

Using Apollo at the i5k Workspace@NAL

Monica Poelchau, USDA-ARS NAL

February 16th, 2021



Agenda

- Basic RNA-Seq evaluation
- Basic structural changes- splitting and merging a model, adding and removing exons
- UTRs –when and how to add and adjust
- Changing translation start and stop sites, and open reading frames
- Non-canonical splice sites
- Annotating isoforms

Other resources

- An additional Apollo webinar with a worked example:
<https://www.youtube.com/watch?v=dol99KExLgY&feature=youtu.be>
- Monica Munoz-Torres from the Apollo group has a number of comprehensive tutorials:
 - <https://www.slideshare.net/MonicaMunozTorres/presentations>
 - I recommend these slides if you need more background:
 - <https://www.slideshare.net/MonicaMunozTorres/apollo-workshop-at-ksu-2015>
 - If you are new to Apollo, or need a refresher, I **highly recommend** that you review one of her presentations
- The official Apollo annotation guide:
 - <https://genomearchitect.readthedocs.io/en/latest/UsersGuide.html>
- I5k Workspace manual annotation landing page:
<https://i5k.nal.usda.gov/manual-annotation-and-apollo>
- Other manual curation tutorials:
<http://genomecuration.github.io/genometrain/d-feature-curation-crossing/>
- VEuPathDB Apollo training webinar:
<https://veupathdb.org/veupathdb/app/static-content/webinars.html#apollo>

Basic RNA-Seq evaluation

RNA-Seq tracks

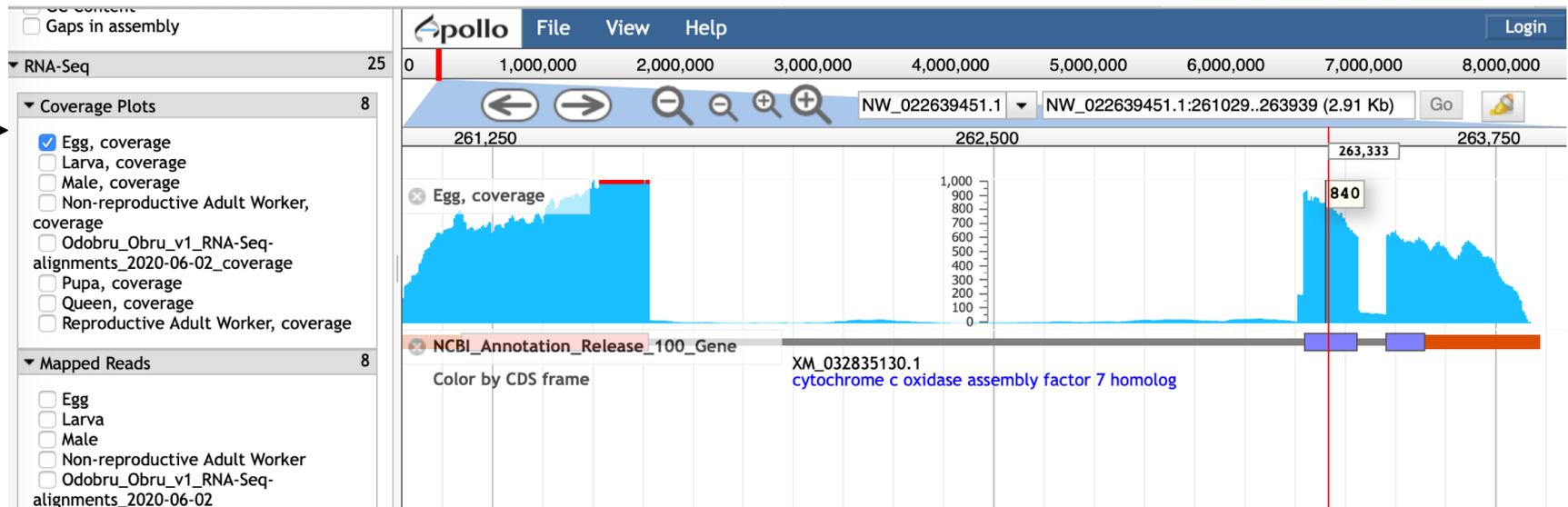
- **Coverage plots:**
Histogram of the number of mappings at each nucleotide
- **Mapped reads:**
Individual glyphs of each mapped read.
- **Junction reads:** Show where mapped reads are spliced.

The screenshot shows a genomic browser interface with the following tracks and settings:

- Available Tracks**
 - filter by text
 - 0. Reference Assembly (3)
 - GC Content
 - Gaps in assembly
 - BLAST+ Results
 - NCBI Annotation Release 100 (2)
 - NCBI_Annotation_Release_100_Gene
 - NCBI_Annotation_Release_100_Pseudogene
 - RNA-Seq (25)
 - Coverage Plots (8)
 - Egg, coverage
 - Larva, coverage
 - Male, coverage
 - Non-reproductive Adult Worker, coverage
 - Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02_coverage
 - Pupa, coverage
 - Queen, coverage
 - Reproductive Adult Worker, coverage
 - Mapped Reads (8)
 - Egg
 - Larva
 - Male
 - Non-reproductive Adult Worker
 - Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02
 - Pupa
 - Queen
 - Reproductive Adult Worker
 - Splice junctions (8)
 - Egg, junction reads
 - Larva, junction reads
 - Male, junction reads
 - Non-reproductive Adult Worker, junction reads
 - Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02_junctions
 - Pupa, junction reads
 - Queen, junction reads
 - Reproductive Adult Worker, junction reads
 - Transcriptome Assembly (1)

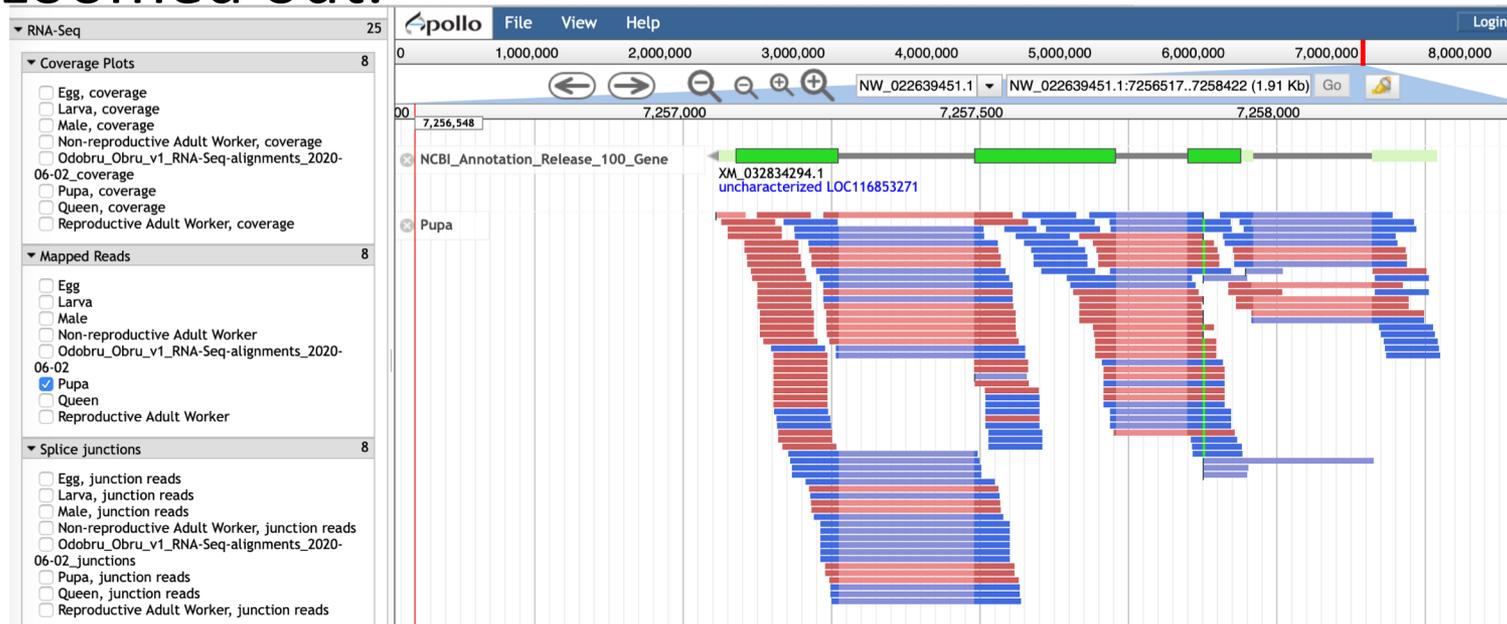
RNA-Seq tracks

- **Coverage plots:** Histogram of the number of mappings at each nucleotide; hover over the blue area to see the value



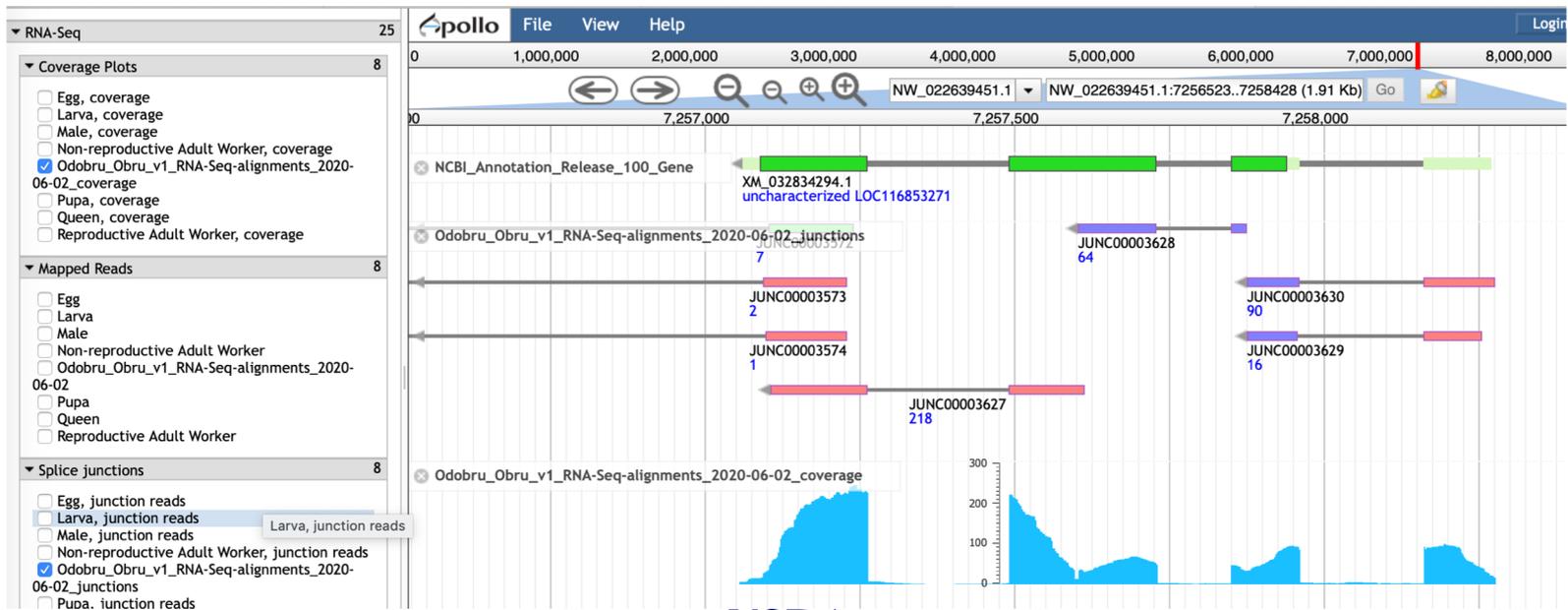
RNA-Seq tracks

- **Mapped reads:** Individual glyphs of each mapped read. Show mapped and spliced areas, and SNPs/indels. Informative, but hard to work with when zoomed out.

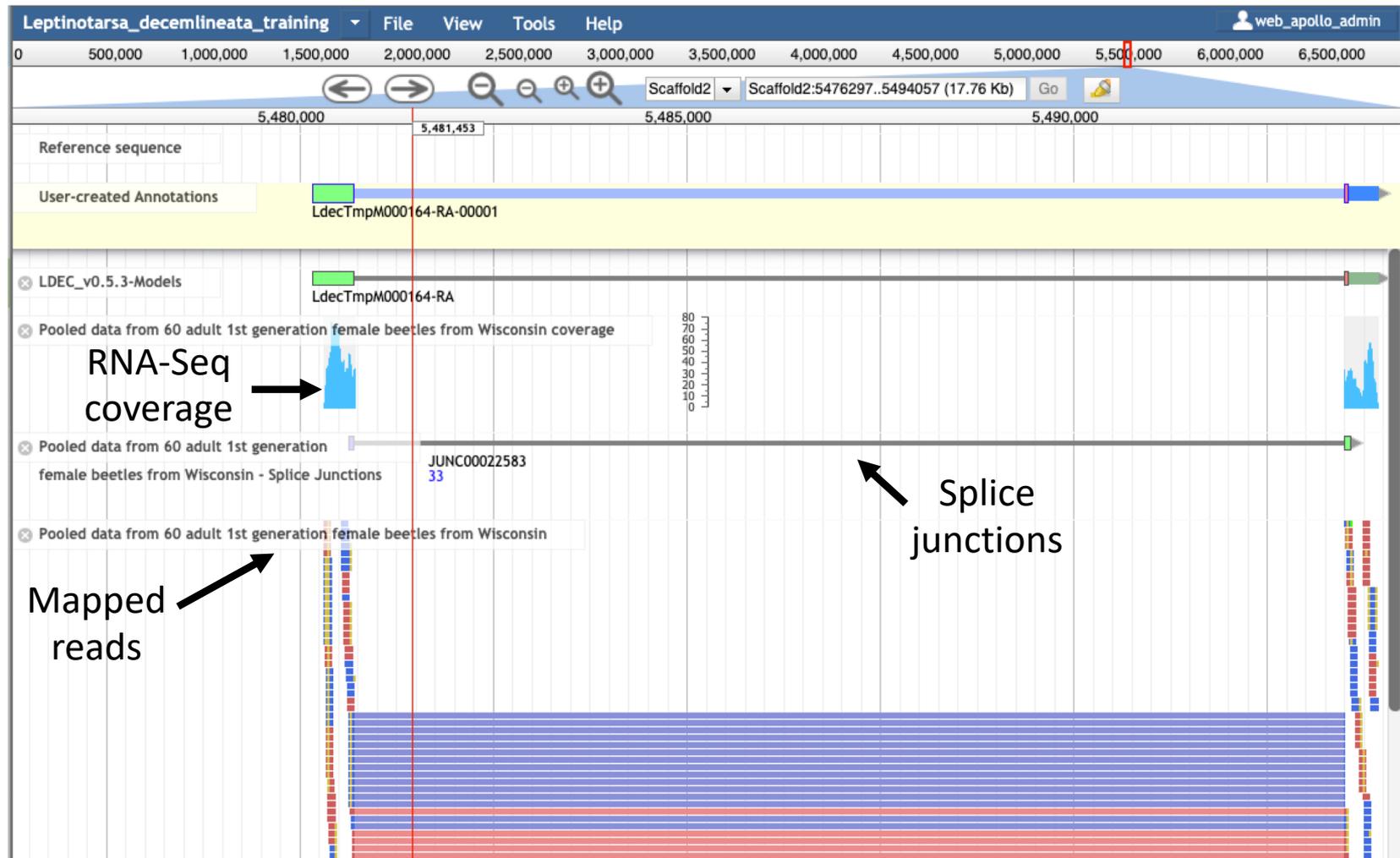


RNA-Seq tracks

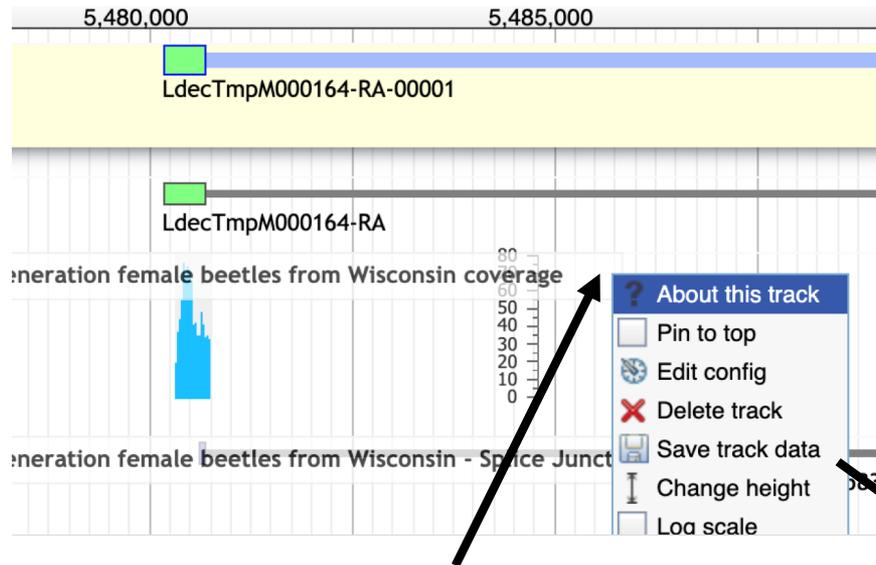
- **Junction reads:** Useful combined with coverage plots; show where mapped reads are spliced. The blue numbers show the 'score' –the number of mappings that support the splice junction.



A simple case



A simple case



Information about methods

About track: Pooled data from 60 adult 1st generation female beetles from Wisconsin coverage

Name Pooled data from 60 adult 1st generation female beetles from Wisconsin coverage

Publication status Unpublished - please follow Toronto/Ft. Lauderdale conditions of data re-use.

File provider Justin Clements and Dr. Sean Schoville UW Madison

Data provider Justin Clements

Sequencing platform Illumina Hi-seq 200 bp

Alignment method Tophat2

Data source NA

Track type JBrowse/View/Track/Wiggle/XYPlot

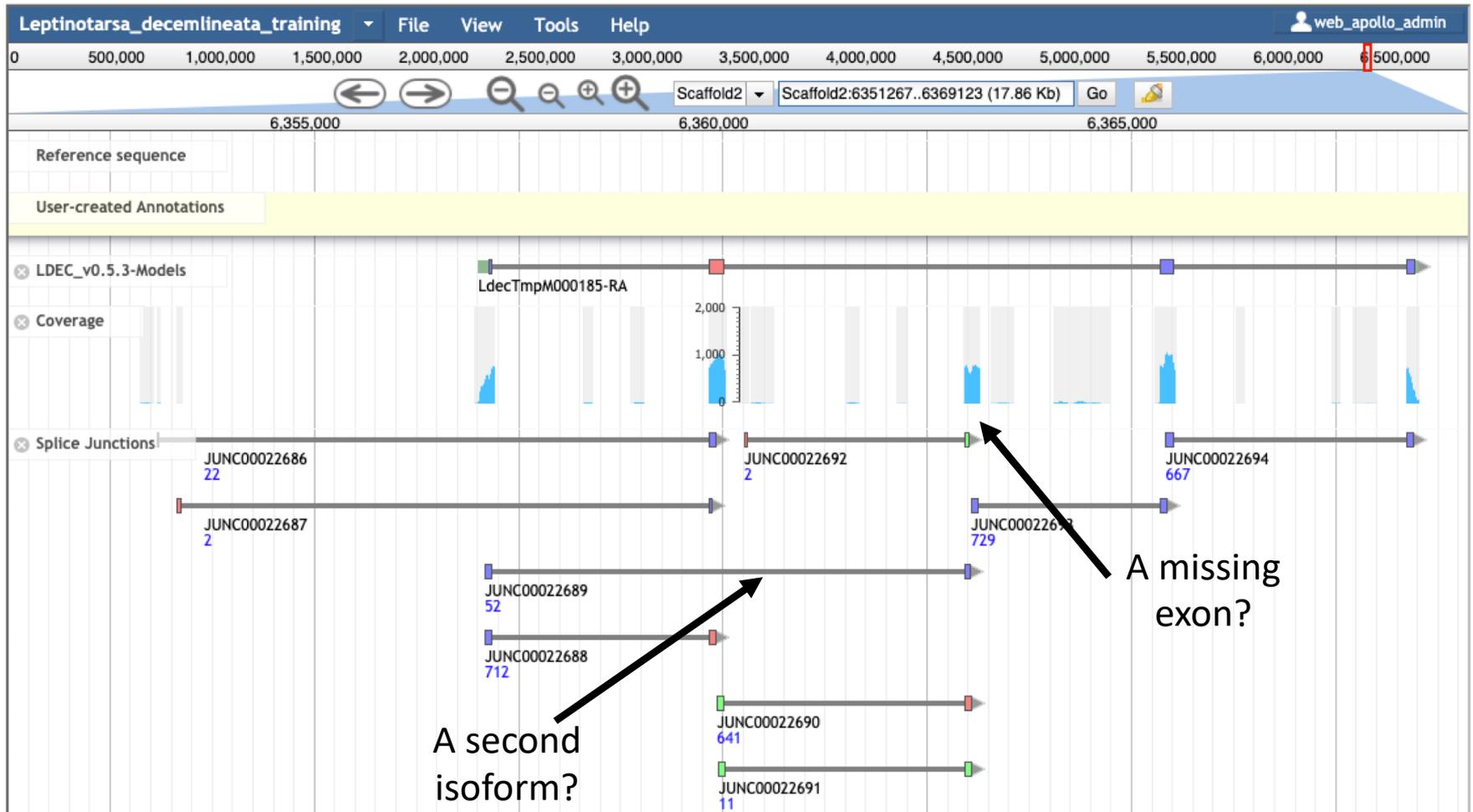
Category Transcriptome/Coverage plots (BigWig)

Stats (current reference sequence) (7)

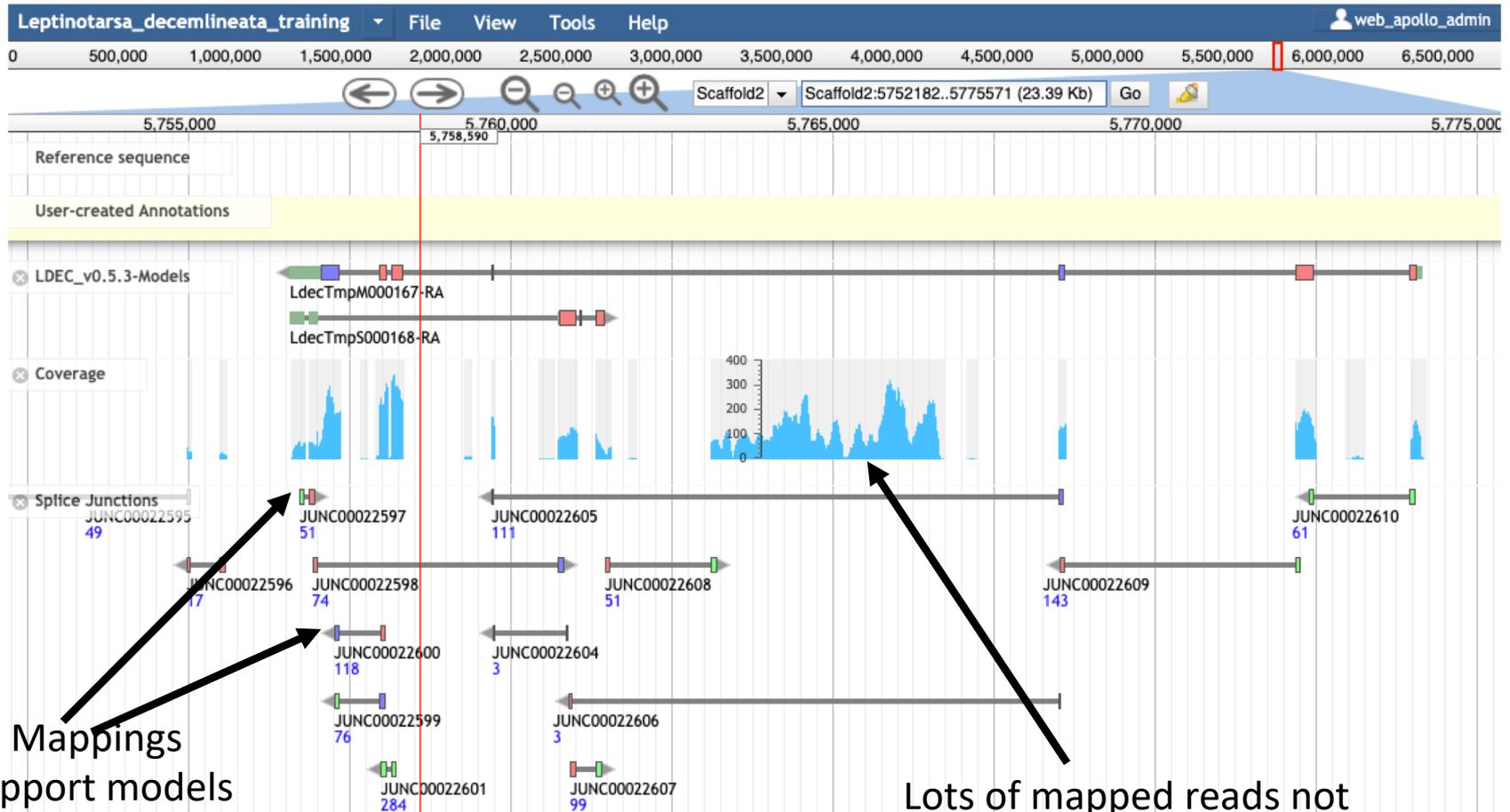
Name	Value
basesCovered	61640461
scoreMax	6163708
scoreMean	249.53225598036977
scoreMin	1

Select 'About this track' from drop-down menu

A more complex case



A really messy case



Mappings support models on forward and reverse strands

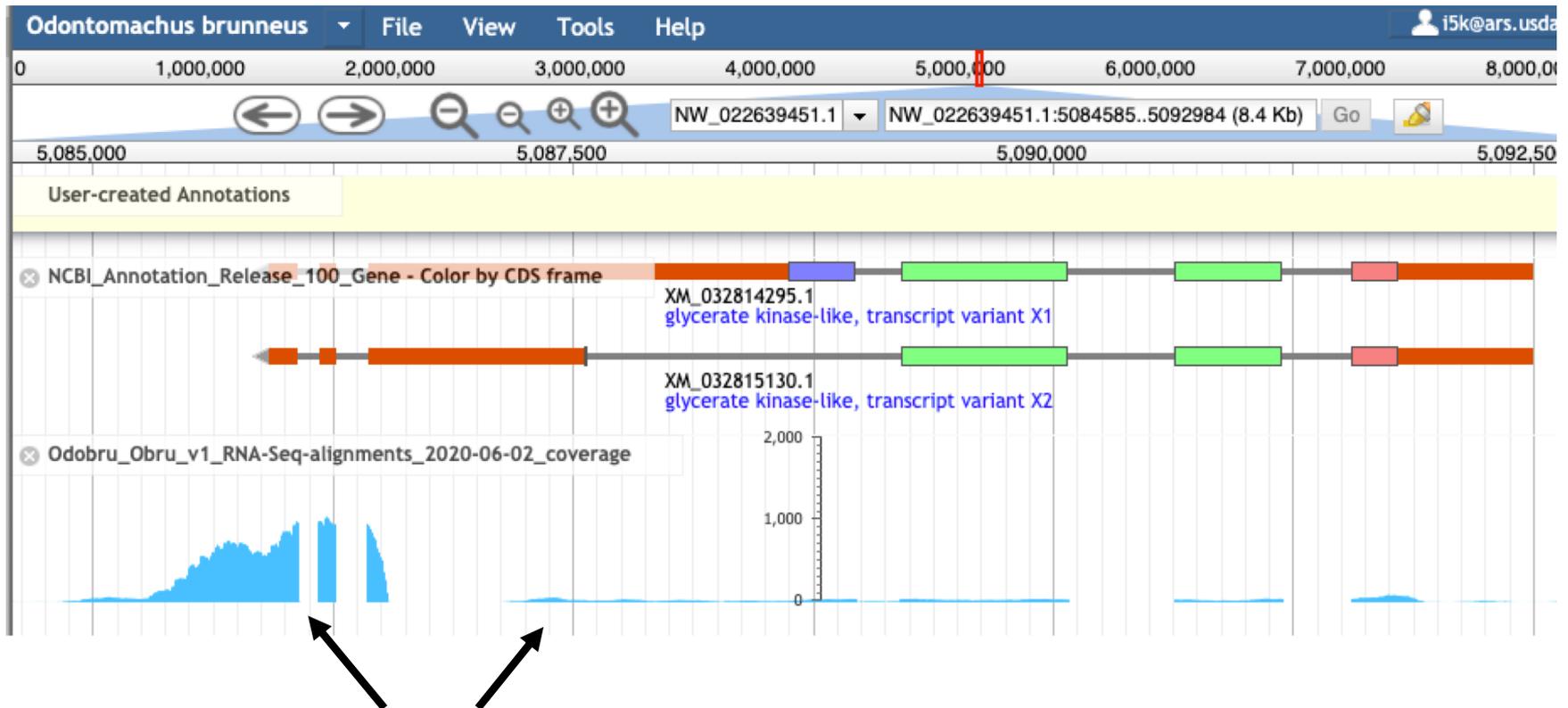
Lots of mapped reads not reflected in gene model

Basic structural changes –
splitting and merging a
model, adding and
removing exons

Annotation Example

- Glycerate kinase-like in the trap-jaw ant *Odontomachus brunneus*
- More information about the trap-jaw ant genome assembly: <https://i5k.nal.usda.gov/odontomachus-brunneus>
- *Odontomachus brunneus* Apollo URL: https://apollo.nal.usda.gov/apollo/4006447/jbrowse/index.html?loc=NW_022639451.1%3A5084490..5093717&tracks=DNA%2CAnnotations%2CNCBI_Annotation_Release_100_Gene-CBT&highlight=

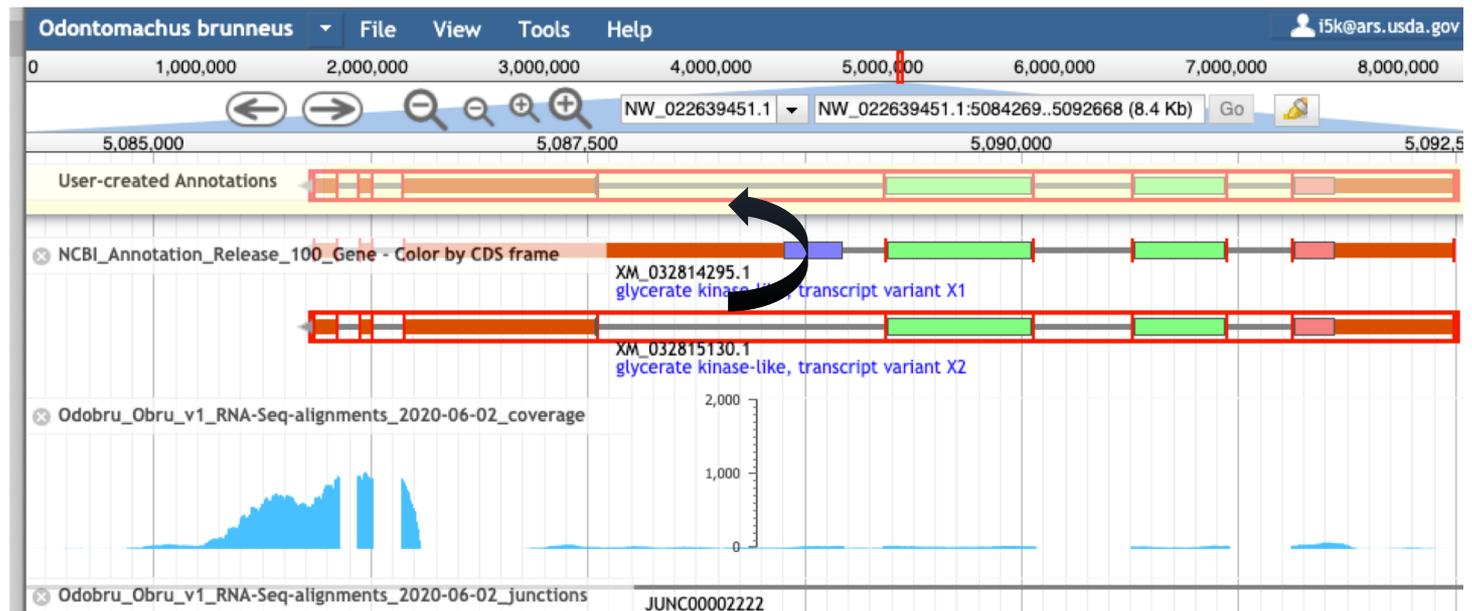
RNA-Seq evaluation



- Very different coverage between UTR and CDS
- No RNA-Seq coverage between high and low expression areas
 - 2 separate models?

Create new model in user-created annotations track

Drag evidence to UcA track (or right-click and select “create annotation”)



Split model

Select exons on which to split the model using the 'shift' key

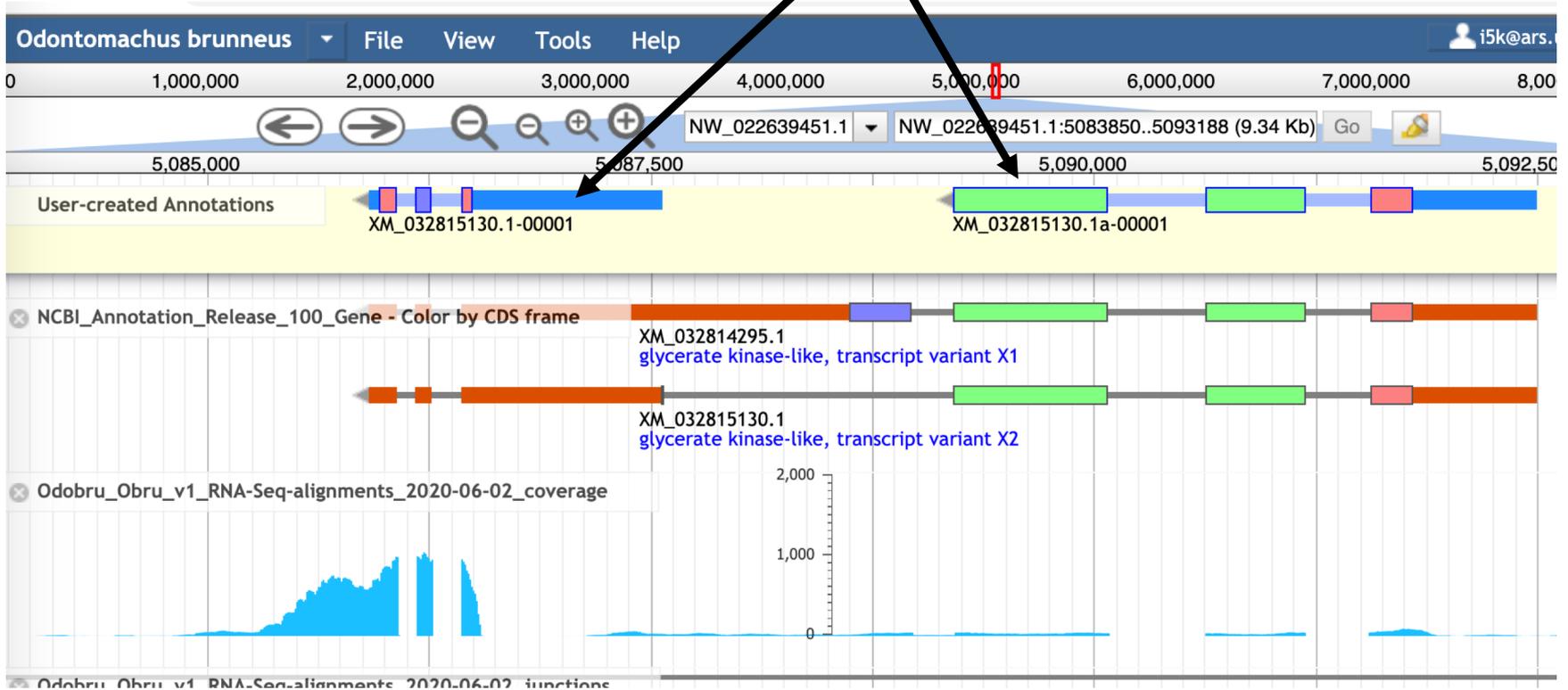
Right-click on the model while continuing to hold shift to get the drop-down menu

The screenshot displays the genome browser interface for *Odontomachus brunneus*. The top menu bar includes 'File', 'View', 'Tools', and 'Help'. The main area shows a genomic track with coordinates from 1,000,000 to 8,000,000. A specific region is highlighted with a red vertical line at approximately 5,090,000. Below the coordinate track, there are several tracks: 'User-created Annotations' showing a blue bar for XM_032815130.1-00001; 'NCBI_Annotation_Release_100_Gene - Color by CDS frame' showing two transcripts, XM_032814295.1 (glycerate kinase-like, transcript variant X1) and XM_032815130.1 (glycerate kinase-like, transcript variant X2); and 'Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02_coverage' showing a blue histogram. A context menu is open over the transcript model, listing options such as 'Get Sequence', 'Get GFF3', 'Zoom to Base Level', 'View in Annotator Panel', 'Edit Information (alt-click)', 'Change annotation type', 'Associate Transcript to Gene', 'Dissociate Transcript from Gene', 'Delete', 'Merge', 'Split', 'Duplicate', 'Make Intron', 'Move to Opposite Strand', 'Set Translation Start', and 'Set Translation End'. The 'Split' option is highlighted in blue. Red boxes highlight two exons in the 'User-created Annotations' track, and black arrows point from the text above to these exons and the 'Split' option in the menu.

Select 'split'

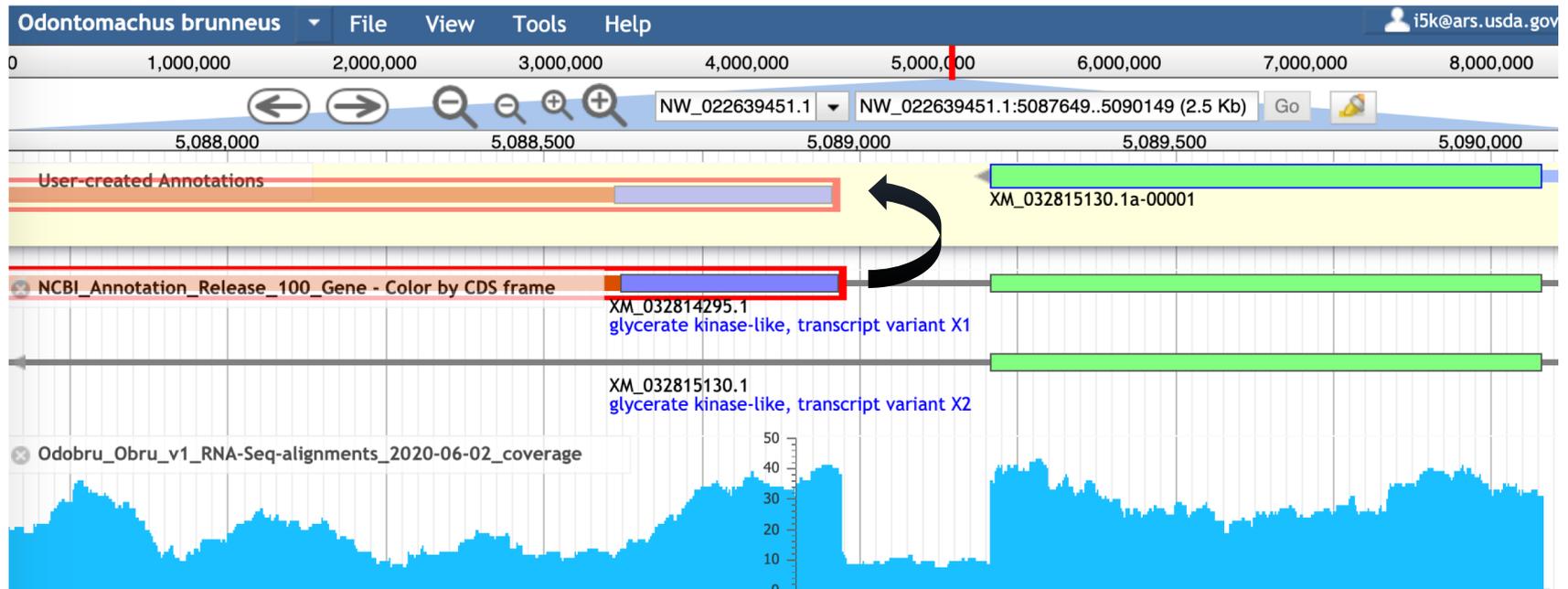
Split model

You now have 2 models! Let's start fixing the model on the right – it needs a 3' exon.



Add an exon

Zoom in, select the missing exon,
drag up to Uca track



Merge exons

Shift-select both exons, shift-right click, then select 'merge' from the dropdown menu

The screenshot displays the Genes and Genomes browser interface for *Odontomachus brunneus*. The top navigation bar includes 'File', 'View', 'Tools', and 'Help'. The main view shows a genomic track with coordinates from 1,000,000 to 8,000,000. A zoomed-in view shows coordinates from 5,088,000 to 5,090,000. A red box highlights two exons, XM_032814295.1-00001 and XM_032814295.1, which are selected. A context menu is open over the selected exons, with the 'Merge' option highlighted. The menu options include: Get Sequence, Get GFF3, Zoom to Base Level, View in Annotator Panel, Edit Information (alt-click), Change annotation type, Associate Transcript to Gene, Dissociate Transcript from Gene, Delete, Merge, Split, Duplicate, Make Intron, Move to Opposite Strand, Set Translation Start, and Set Translation End. The track also shows NCBI Annotation Release 100 Gene data and RNA-Seq coverage for Odobru_Obru_v1.

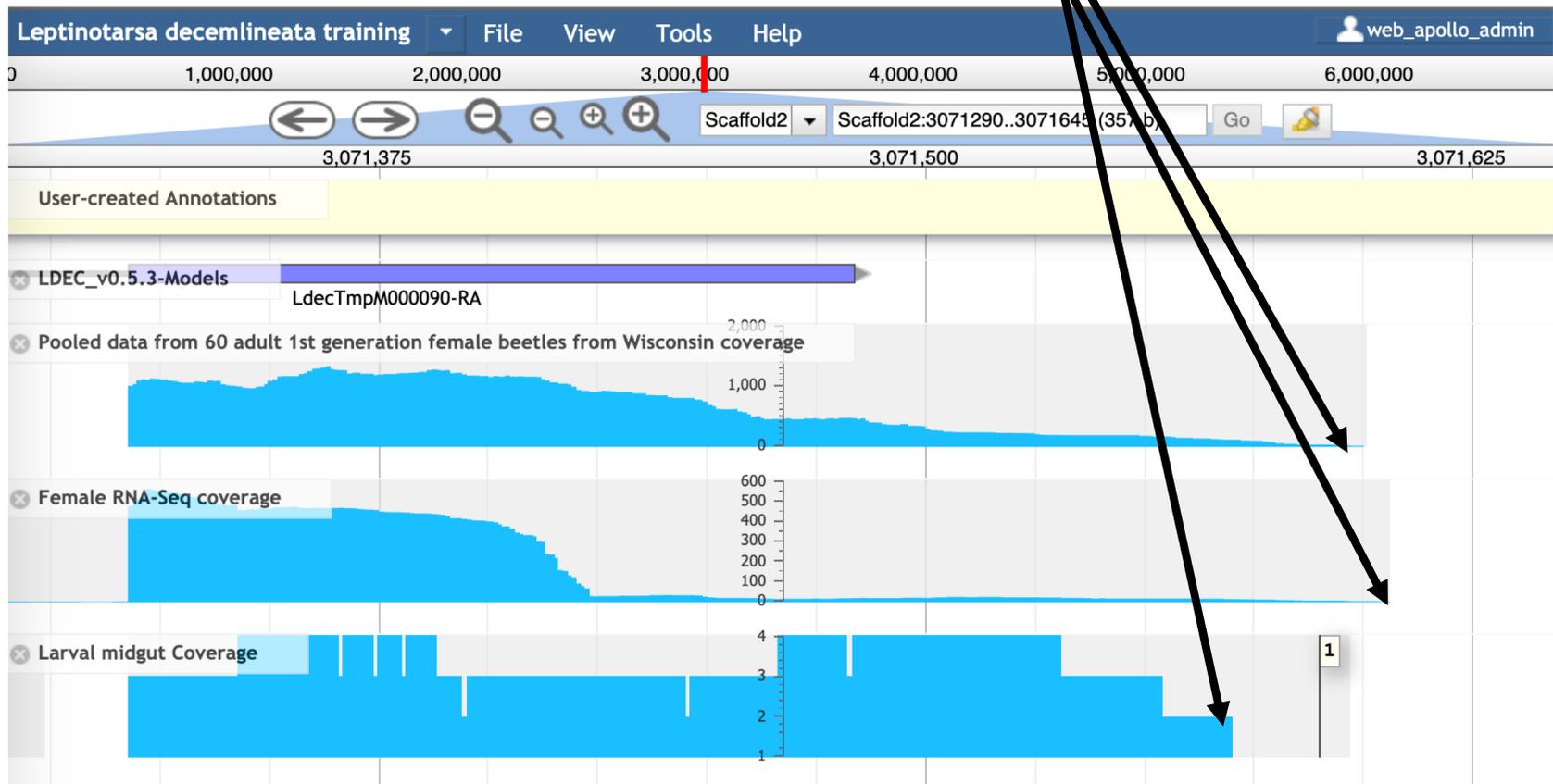
UTRs – how and when to
add or adjust

Adding or adjusting UTR boundaries

- When should you add or change UTRs?
 - Only if you have RNA-Seq evidence with sufficient coverage relative to the rest of the model
 - Adding or changing UTRs is helpful, but not necessary if you're only interested in the protein sequence
 - Deciding where the UTR ends is usually a judgement call
- Apollo tools for gene boundary changes:
 - Manual edge-matching to available evidence
 - Automated edge-matching to available evidence

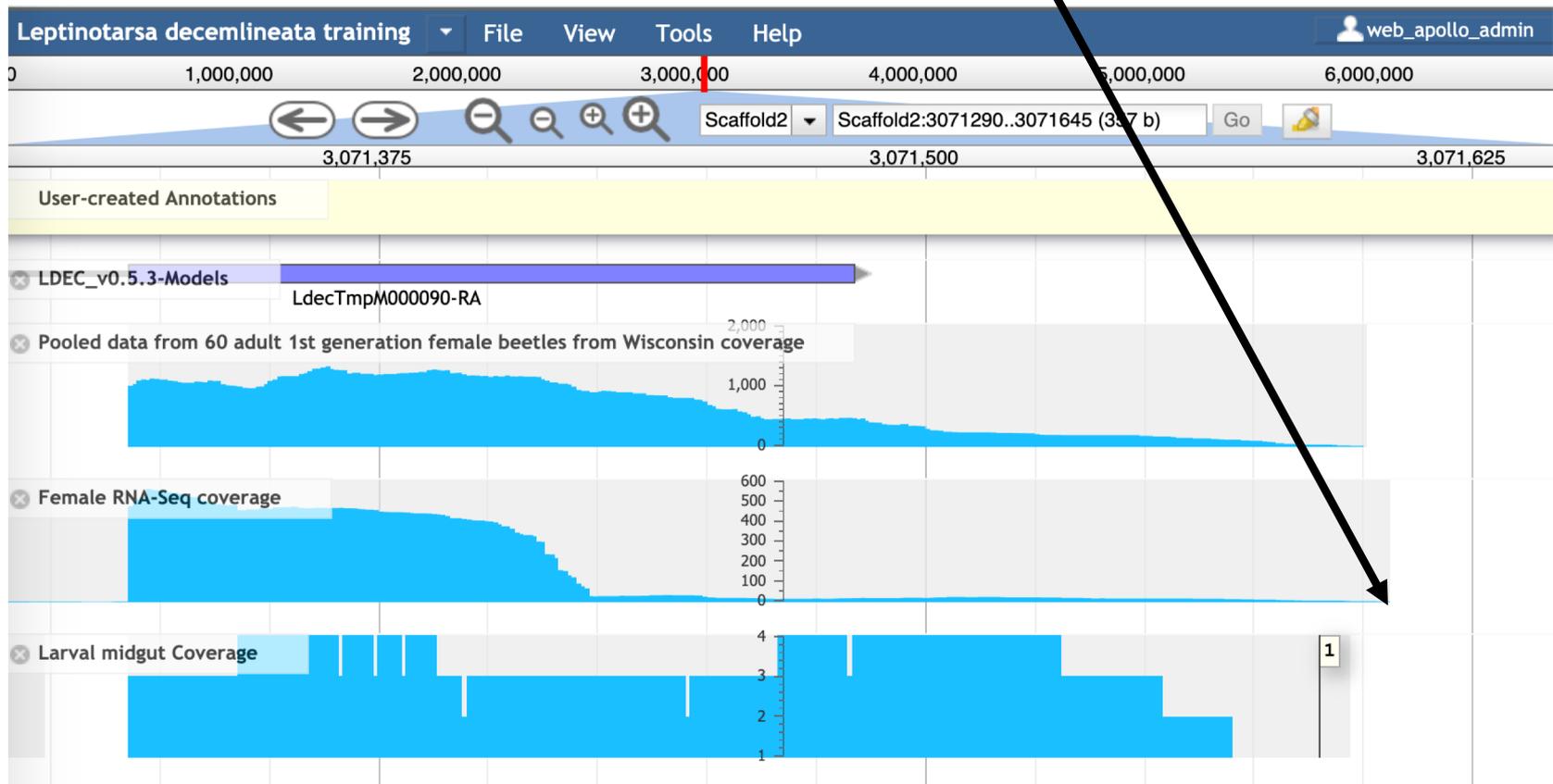
Adjusting gene boundaries

RNA-Seq evidence ends in different places for each track – how do you decide?

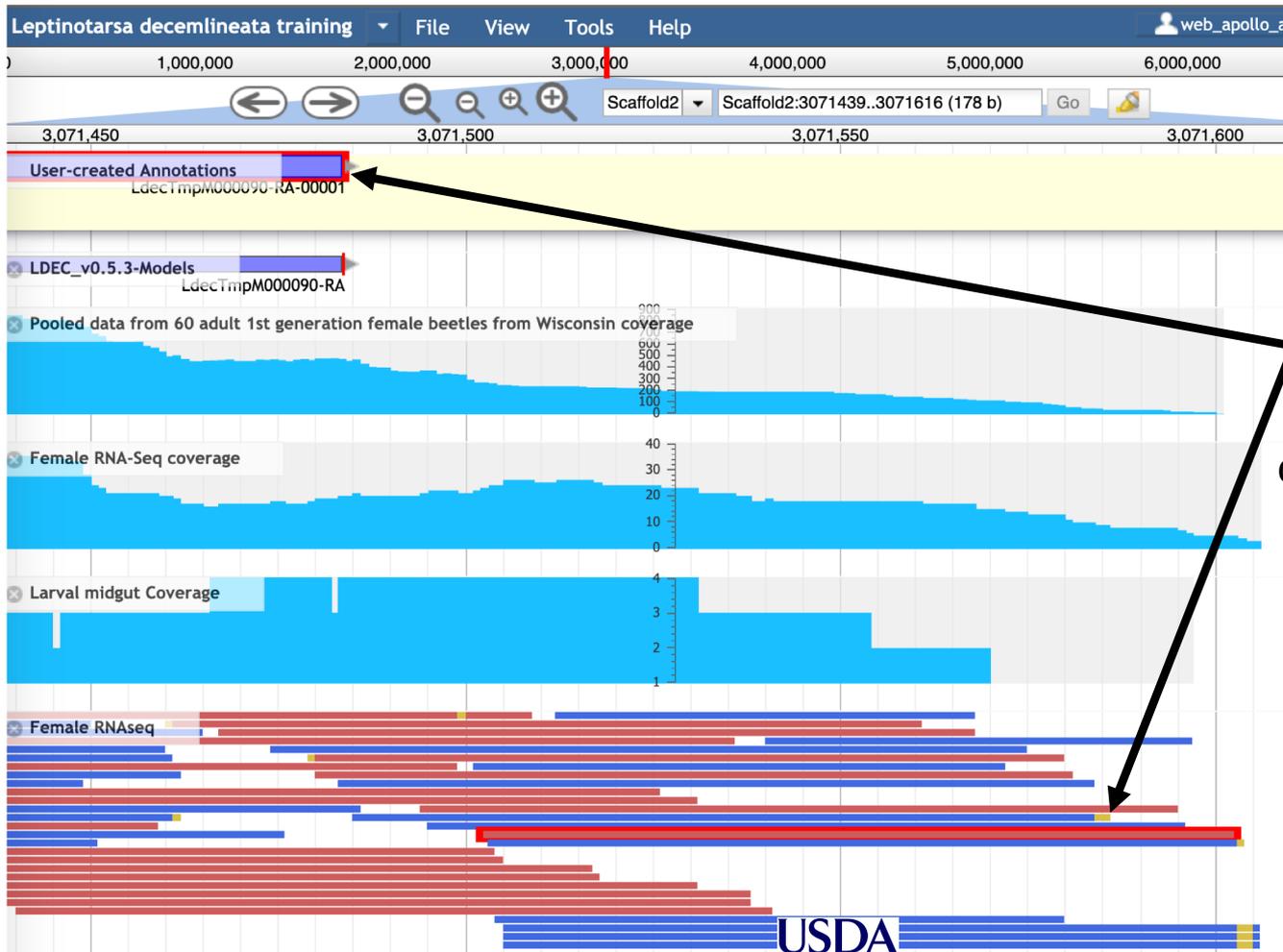


Adjusting gene boundaries

Pick the longest boundary available, and note which track you used in the 'Comments' section



Adjusting gene boundaries



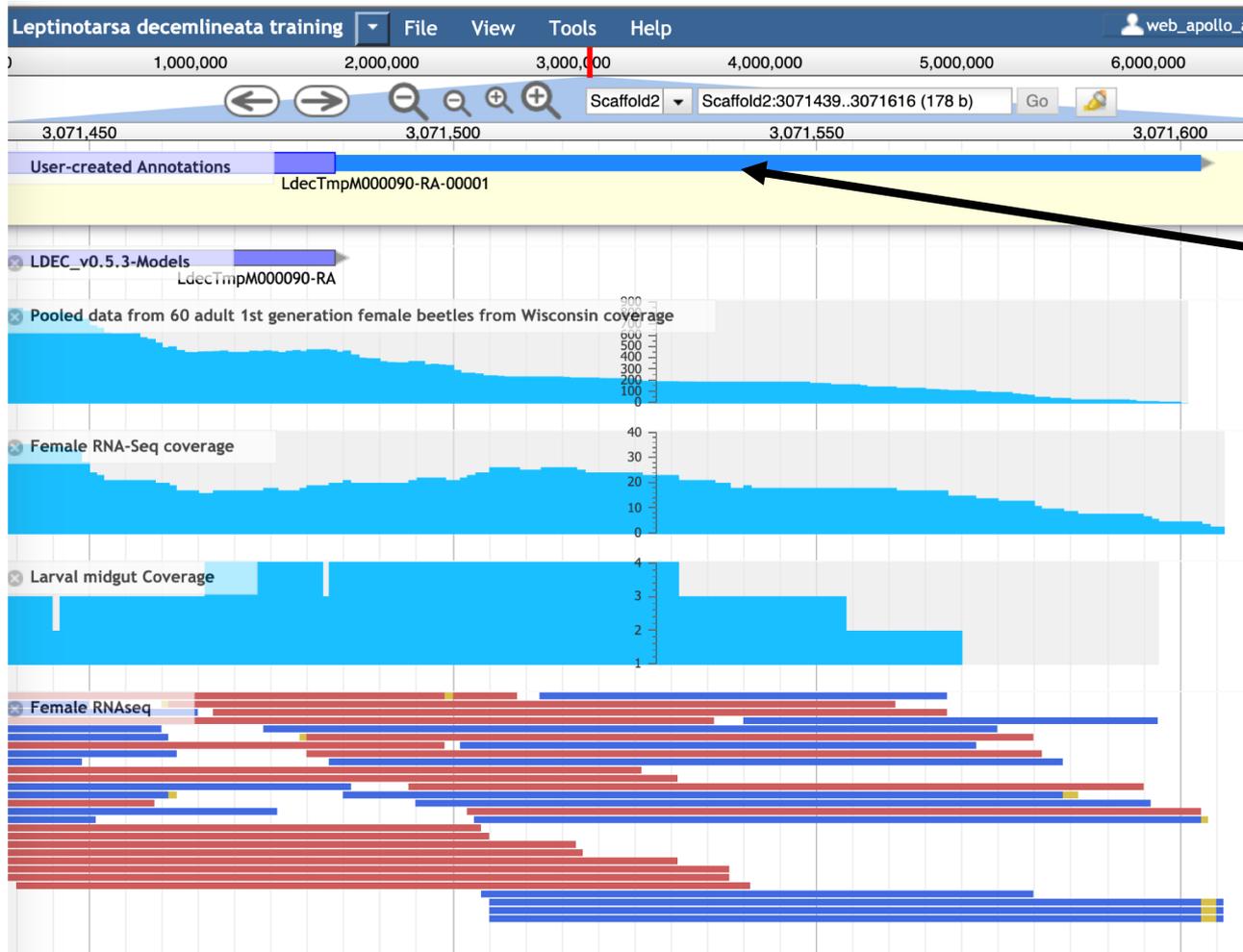
One way to change the boundary: find a mapped read on the same strand as the model; hold shift and click on the read and the model to highlight them both

Adjusting gene boundaries

The screenshot displays the Apollo genome browser interface for *Leptinotarsa decemlineata*. The top navigation bar includes 'File', 'View', 'Tools', and 'Help'. The main view shows a genomic track for Scaffold2, with coordinates ranging from 1,000,000 to 6,000,000. A specific region is highlighted, with coordinates 3,071,450 to 3,071,600. A context menu is open over a user-created annotation track, listing various actions such as 'Get Sequence', 'Zoom to Base Level', 'Delete', 'Merge', 'Split', 'Duplicate', 'Make Intron', 'Move to Opposite Strand', 'Set Translation Start', 'Set Translation End', 'Set Longest ORF', 'Set Readthrough Stop Codon', 'Set as 5' end', 'Set as 3' End', 'Set both Ends', 'Set to Downstream Splice Donor', 'Set to Upstream Splice Donor', 'Set to Downstream Splice Acceptor', 'Set to Upstream Splice Acceptor', 'Undo', 'Redo', and 'Show History'. The 'Set as 3' End' option is highlighted in blue. A black arrow points from the text on the right to this option.

Right-click on model in user-created annotations track, and select 'Set as 3' end' from the drop-down menu

Adjusting gene boundaries



New UTR is there!

Adjusting gene boundaries

The screenshot displays the Apollo genome browser interface for *Leptinotarsa decemlineata*. Two 'Information Editor' windows are open, showing the 'Attributes' section with 'Tag' and 'Value' fields, and 'Add' and 'Delete' buttons. The right window also shows 'PubMed IDs', 'Gene Ontology IDs', and 'Comments' sections. A comment is added to the 'Comments' field: 'Added 3' UTR based on Female RNA-Seq'. An arrow points from the text 'Add comment explaining UTR addition' to this comment.

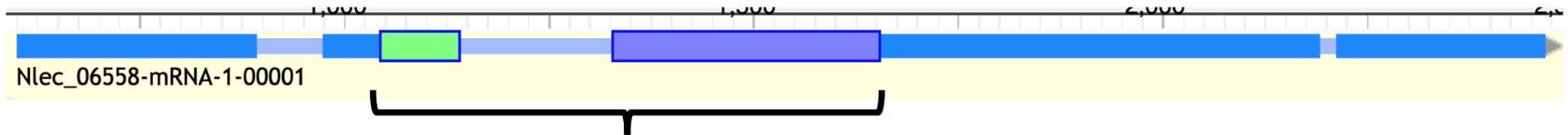
Add comment explaining UTR addition

Starts, stops, open
reading frames

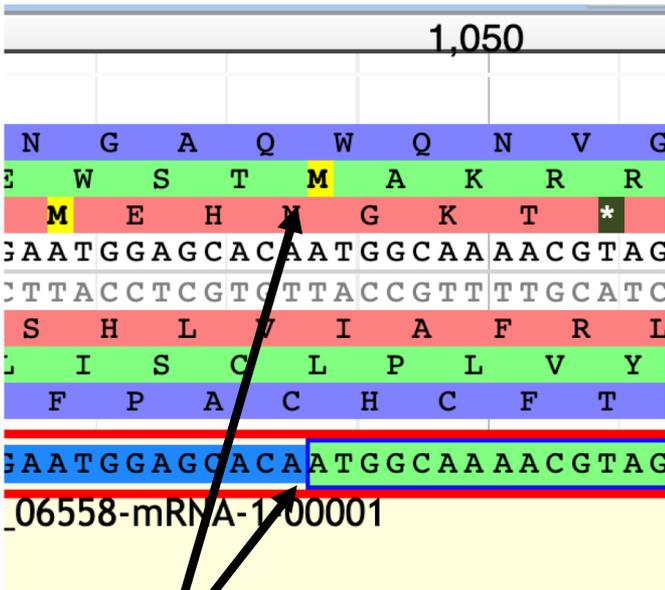
Setting the sequence start, stop, and open reading frame (ORF)

- Apollo will automatically calculate the longest possible ORF that includes canonical 'Start' and 'Stop' signals
(<https://genomearchitect.readthedocs.io/en/latest/UsersGuide.html#start-and-stop-sites>)
- However, in some fringe cases, you will need to double-check
- You can change a model's start and stop sites if needed

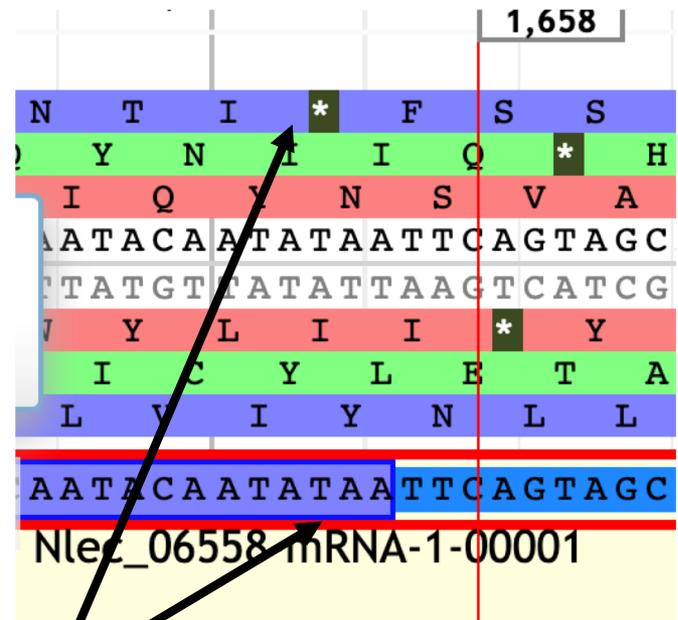
Starts, stops, ORFs



Open reading frame (ORF): translated region

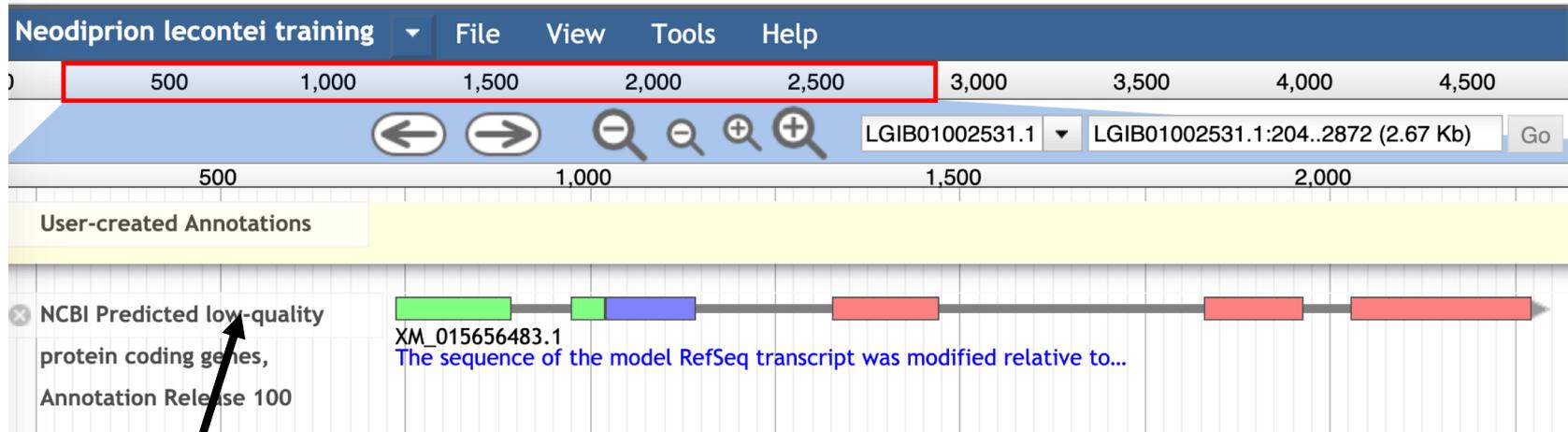


Translation start at Methionine



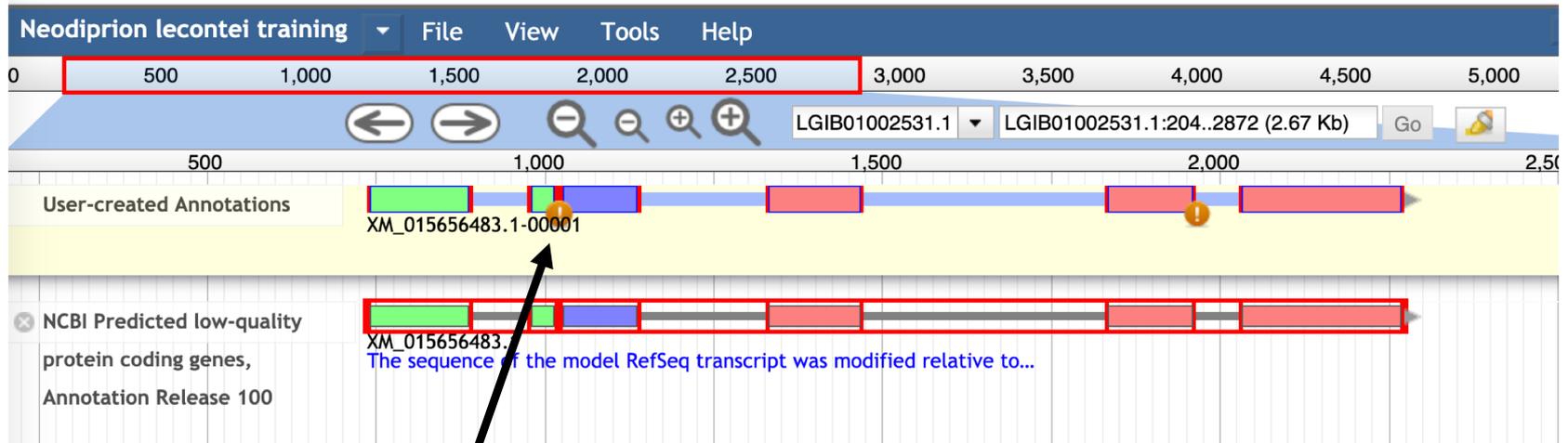
Translation stop at stop codon

Starts, stops, ORFs



This is a 'low quality' protein coding gene from NCBI – it will likely show some problems in Apollo

Starts, stops, ORFs



We can see a non-canonical splice site in the Uca (more on that later). Let's zoom to the start of the model.

Starts, stops, ORFs

Neodiprion lecontei training | File | View | Tools | Help | web_apollo_admin

0 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000

← → 🔍 - +

LGIB01002531.1 | LGIB01002531.1:703..832 (131 b) | Go

725 750 775 800 825

Reference sequence

V R L N Y Y P R Y L R Y G C Q Y C D T Q I E A R R A V * G C * N I D C L * F M E N V

T P E L I L S T L S Q I R M P I L R Y A N * G P E S C I R L L E Y * L P L I H G K C

Y A * I D I I H A I S D T D A N T A I R K L R P G E L Y K V A R I L T A S D S W K M L

GTACGCCTGAATTGATATTATCCACGCTATCTCAGATACGGATGCCAACTACTGCGATACGCAATTGAGGCCCGGAGAGCTGTATAAGGTTGCTAGAAATATTGACTGCCTCTGATTCATGGAAATGT?

CATCGGACTTAACATAAATAGGTGCGATAGAGTCTATGCCTACGGTTATGACGCTATGCCCTAACTCCGGGCCTCTCGACATATTCCAACGATCTTATAACTGACGGAGACTAAGTACCTTTTACAT?

L V G S N I N D V S D * I R I G I S R Y F Q P G S L Q I L N S S Y Q S G R I * P F H

Y A Q I S I I W A I E S V S A L V A I T L N L G P S S Y L T A L I N V A E S E H F I

T R R F Q Y * G R * R L Y P H W Y Q S V C I S A R L A T Y P Q * F I S Q R Q N M S F T

User-created Annotations

ATACGGATGCCAACTACTGCGATACGCAAAATTGAGGCCCGGAGAGCTGTATAAGGTTGCTAGAAATATTGACTGCCTCTGATTCATGGAAATGT?

XM_015656483.1-00001

NCBI Predicted low-quality protein coding genes, Annotation Release 100

XM_015656483.1

The sequence of the model RefSeq transcript was modified relative to...

Apollo shows this model in the green reading frame – however, we can see a stop pretty early on in the genome sequence - but that’s not reflected in the Apollo model! It looks like the pink reading frame doesn’t have stops.

Starts, stops, ORFs

Sequence 750

```
>88802400-725a-4c8b-9ec9-1943fde9749c (sequence:exon) 10 residues [LGIB01002531.1:737-893 + strand]
[peptide]
IRMPILRYAN
```

Sure enough, the protein sequence is suspiciously short

Peptide sequence
 cDNA sequence
 CDS sequence
 Genomic sequence
 Genomic sequence +/- bases

Starts, stops, ORFs

The screenshot shows a genome browser interface for 'Neodiprion lecontei training'. The top navigation bar includes 'File', 'View', 'Tools', and 'Help'. A scale bar at the top indicates positions from 0 to 3,000. The main area displays a reference sequence with multiple reading frames highlighted in different colors. A pink reading frame is selected, and a context menu is open over it. The menu options include: 'Get Sequence', 'Get GFF3', 'Zoom to Base Level', 'View in Annotator Panel', 'Edit Information (alt-click)', 'Change annotation type', 'Associate Transcript to Gene', 'Dissociate Transcript from Gene', 'Delete', 'Merge', 'Split', 'Duplicate', 'Make Intron', 'Move to Opposite Strand', 'Set Translation Start', 'Set Translation End', 'Set Longest ORF', 'Set Readthrough Stop Codon', and 'Set as 5' end'. A red box highlights the 3rd nucleotide (U) in the UCA codon of the pink reading frame. A black arrow points from the text on the right to this nucleotide, and another black arrow points from the text to the 'Set Translation Start' menu option.

Let's set the translation start in the pink reading frame – click on the 3rd nucleotide in the Uca, right-click, and select 'Set Translation Start'

Starts, stops, ORFs

Neodiprion lecontei training File View Tools Help web_apollo_admin

0 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500 5,000 5,500 6,000

725 750 775 800 825

Reference sequence

User-created Annotations

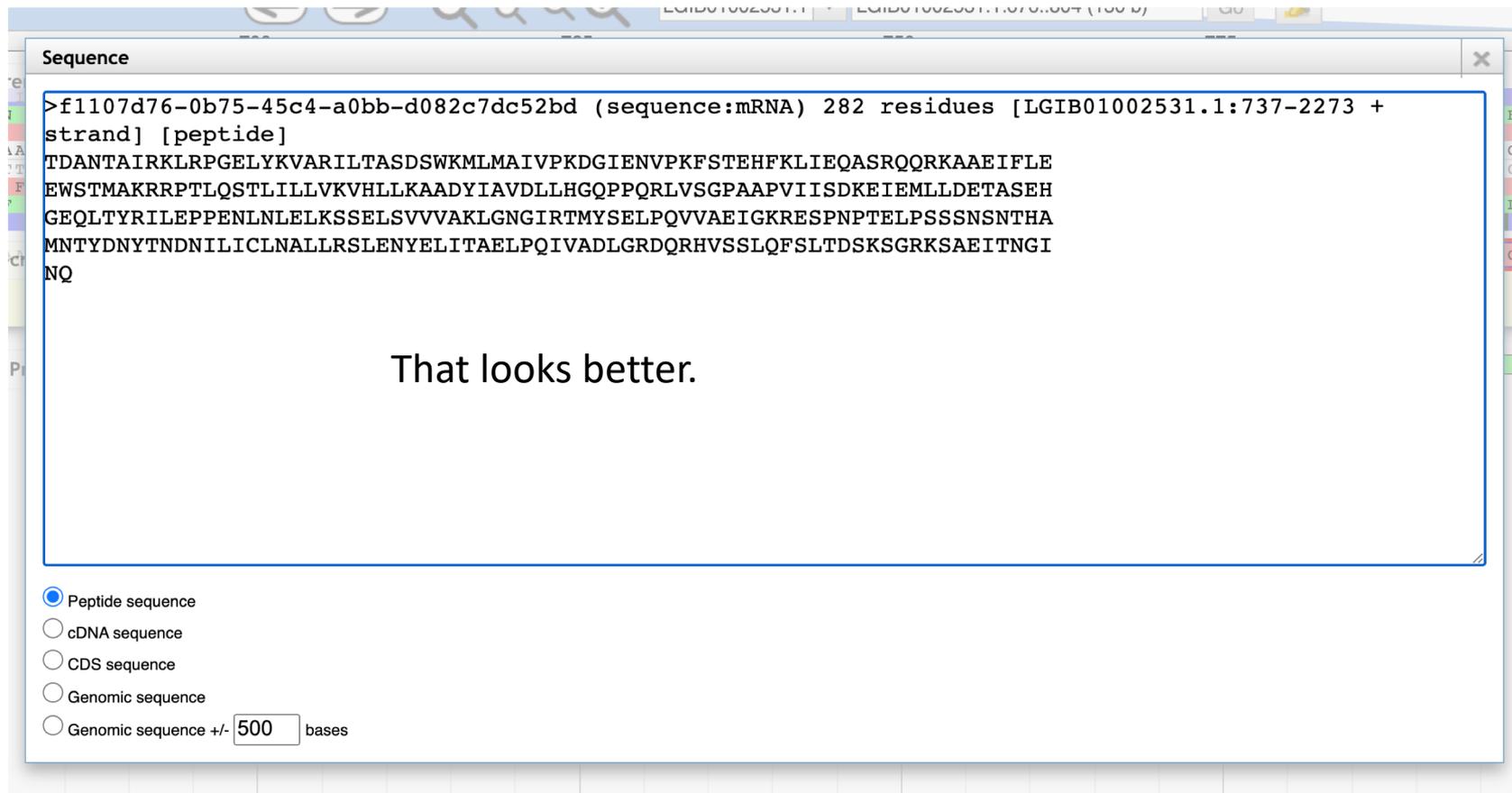
NCBI Predicted low-quality protein coding genes, Annotation Release 100

XM_015656483.1-00001

XM_015656483.1
The sequence of the model RefSeq transcript was modified relative to...

We're in the pink reading frame now – let's check the protein sequence

Starts, stops, ORFs



The screenshot shows a software window titled "Sequence" with a close button in the top right corner. The main content area contains the following text:

```
>f1107d76-0b75-45c4-a0bb-d082c7dc52bd (sequence:mRNA) 282 residues [LGIB01002531.1:737-2273 + strand] [peptide]  
TDANTAIRKLRPGELYKVARILTASDSWKMLMAIVPKDGIENVPKFSTEHFKLIEQASRQQRKAAEIFLE  
EWS'TMAKRRPTLQSTLILLVKVHLLKAADYIAVDLLHGQPPQRLVSGPAAPVVISDKEIEMLLDETASEH  
GEQLTYRILEPPENLNLELKSSELSVVAKLNGGIRTMYSERPQVVAEIGKRESPNPTELPSSSNSNTHA  
MNTYDNYTNDNILICLNALLRSLENYELITAE LPQIVADLGRDQRHVSSLQFSLTDSKSGRKS AEITNGI  
NQ
```

Below the sequence, there is a radio button menu with the following options:

- Peptide sequence
- cDNA sequence
- CDS sequence
- Genomic sequence
- Genomic sequence +/- bases

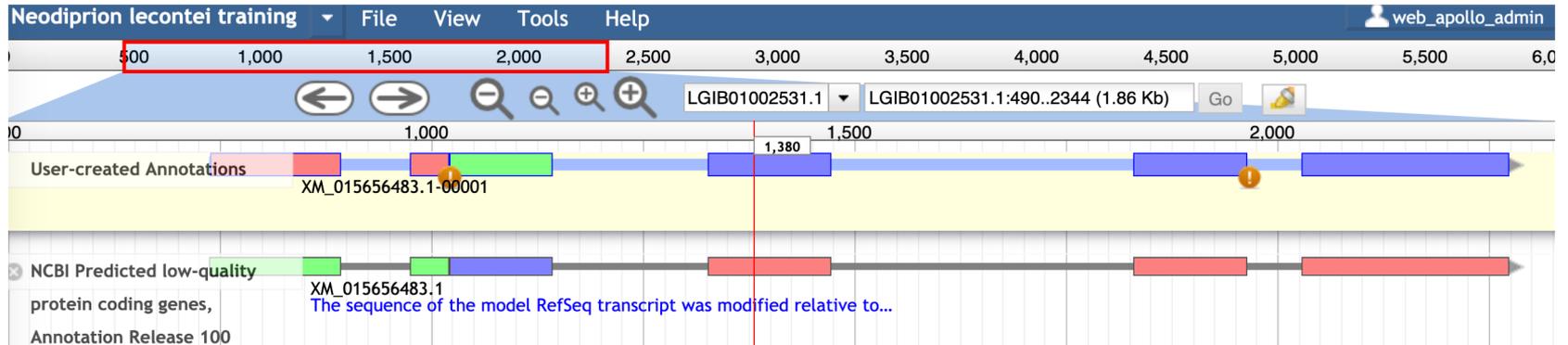
In the center of the main content area, the text "That looks better." is displayed.

Starts, stops, ORFs

The screenshot shows a genome browser interface for 'Neodiprion lecontei training'. The top menu bar includes 'File', 'View', 'Tools', and 'Help'. A scale bar at the top indicates positions from 0 to 3,500. Below the scale, a sequence alignment is shown with a red box highlighting a region between approximately 500 and 2,000. A context menu is open over the alignment, listing various actions such as 'Get Sequence', 'Zoom to Base Level', and 'Set Longest ORF'. The 'Set Longest ORF' option is highlighted in blue. The menu also includes options for 'Set Translation Start', 'Set Translation End', and 'Set Readthrough Stop Codon'. The background shows 'User-created Annotations' and 'NCBI Predicted low-quality protein coding genes' with various colored bars representing different features.

Sometimes it can be hard to tell what the protein sequence should be – in that case you can right-click and select ‘Set Longest ORF’

Starts, stops, ORFs



This also fixed the reading frame.

```
Sequence
>9190e15d-dcee-45c8-9236-5d7babfca448 (sequence:mRNA) 282 residues [LGIB01002531 strand] [peptide]
TDANTAIRKLRPGELYKVARILTASDSWKMLMAIVPKDGIENVPKFSTEHFKLIEQASRQQRKAAEIFLE
EWSTMAKRRPTLQSTLILLVKVHLLKAADYIAVDLLHGQPPQRLVSGPAAPVIISDKEIEMLLDETASEH
GEQLTYRILEPPENLNLELKSSSELSVVVAKLGNIGRTMYSELPQVVAEIGKRESPNPTELPSSSNSNTHA
MNTYDNYTNDNILICLNALLRSLENYELITAELPQIVADLGRDQRHVSSLQFSLTDSKSGRKS AEITNGI
NQ
```

Starts, stops, ORFs

The screenshot shows a genome browser interface for the Neodiprion lecontei training dataset. The top menu includes File, View, Tools, and Help. A scale bar at the top indicates positions from 0 to 4,000. The reference sequence is displayed in a multi-line format with amino acid translations above the nucleotide sequence. A context menu is open over a specific nucleotide (G) at position 2,229. The menu options include: Get Sequence, Get GFF3, Zoom to Base Level, View in Annotator Panel, Edit Information (alt-click), Change annotation type, Associate Transcript to Gene, Dissociate Transcript from Gene, Delete, Merge, Split, Duplicate, Make Intron, Move to Opposite Strand, Set Translation Start, Set Translation End (highlighted), and Set Largest ORF. The 'User-created Annotations' section shows a yellow bar for XM_015656483.1-00001. The 'NCBI Predicted low-quality' section shows a red bar for XM_015656483.1, with the description 'The sequence of the model RefSeq protein coding genes, Annotation Release 100'.

Similarly, if you have evidence to change the translation end, you can click on the corresponding nucleotide, right-click, and select 'Set Translation End'

Starts, stops, ORFs

Neodiprion lecontei training File View Tools Help

0 500 1,000 1,500 2,000 2,500 3,000 3,500 4,000 4,500

← → 🔍 - 🔍 + 🔍 + LGIB01002531.1 LGIB01002531.1:2182..2311 (131 b)

2,200 2,225 2,250 2,275

Reference sequence

D Q R H V S S L Q F S L T D S K S G R K S A E I T N G I N Q * V S

S T P R * F F A I Q S H R F K I W K K I C R N Y * R Y * P I G V

I N A T L V L C N S V S P I Q N L E E N L P K L L T V L T N R C L

GATCAACGCCACGTTAGTTCTTTGCAATTCAGTCTCACCGATTCAAAATCTGGAAGAAAATCTGCCGAAATTACTAACGGTATTAACCAATAGGTGTC

CTAGTTGCGGTGCAATCAA GAAACGTTAAGTCAGAGTGGCTAAGTTTTAGACCTTCTTTTAGACGGCTTAAATGATTGCCATAATTGGTTATCCACAGI

P D V G R * N K A I * D * R N L I Q F I Q R F * * L R Y * G I P T

I L A V N T R Q L E T E G I * F R S S F R G F N S V T N V L L H I

S * R W T L E K C N L R V S E F D P L F D A S I V L P I L W Y T D

User-created Annotations

XM_015656483.1-00001

NCBI Predicted low-quality

protein coding genes, Annotation Release 100

XM_015656483.1
The sequence of the model RefSeq transcript was modified relative to...

Now the sequence after the translation end is 3' UTR.

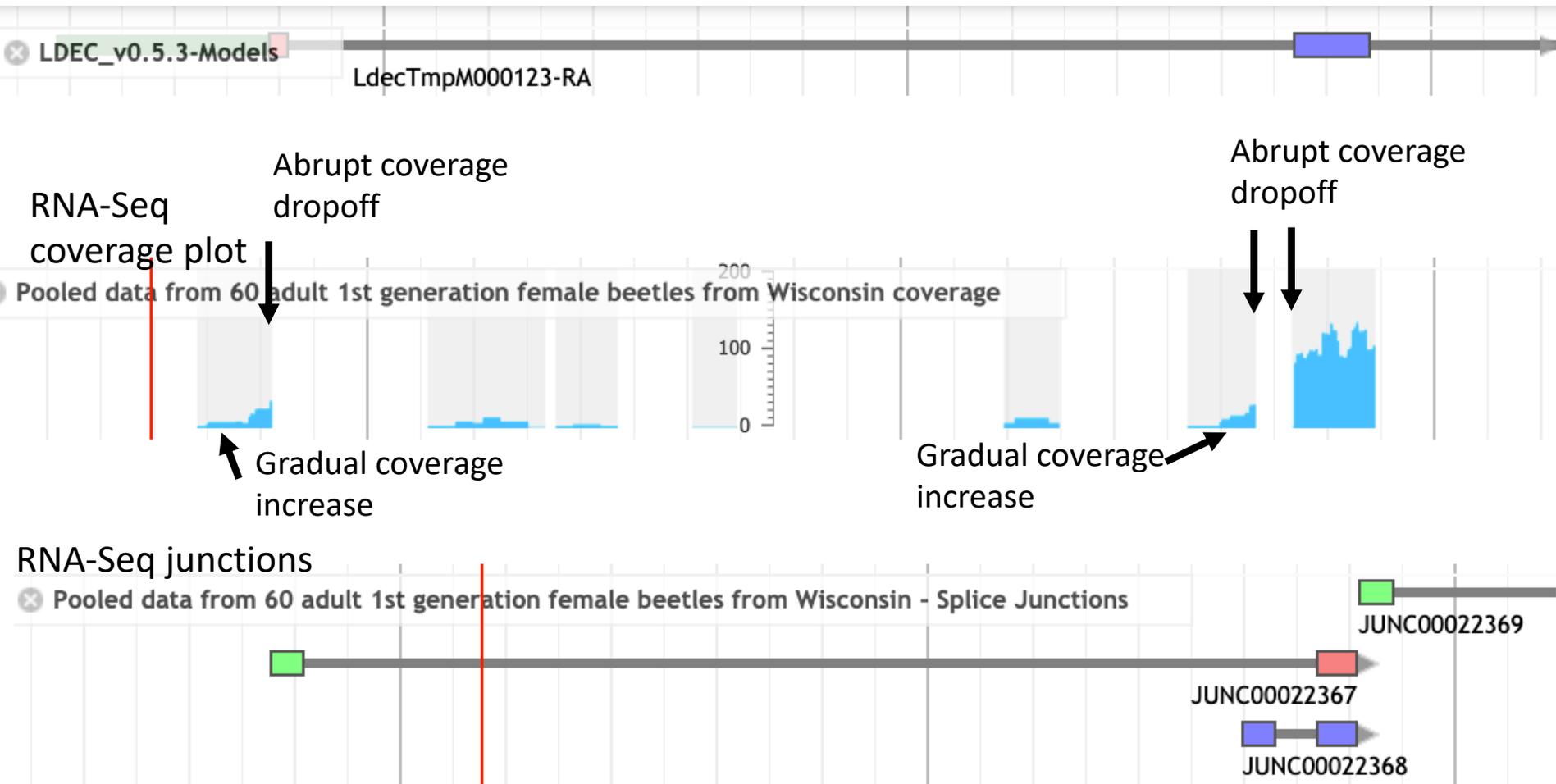
Annotating isoforms

Isoform annotation example

- In our experience, lots of mapped RNA-Seq reads are critical for good manual isoform annotation
- Before evaluating RNA-Seq for isoforms, it helps to understand how to interpret gradual and abrupt drops in coverage
 - Gradual – usually means 5' start or 3' end of expression
 - Abrupt – usually means splice junction
- Checking junction reads (if available) is incredibly useful

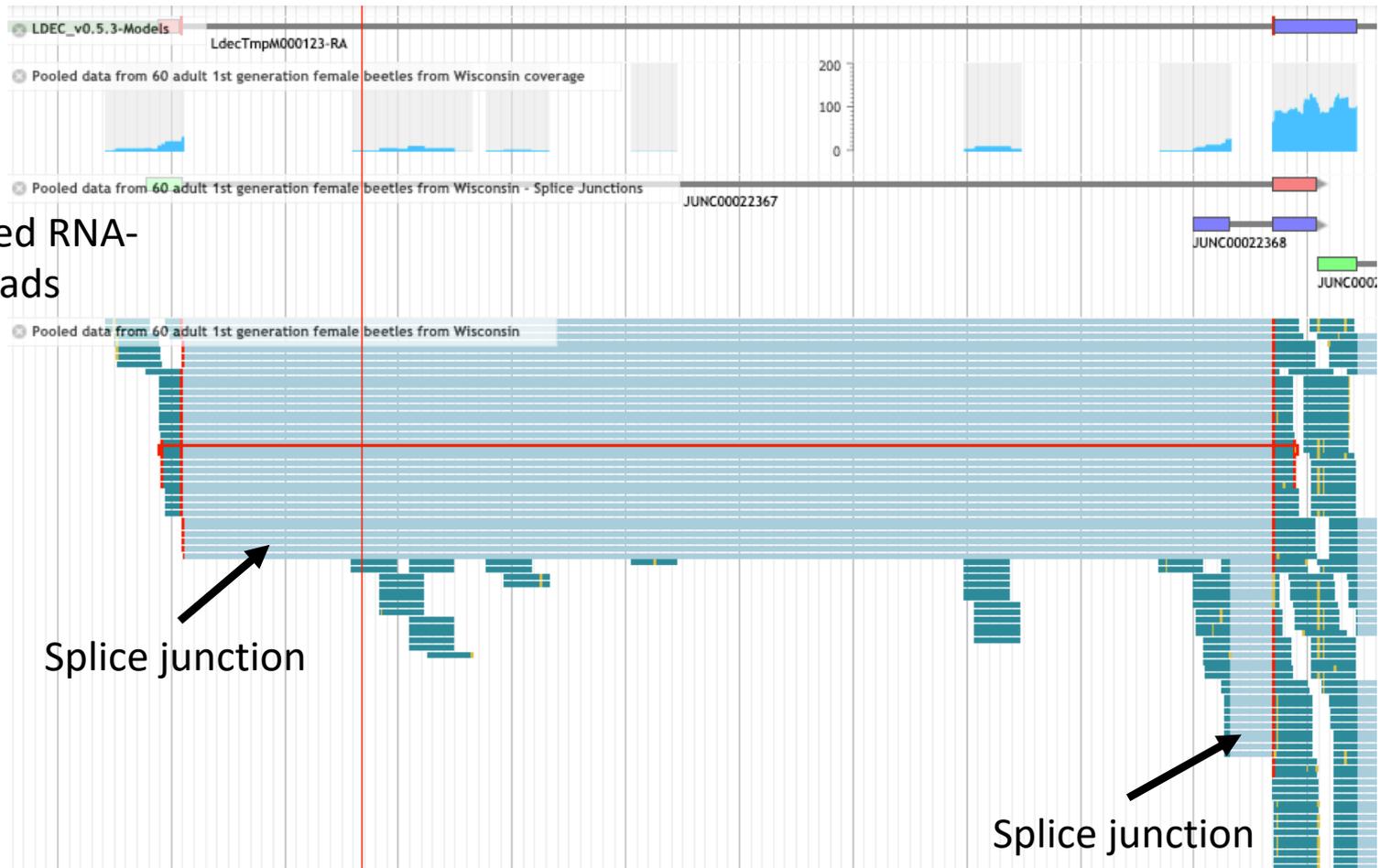
Isoform annotation example

5' end of MAKER tyrosine
protein kinase gene prediction



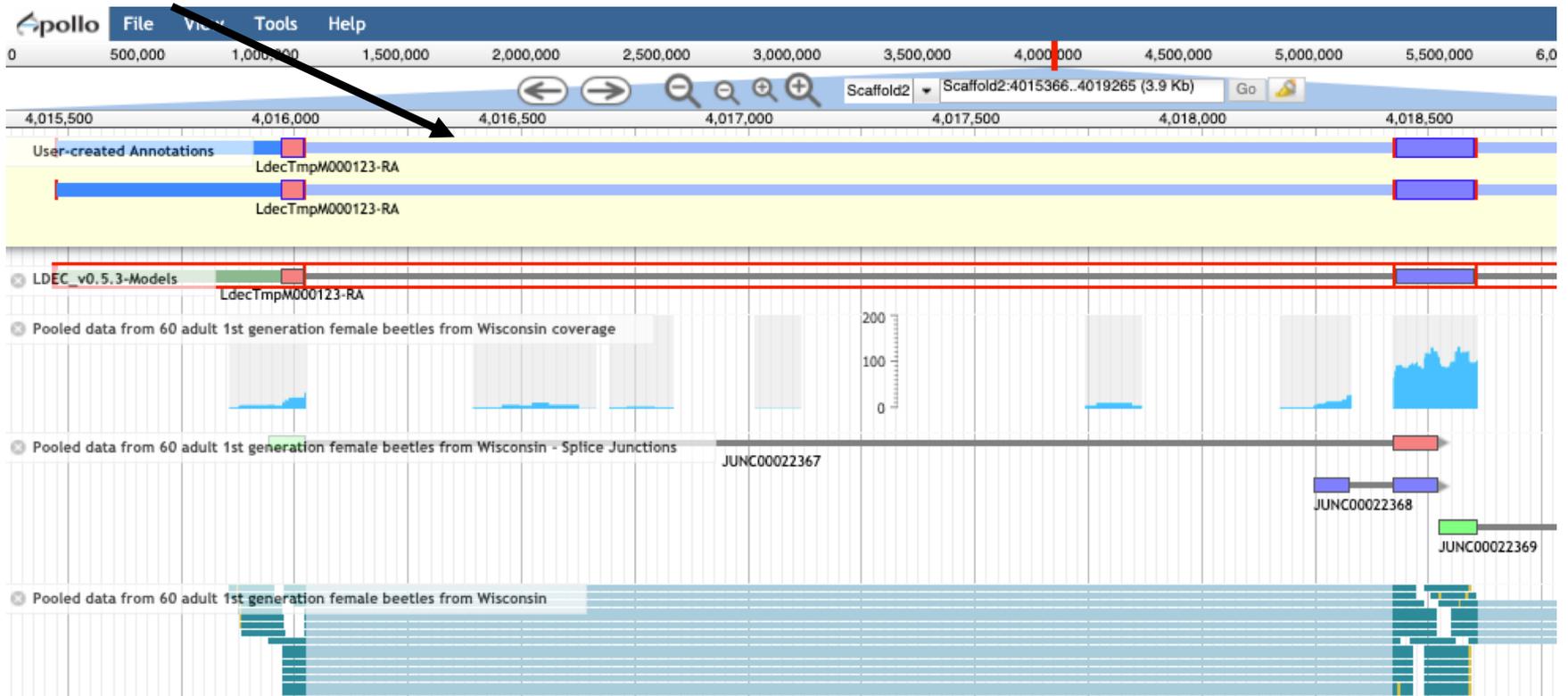
Isoform annotation example

Mapped RNA-Seq reads



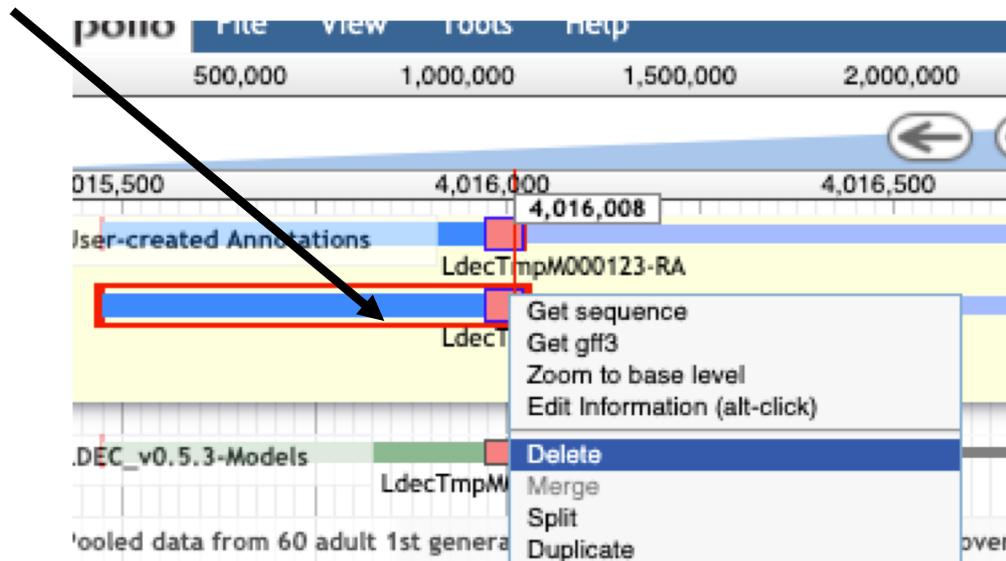
Isoform annotation example

Create 2 isoforms
from Maker model



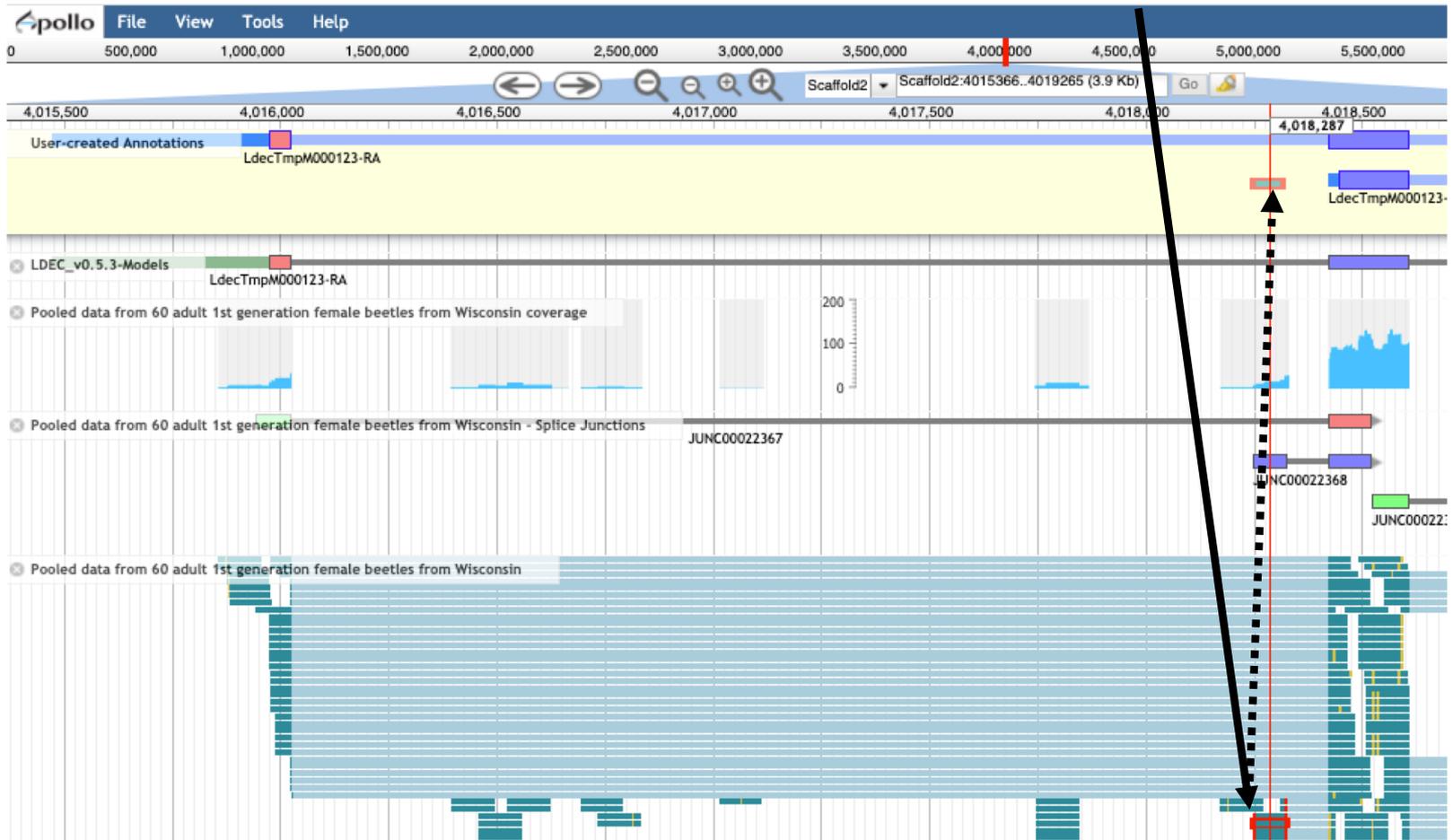
Isoform annotation example

Select and delete 5' exon from one of the isoforms



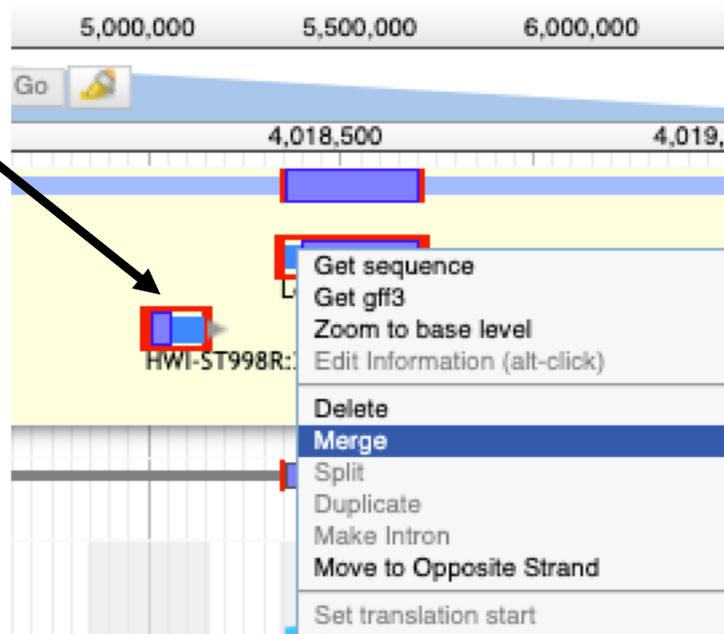
Isoform annotation example

Add a new 5' exon from mapped RNA-Seq evidence



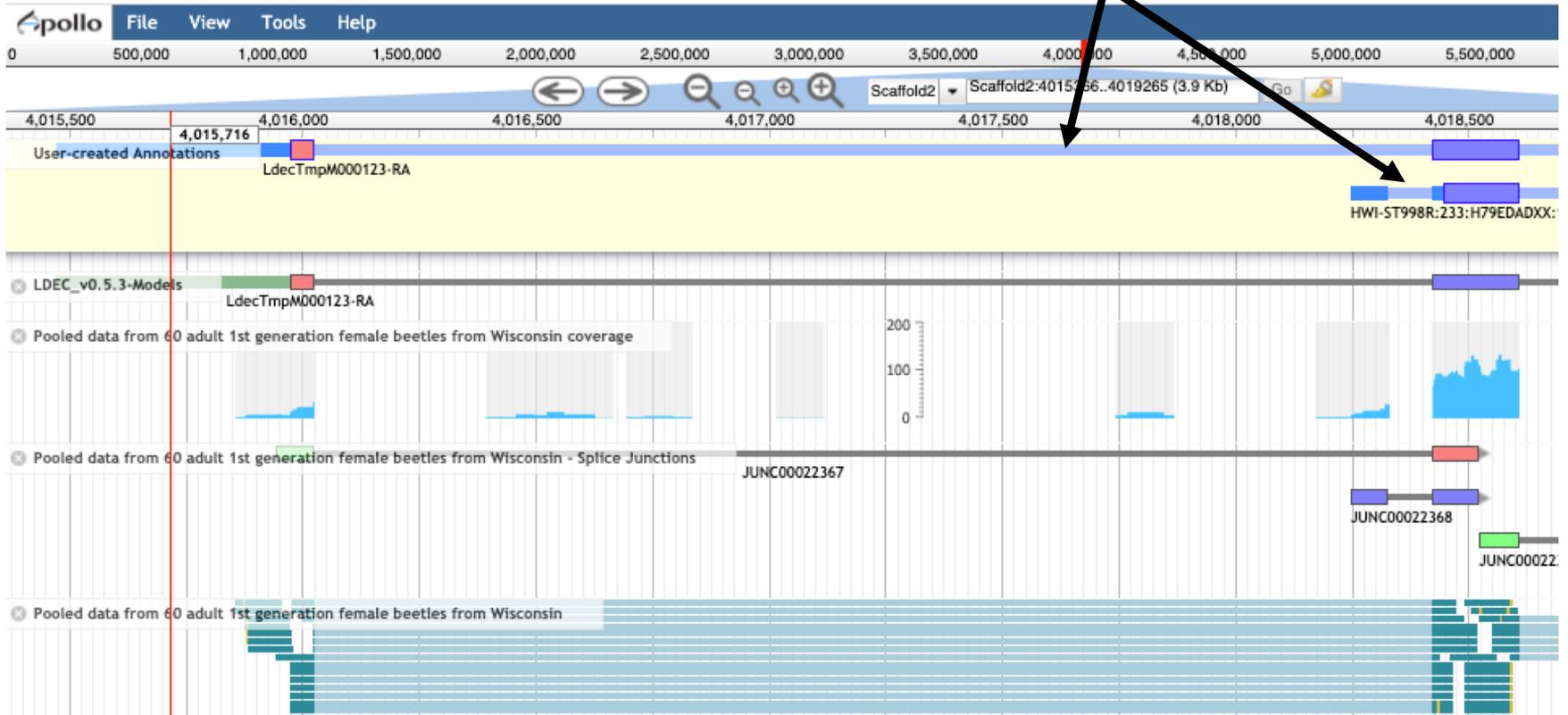
Isoform annotation example

Merge the new 5' exon with the rest of the model



Isoform annotation example

2 isoforms supported by RNA-Seq evidence

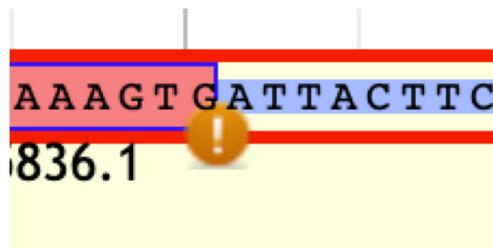
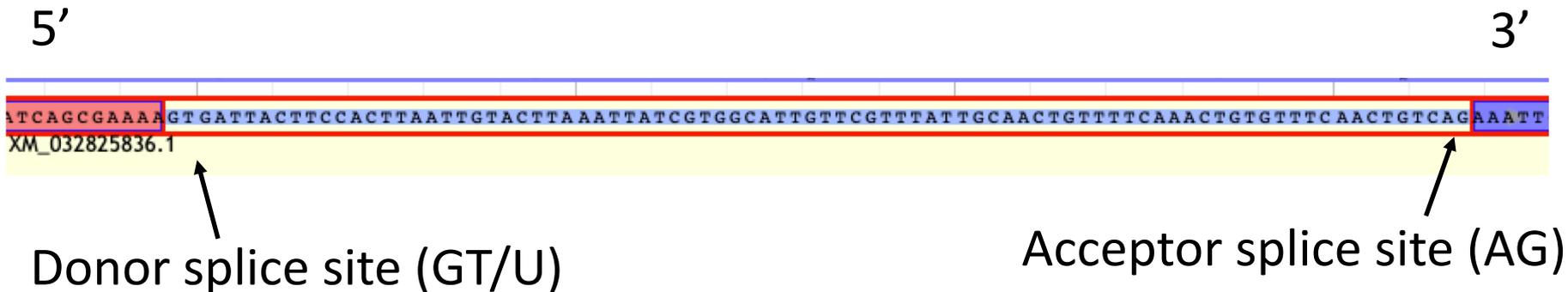


Non-canonical splice sites

Splice sites

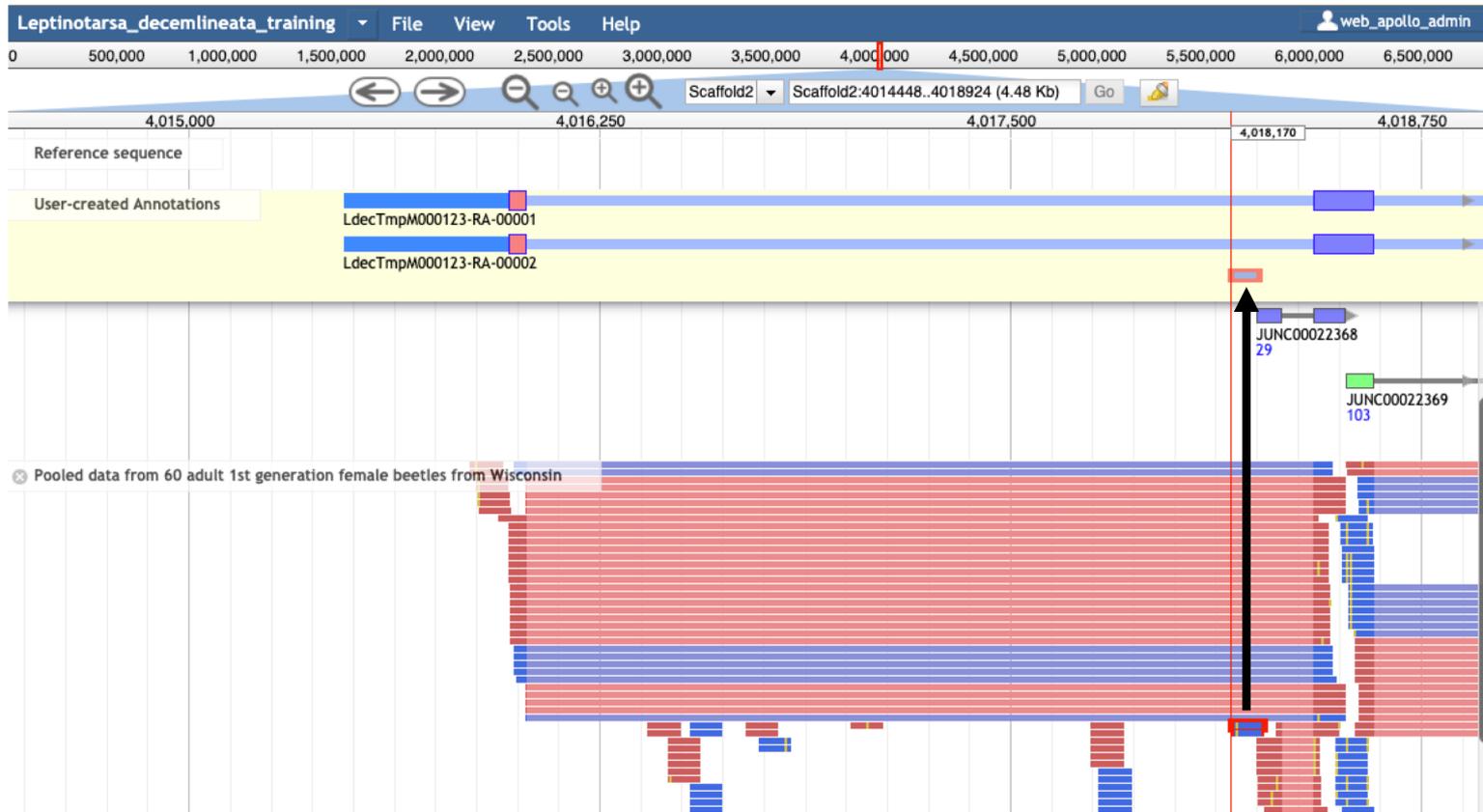
Introns are removed from primary transcripts by cleavage at conserved sequences called **splice sites**. These sites are found at the 5' and 3' ends of introns.

(<https://www.nature.com/scitable/topicpage/rna-splicing-introns-exons-and-spliceosome-12375/>)



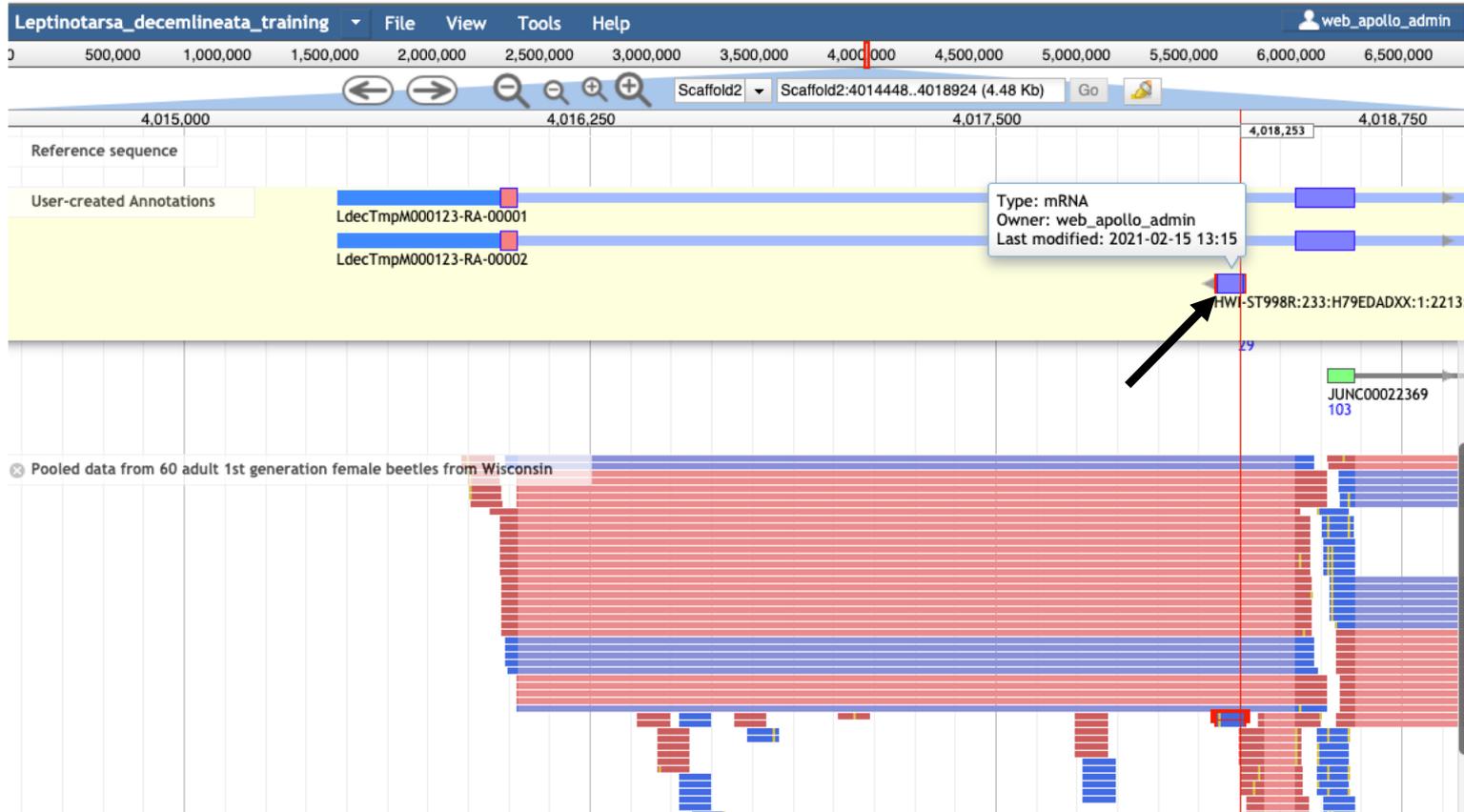
'Non-canonical' splice sites – non-conserved, and possibly erroneous sites – are marked by an exclamation point in Apollo.

Fixing non-canonical splice sites



Let's revisit the previous 'Isoform annotation' example. Let's add the new exon using different starting evidence.

Fixing non-canonical splice sites



That starting evidence was mapped to the reverse strand – let's flip it to the forward strand

Fixing non-canonical splice sites

The screenshot displays the Apollo genome browser interface for the species *Leptinotarsa decemlineata*. The top navigation bar includes a menu (File, View, Tools, Help) and a user profile (web_apollo_admin). A coordinate scale at the top shows positions from 0 to 6,500,000. Below this, a track for Scaffold2 shows a region from 4,015,000 to 4,018,750. The 'User-created Annotations' track contains two transcripts: LdecTmpM000123-RA-00001 and LdecTmpM000123-RA-00002. A 'Pooled data from 60 adult 1st generation female beetles from Wisconsin' track shows red evidence bars. A context menu is open over a red evidence bar, listing actions such as 'Get Sequence', 'Change annotation type', and 'Move to Opposite Strand'. An arrow points to the 'Move to Opposite Strand' option.

Right-click on the evidence and select 'Move to Opposite Strand'

Fixing non-canonical splice sites

The screenshot displays the Apollo genome browser interface for the species *Leptinotarsa decemlineata*. The top navigation bar includes 'File', 'View', 'Tools', and 'Help', along with the user 'web_apollo_admin'. A genomic scale at the top shows coordinates from 0 to 6,500,000. The current view is centered on Scaffold2, specifically Scaffold2:4014448..4018924 (4.48 Kb). The reference sequence is visible at the top, with coordinates 4,015,000, 4,016,250, 4,017,500, and 4,018,750. Below the reference sequence, the 'User-created Annotations' track shows two transcripts: 'LdecTmpM000123-RA-00001' and 'LdecTmpM000123-RA-00002'. The first transcript is highlighted with a red box, and a context menu is open over it. The menu options include: 'Get Sequence', 'Get GFF3', 'Zoom to Base Level', 'View in Annotator Panel', 'Edit Information (alt-click)', 'Change annotation type' (with a sub-menu for 'Associate Transcript to Gene' and 'Dissociate Transcript from Gene'), 'Delete' (highlighted in blue), 'Merge', 'Split', 'Duplicate', 'Make Intron', and 'Move to Opposite Strand'. A black arrow points from the text below to the 'Delete' option. Other annotations include 'HWI-ST998R:233:H79EDADXX:1:2213:29' and 'JUNC00022369 103'. At the bottom left, there is a note: 'Pooled data from 60 adult 1st generation female beetles from Wis'.

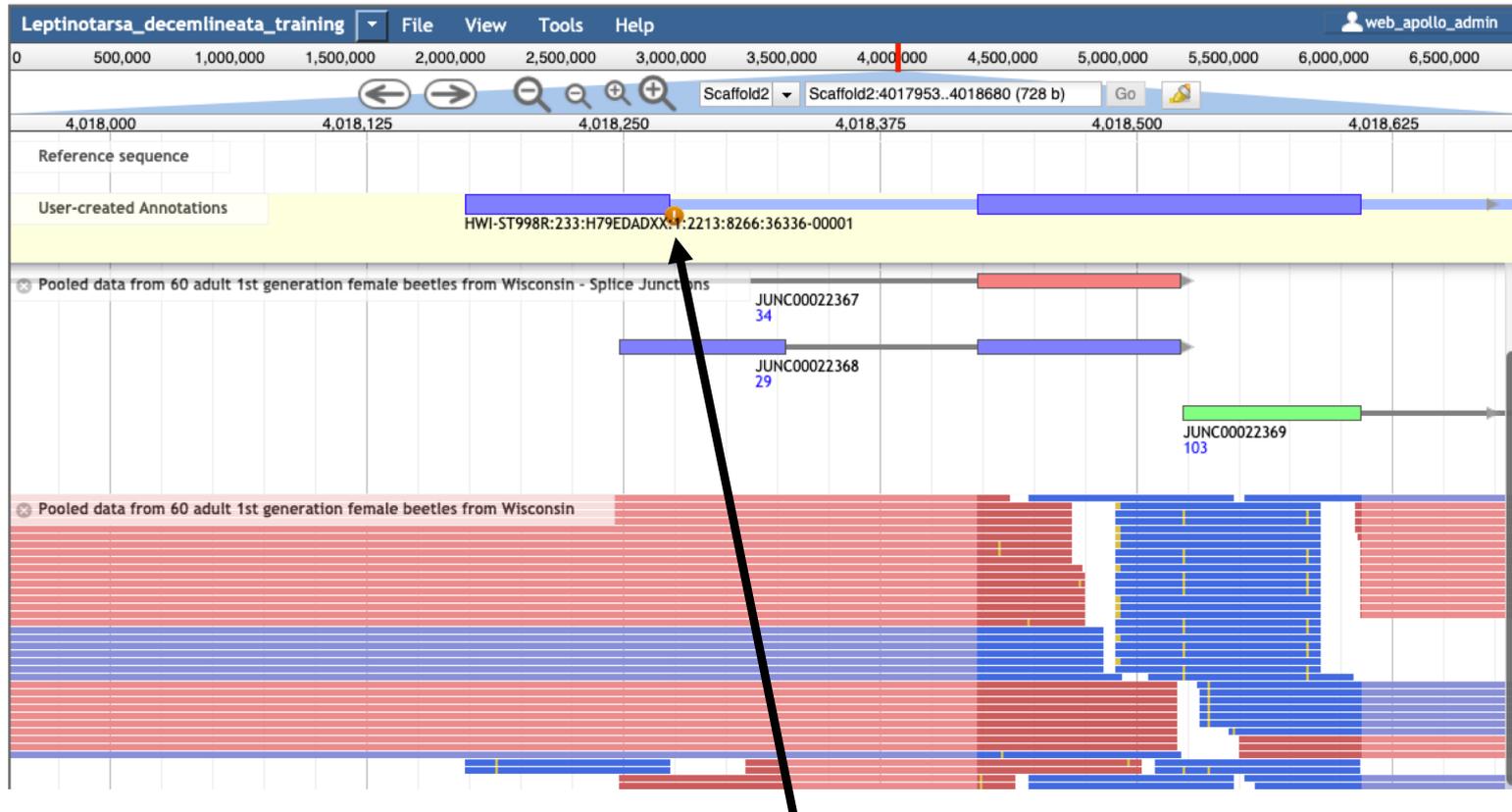
Delete the 5' exon as previously (no support for it for this isoform)

Fixing non-canonical splice sites

The screenshot shows the Apollo genome browser interface for the species *Leptinotarsa decemlineata*. The top navigation bar includes 'File', 'View', 'Tools', and 'Help'. The main view displays genomic coordinates from 4,018,000 to 4,018,625. A reference sequence is shown at the top. Below it, 'User-created Annotations' includes a transcript 'LdecTmpM000123-RA-00001' and a junction 'JUNC00022368' with coordinates 'HWI-ST998R:233:H79EDADXX:1:2213:8266:36336-00001'. A context menu is open over the junction, listing options such as 'Get Sequence', 'Zoom to Base Level', and 'Merge'. The 'Merge' option is highlighted in blue. A black arrow points from the text 'Merge the new 5' exon with the model' below to the 'Merge' option in the menu.

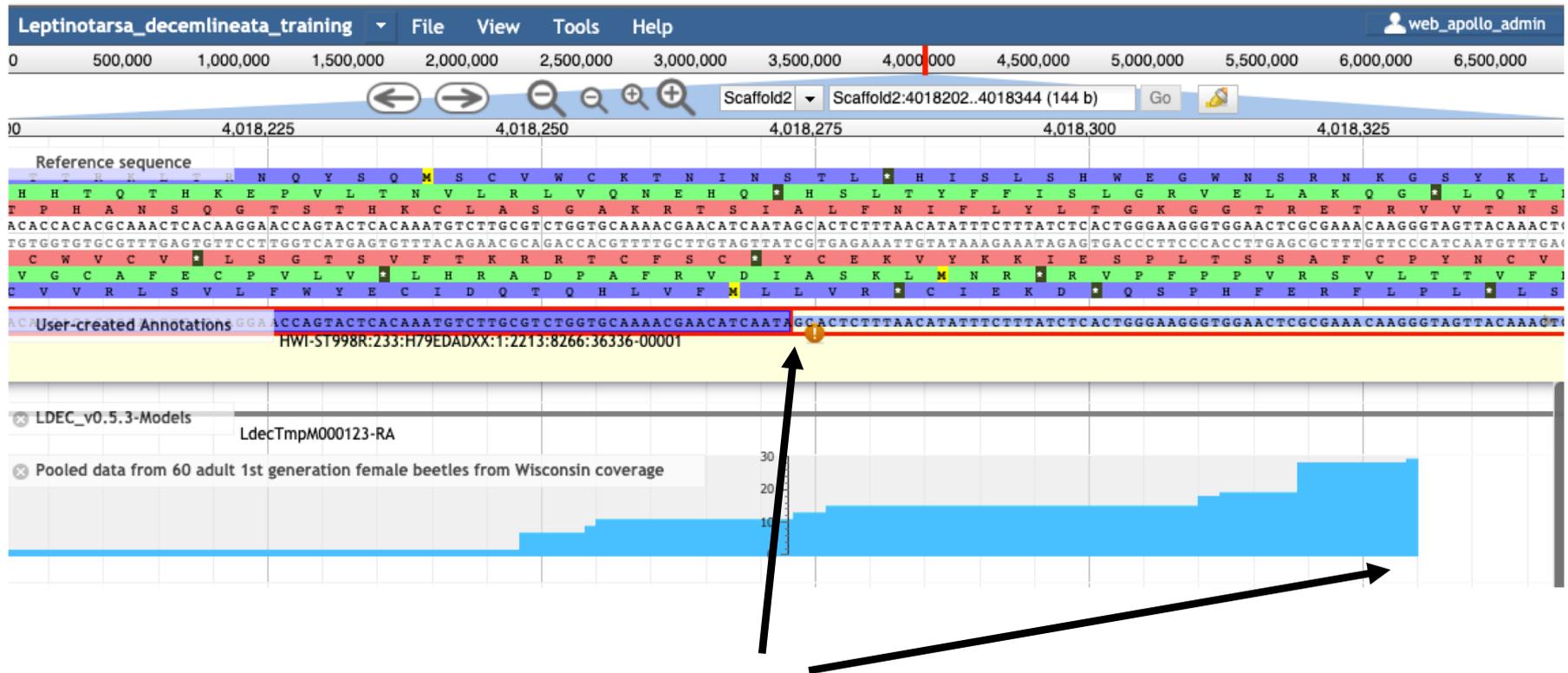
Merge the new 5' exon with the model

Fixing non-canonical splice sites



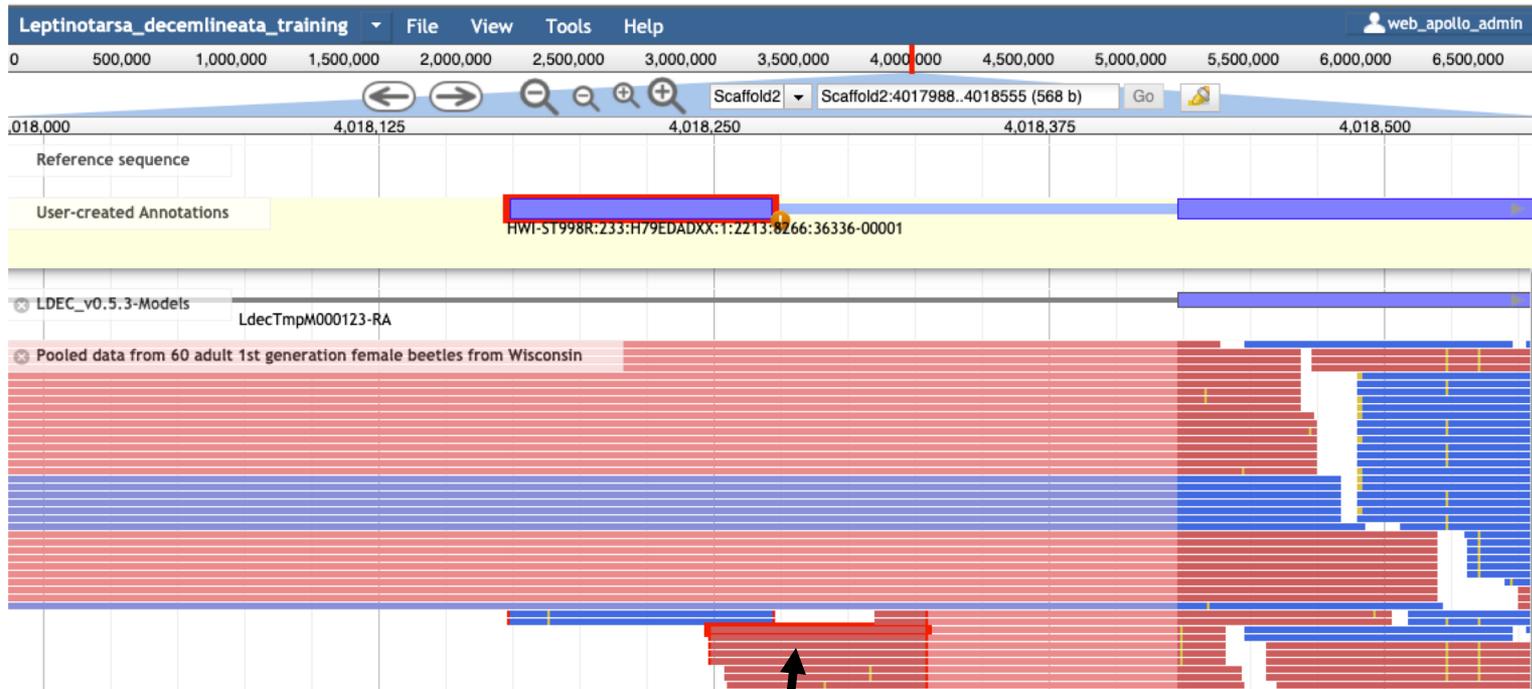
Now that the models are merged, Apollo shows us a non-canonical splice site

Fixing non-canonical splice sites



- Canonical donor splice sites should be G(T/U) – this is GC
- The coverage track shows us that the exon needs to extend further to the right

Fixing non-canonical splice sites



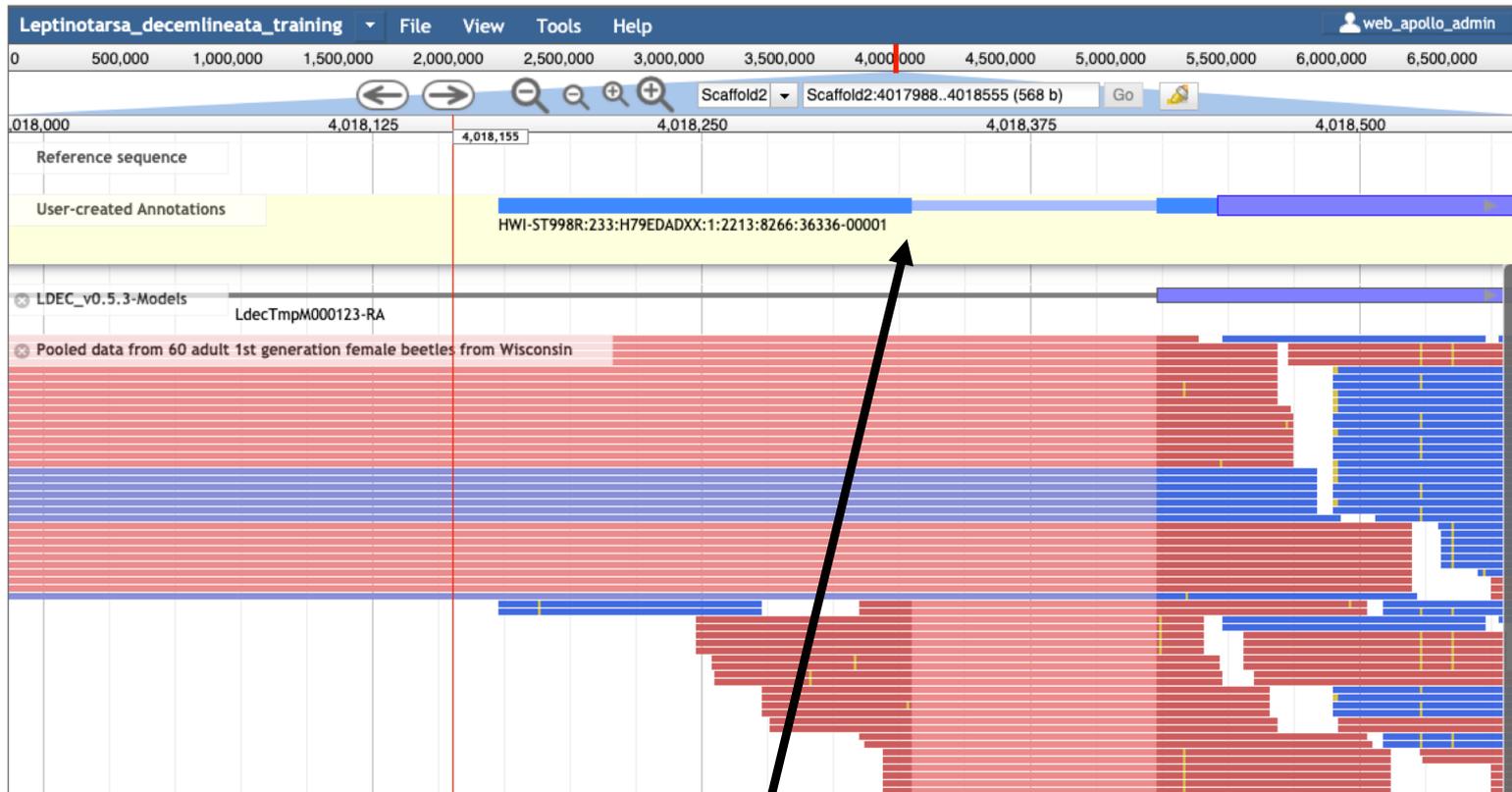
Let's use another RNA-Seq read to extend the exon to the actual splice site

Fixing non-canonical splice sites

The screenshot shows the Apollo genome browser interface. The top navigation bar includes 'Leptinotarsa_decemlineata_training', 'File', 'View', 'Tools', and 'Help'. The main view displays a genomic track for Scaffold2:4017988..4018555 (568 b). The track shows a reference sequence, user-created annotations, and a transcript model (LdecTmpM000123-RA) with exons in red and introns in blue. A context menu is open over a specific exon and an RNA-Seq read (HWI-ST998R:Z33:H79EDADXX:1:2). The menu options include 'Get Sequence', 'Get GFF3', 'Zoom to Base Level', 'View in Annotator Panel', 'Edit Information (alt-click)', 'Change annotation type', 'Associate Transcript to Gene', 'Dissociate Transcript from Gene', 'Delete', 'Merge', 'Split', 'Duplicate', 'Make Intron', 'Move to Opposite Strand', 'Set Translation Start', 'Set Translation End', 'Set Longest ORF', 'Set Readthrough Stop Codon', 'Set as 5' end', 'Set as 3' End', 'Set as both Ends', 'Set to Downstream Splice Donor', and 'Set to Upstream Splice Donor'. The 'Set as 3' End' option is highlighted, and a black arrow points to it from the text below.

- Shift-click on the model's 5' exon and the RNA-Seq read
 - right-click to open menu
 - Select 'Set as 3' end'

Fixing non-canonical splice sites



Fixed.

Thank you!

The NAL Team

- Chris Childers
- Vern Chapman
- Ming Chen
- Susan McCarthy
- Shang-Yu Chang

- i5k Coordinating Committee
- i5k Pilot Project
- Apollo & JBrowse Development Teams
- GMOD/Tripal community
- All of our users and contributors!

Contact us:

- <https://i5k.nal.usda.gov/contact>
- i5k@ars.usda.gov
- Monica.Poelchau@usda.gov
- Christopher.Childers2@usda.gov