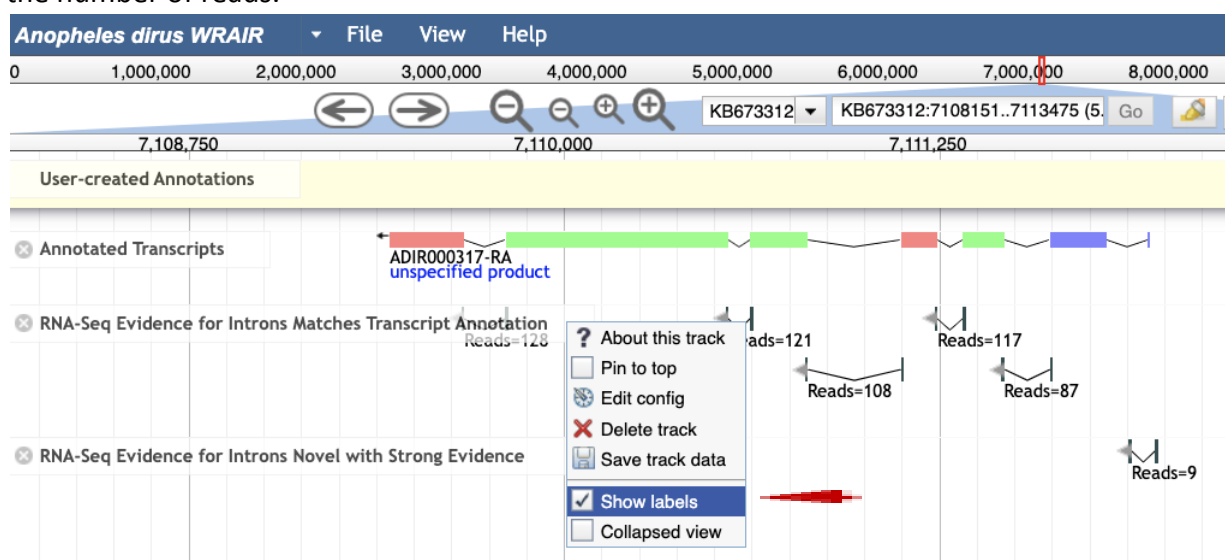


3) Adding Draggable annotation

Click on Tracks and select from Draggable Annotation **Annotated Transcripts** and **RNA-Seq evidence for Introns**.

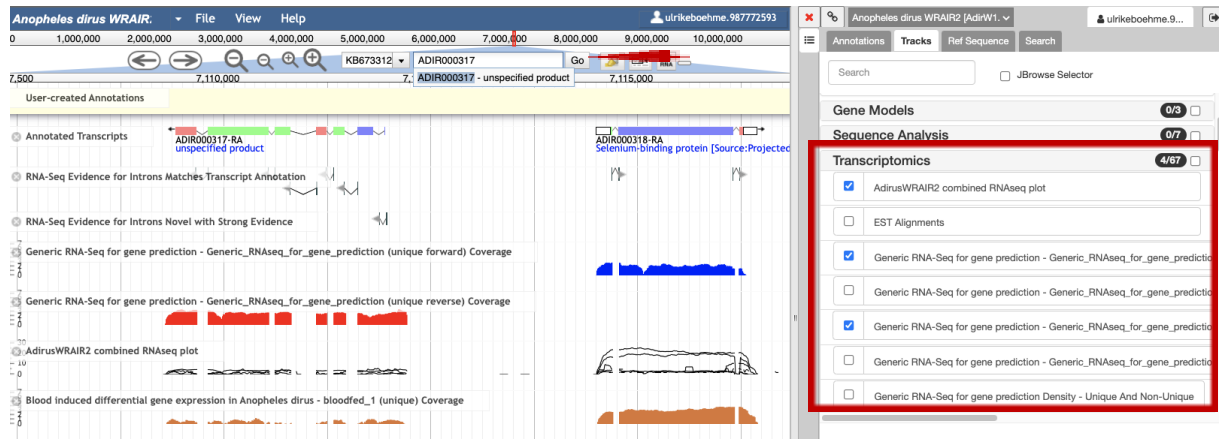


All tracks have a drop-down menu. Choose **Show labels** from the drop-down menu to see the number of reads.



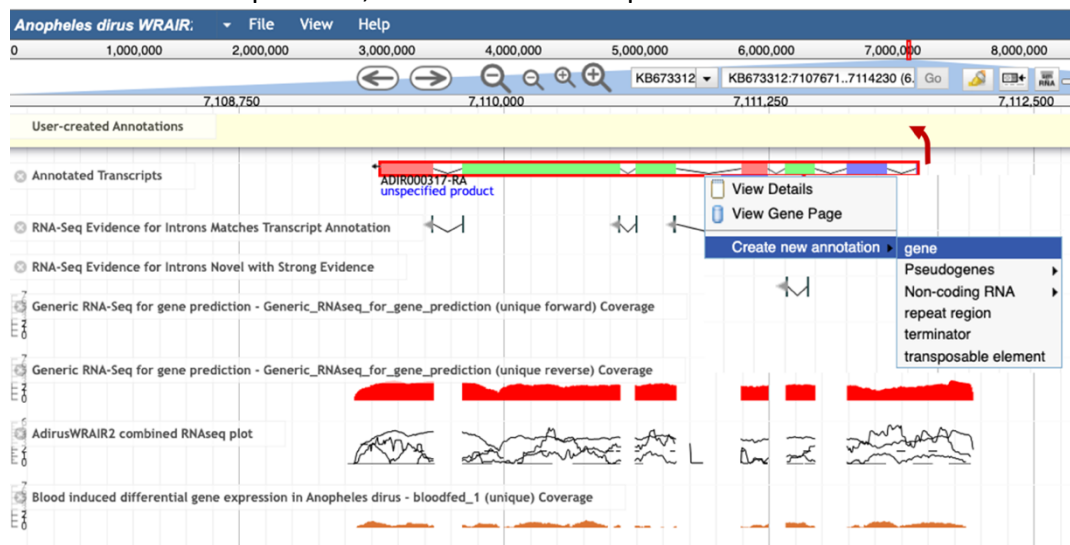
4) Find your gene or location of interest and add additional evidence

Find your gene of interest by typing the gene ID or location into the search bar. Choose from the Transcriptomics menu additional evidence tracks. Hint: A track that is useful is the combined RNAseq plot.



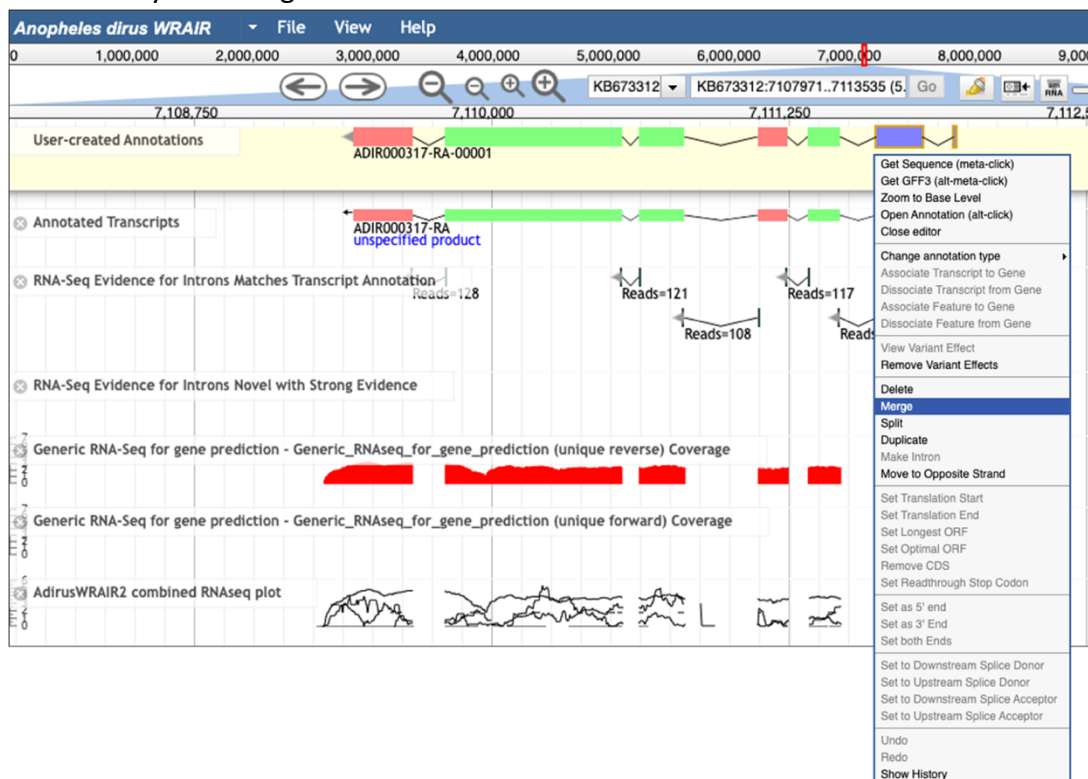
5) Add the gene model into the User-created Annotations area

Select the gene model by clicking on one of the introns. Drag and drop the gene model into the User-created Annotations track. Alternatively, you can also right click on the gene in the Annotated Transcripts track, select from the drop-down menu **Create new annotation > gene**.

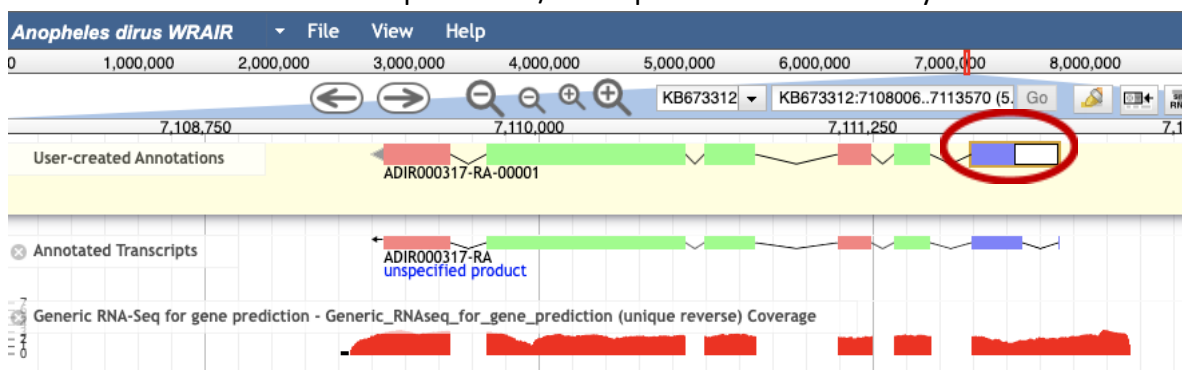


6) Modify the gene model

With a right-click you can access the menu. For this example two exons were selected with the shift-key and merged.



To create UTR's, point your mouse at the edge of the feature, a little arrow will appear, then extend the exon to the transcription start/end. Apollo will automatically create UTRs.



7) Add functional annotation

With a right-click open the menu and select Get Sequence.

The screenshot displays the Anopheles dirus WRAIR genome browser. The top navigation bar includes 'File', 'View', and 'Help' menus. A scale bar at the top indicates genomic coordinates from 0 to 9,000,000. The main view shows several tracks: 'User-created Annotations' with a gene model 'ADIR000317-RA-00001', 'Annotated Transcripts' with 'ADIR000317-RA unspecified product', 'Generic RNA-Seq for gene prediction', and 'AdirusWRAIR2 combined RNAseq plot'. A right-click context menu is open over the gene model, listing options such as 'Get Sequence (meta-click)', 'Get GFF3 (alt-meta-click)', 'Zoom to Base Level', 'Open Annotation (alt-click)', 'Close editor', 'Change annotation type', 'Associate Transcript to Gene', 'Dissociate Transcript from Gene', 'Associate Feature to Gene', 'Dissociate Feature from Gene', 'View Variant Effect', 'Remove Variant Effects', 'Delete', and 'Merge'. Below the main view, a 'Sequence' window is open, showing the nucleotide sequence of the selected gene model.

Sequence

Copy to Clipboard Search Sequence Copied!

```
>0fe2d434-66e9-4dee-9889-87e23055fcf2 (sequence: mRNA) 534
residues [KB673312:7109276-7112195 - strand] [peptide]
MLEEEDSDRIKICPCEKESAKKSCAMVYCSVENQGDHKKKTCPPKICSNQ
YGRFLVATROIKGGEVLKESPLVHGPAQITGVCVGLQGLEENKYCECHGMPVCKR
SCQDHPAQACEKFTIARGSKI SIQHYVPIPTVQCMLPLCLLAEEDPARWDALLRL
SHEDERRGSGQWRNDREGVAKLIPRFKCNKNWDEILRVVGVVQVNGHEVPLETPSSV
ALYNASHLESCHPILTKSFTHNDEIMAPNPIRKDEKLSICTVDVHWFTORLZHLQ
QYKMFCECERCSDPSEYETHFSAIRCSTGQKDKCKGYLLPVDKDCWTDGQCLACSK
VTGAAVQCIQERARMDLEAMQHREDCNKYKHYKSWLAPHQYLVDVKISLSQVIGGN
NPEATKKIPTDKLNNKIKICQDLIALFEKICPAEAPAGATRFELHAALAEGRSTEN
SSAMAAVQLEDELBNNECTRLHLEFVLSKESKICNQAQVNAQTLKTLGIKE
```

Copy to Clipboard. You can find out more about your gene by running BLAST and InterPro

<https://vectorbase.org/vectorbase/app/workspace/blast/new>

<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

<https://www.ebi.ac.uk/interpro>

Select the gene model in the User-created Annotations area and with a right-click open the menu. Choose **Open Annotation**. If you want to add a new product choose **Gene Product**.

Add your gene product with evidence code and metadata, i.e. InterPro domain, blast match, publication. More information on evidence codes can be found here: <https://geneontology.github.io/docs/guide-go-evidence-codes>

Edit Gene Product for ADIR000317-RA

GO Annotation Guidance

Product

SET domain-containing protein

☐ Alternate

Evidence

ECO:0000255

☐ All
ECO
Evidence

ISM

(ECO:0000255):

inferred from
sequence

model

Evidence Code

Info

With

InterPro

:

IPR046341

+ Add

Reference

Prefix

:

ID

Note

+ Add

Save

Cancel

8) Finalise your annotation

Once you are happy with your structural and functional annotation select the status **Finished**.

ADIR000317-RA KB673312 gene 2,920 May 18, 2022

ADIR000317-RA-00001 mRNA 2,920 May 18, 2022

gene: ADIR000317-RA [Link to annotation](#) [Close\(x\)](#)

Details GO Gene Product Provenance DbXref Comment Attributes

Go ID Sync name with transcript ☐ Obsolete Annotations

Type gene Status ☒ No status selected ☐ Deleted

Name ADIR000317-RA

Symbol

Aliases (| separated)

Description

Location 7109276 - 7112195 strand(-) Partial ☐ 5'

Ref Sequence KB673312

Owner ulrikeboehme.987772593

Created May 18, 2022 02:19 PM

Working in groups of two, please select 3-5 genes from the list to correct in Apollo, based on RNA-seq data and any other available evidence.

Unannotated Intron Junctions search

You can also find additional possible incorrect genes by using the search option **Unannotated Intron Junctions**.

You can access the search option from the **Searches** menu.

VectorBase Release 57 21 Apr 2022 Site search, e.g. AGAP004730 or "reductase" or "binding protein" A VEuPathDB Project

My Strategies **Searches** Tools My Workspace Data About Help Contact Us

Search for... expand all | collapse all Filter the searches below...

Overview of Results expand all | collapse all

- Genes
 - Annotation, curation and identifiers
 - Function prediction
 - Gene models
 - Gene Model Characteristics
 - Unannotated Intron Junctions

Take a Tour MapVEU Analyze My Data Downloads How Submit

My Organism Preferences (58 of 58) enabled

News and

Select your organism of interest.

Identify Genes based on Unannotated Intron Junctions

Reset values

Organism

Note that this search is only available for genomes to which RNA sequencing reads have been mapped.

1 selected, out of 38

[select only these](#) | [add these](#) | [clear these](#)

Anopheles dirus x ?

☐ Arthropoda

☐ Insecta

☐ Diptera

☐ Culicidae

☐ Anopheles

☒ Anopheles dirus WRAIR2 [Reference]

[select only these](#) | [add these](#) | [clear these](#)

Minimum unique reads / million (ISRPM) >=

20

Percent of Most Abundant Intron (MAI)

20 to 100

Get Answer