

Using Apollo at the i5k Workspace@NAL

Monica Poelchau, USDA-ARS NAL

August 18th, 2020

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
- Sequence alterations and stop-codon readthroughs
- Annotating Non-coding features

Other resources

- An additional Apollo webinar with more background:
<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:
 - <http://genomearchitect.org/users-guide/>
- 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://eupathdb.org/eupathdb/webinars.jsp#apollo>

Basic RNA-Seq evaluation

RNA-Seq tracks

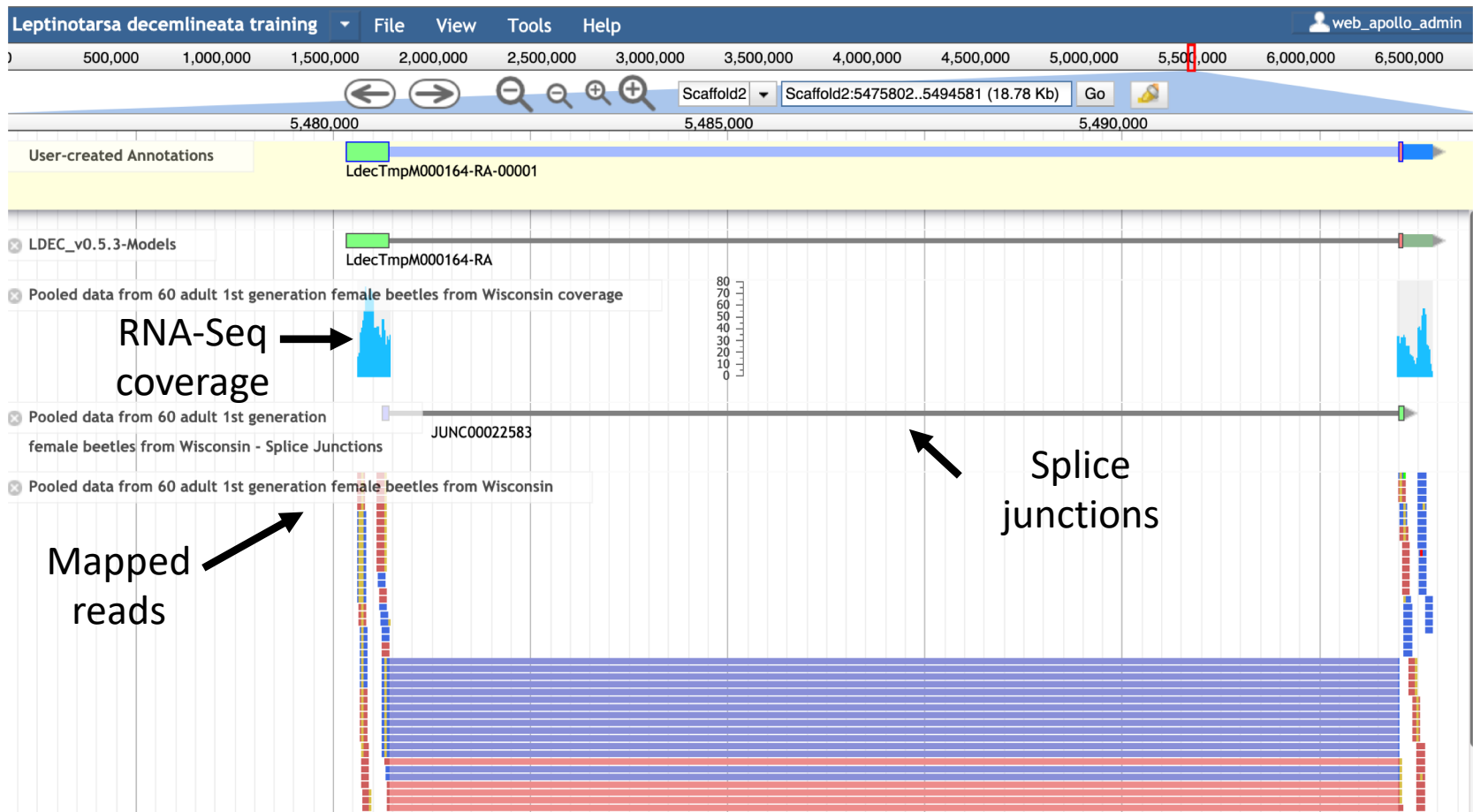
- **Coverage plots:** Histogram of the number of mappings at each nucleotide; hover over the blue area to see the value
- **Mapped reads:** Individual glyphs of each mapped read. Show mapped and spliced areas, and SNPs/indels. Informative, but hard to work with when zoomed out.
- **Junction reads:** Useful combined with coverage plots; show where mapped reads are spliced. Control-click on read and look under 'score' to see how many mapped reads support the splice junction.

The screenshot shows the 'Available Tracks' panel in a genomic browser. The panel is organized into sections, each with a dropdown arrow and a count. The sections are:

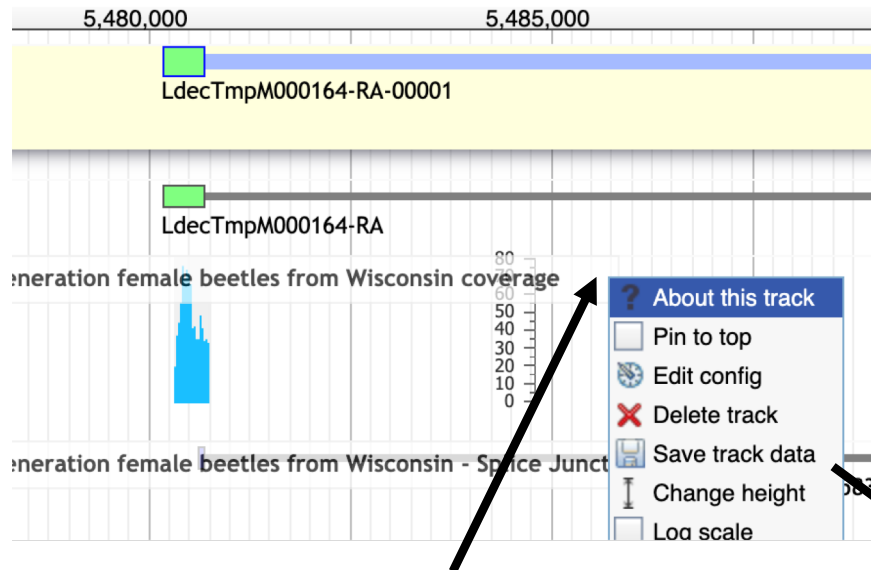
- 0. Reference Assembly** (3 tracks):
 - ☐ GC Content
 - ☐ Gaps in assembly
 - ☒ BLAST+ Results
- NCBI Annotation Release 100** (2 tracks):
 - ☒ NCBI_Annotation_Release_100_Gene
 - ☐ NCBI_Annotation_Release_100_Pseudogene
- RNA-Seq** (25 tracks):
 - Coverage Plots** (8 tracks):
 - ☒ Egg, coverage
 - ☒ 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 tracks):
 - ☐ Egg
 - ☐ Larva
 - ☐ Male
 - ☐ Non-reproductive Adult Worker
 - ☐ Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02
 - ☐ Pupa
 - ☐ Queen
 - ☐ Reproductive Adult Worker
 - Splice junctions** (8 tracks):
 - ☐ 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 track):

Three black arrows point to the 'Coverage Plots', 'Mapped Reads', and 'Splice junctions' sections of the RNA-Seq track.

A simple case



A simple case



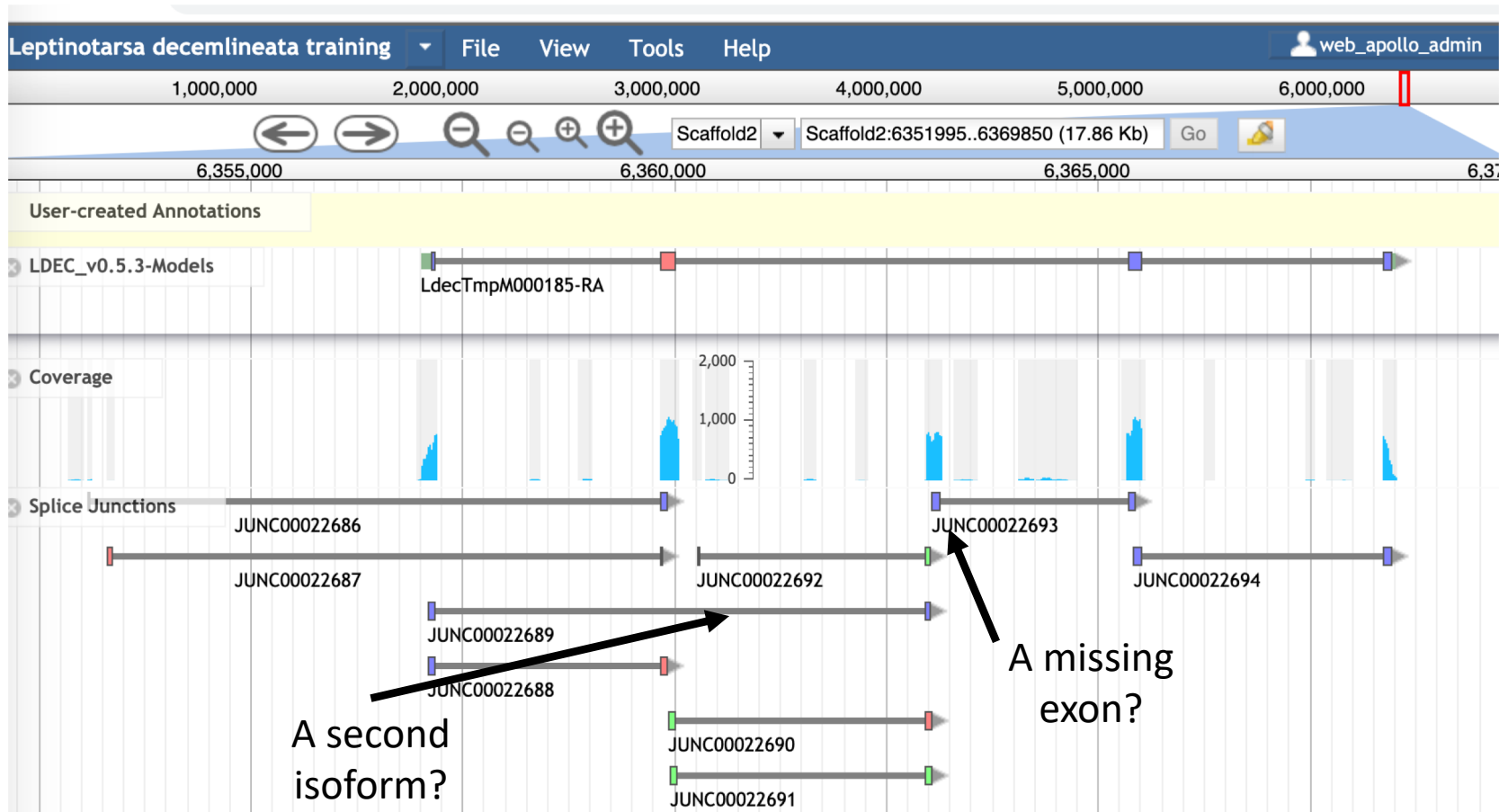
Information about methods

Select 'About this track' from drop-down menu

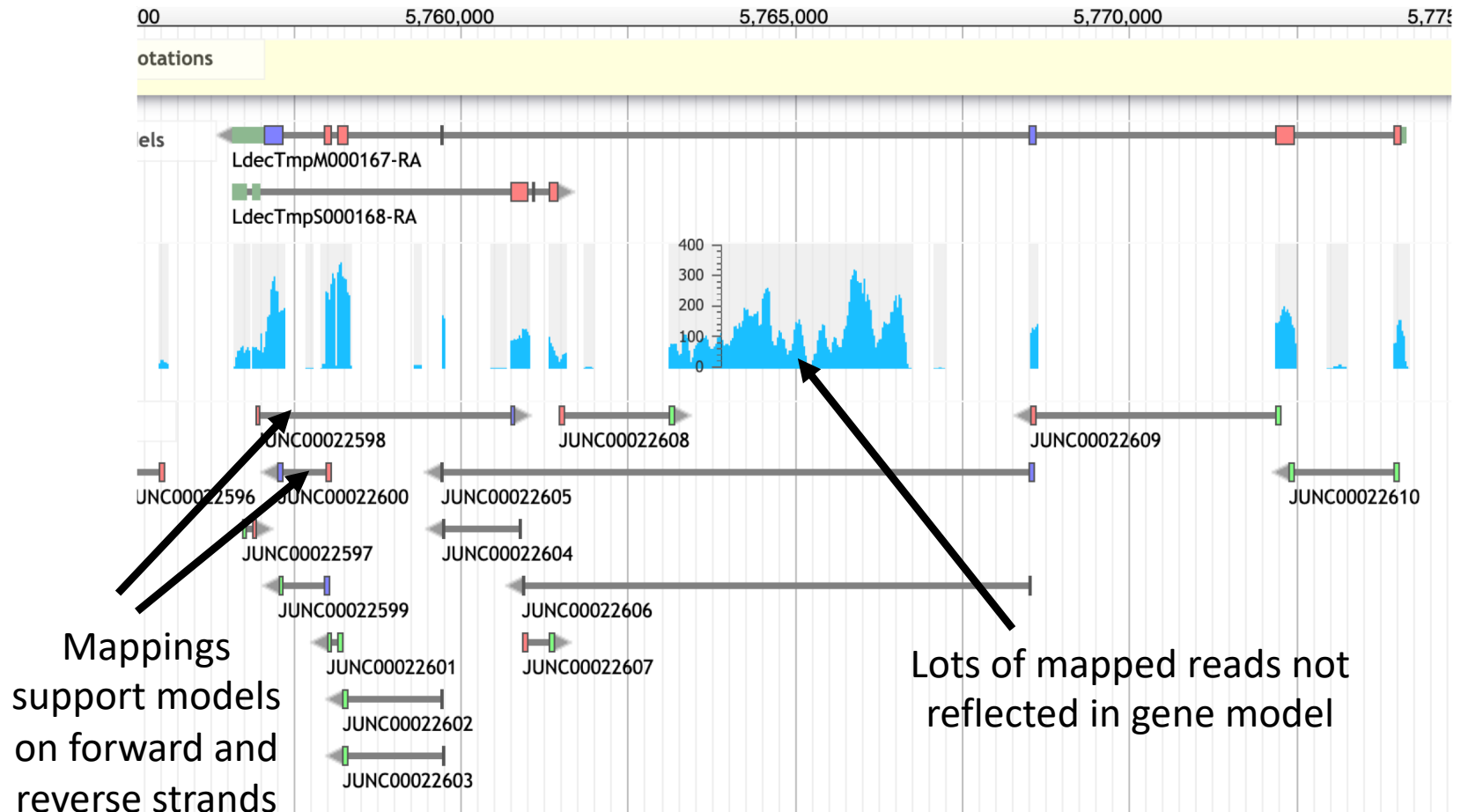
About track: Pooled data from 60 adult 1st generation female beetles from Wisconsin coverage

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

A more complex case



A really messy case

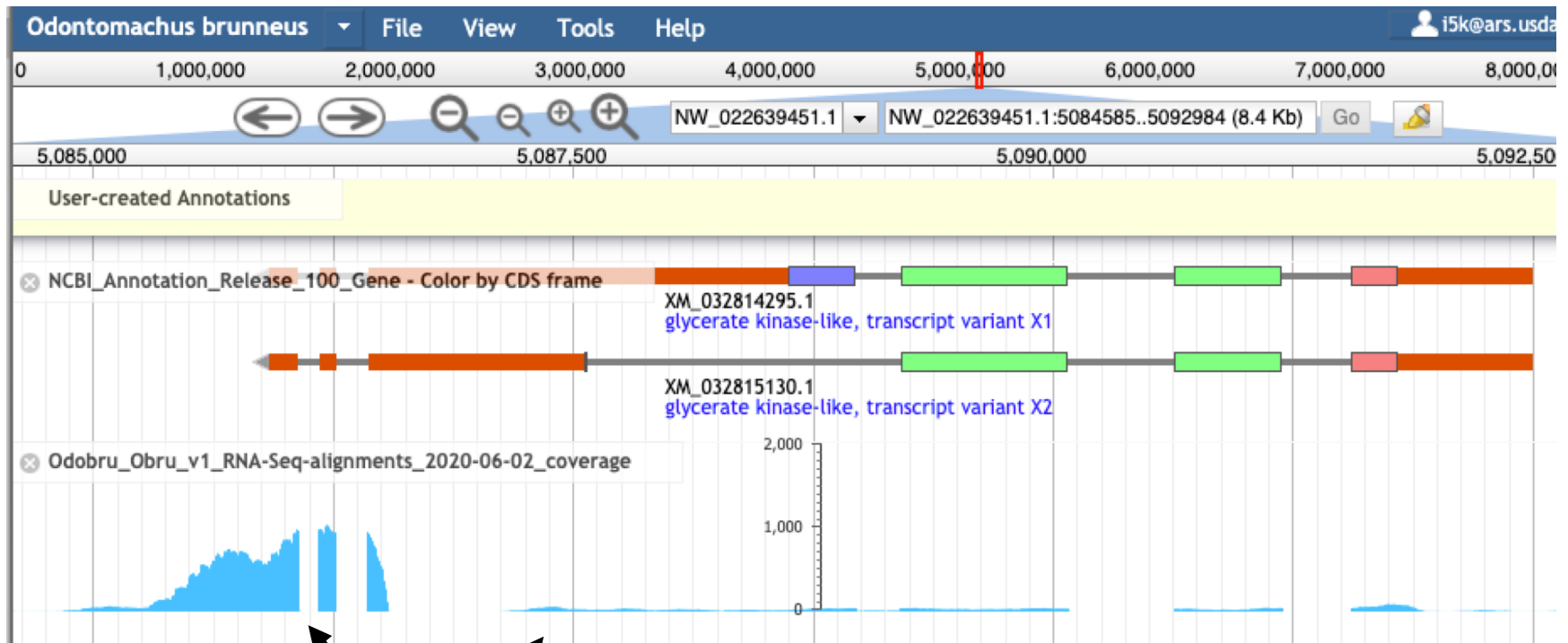


Basic structural changes –
splitting and merging a
model, adding and
removing exons

Annotation Example

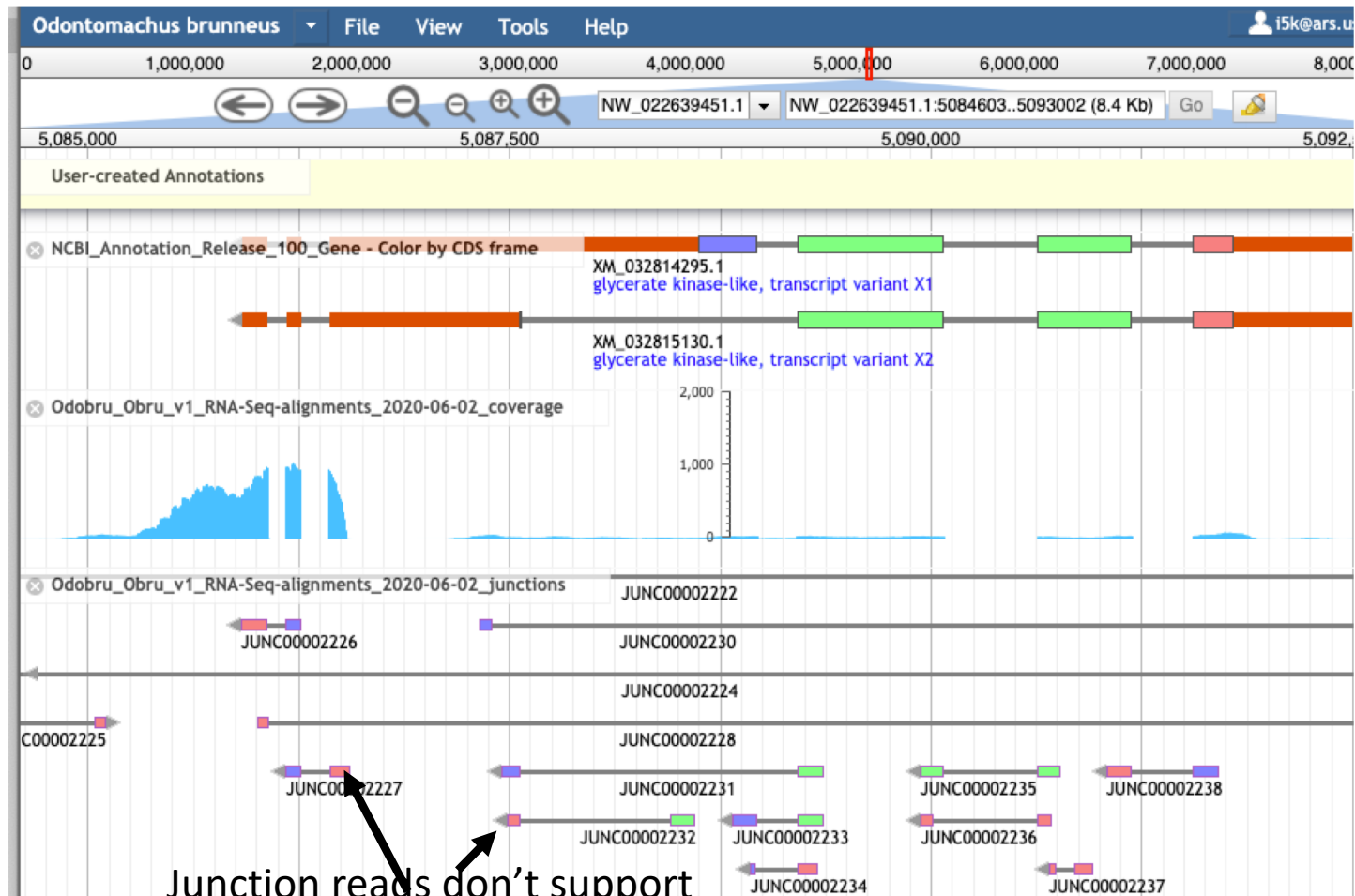
- Glycerate kinase-like in the trap-jaw ant *Odontomachus brunneus*
- More information about the trap-jaw ant genome assembly: <https://i5k.nal.usda.gov/odontomachus-brunneus>
- *Odontomachus brunneus* Apollo URL:
https://apollo.nal.usda.gov/apollo/4006447/jbrowse/index.html?loc=NW_022639451.1%3A5084490..5093717&tracks=DNA%2CAnnotations%2CNCBI_Annotation_Release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

RNA-Seq evaluation



Very different coverage
between UTR and CDS

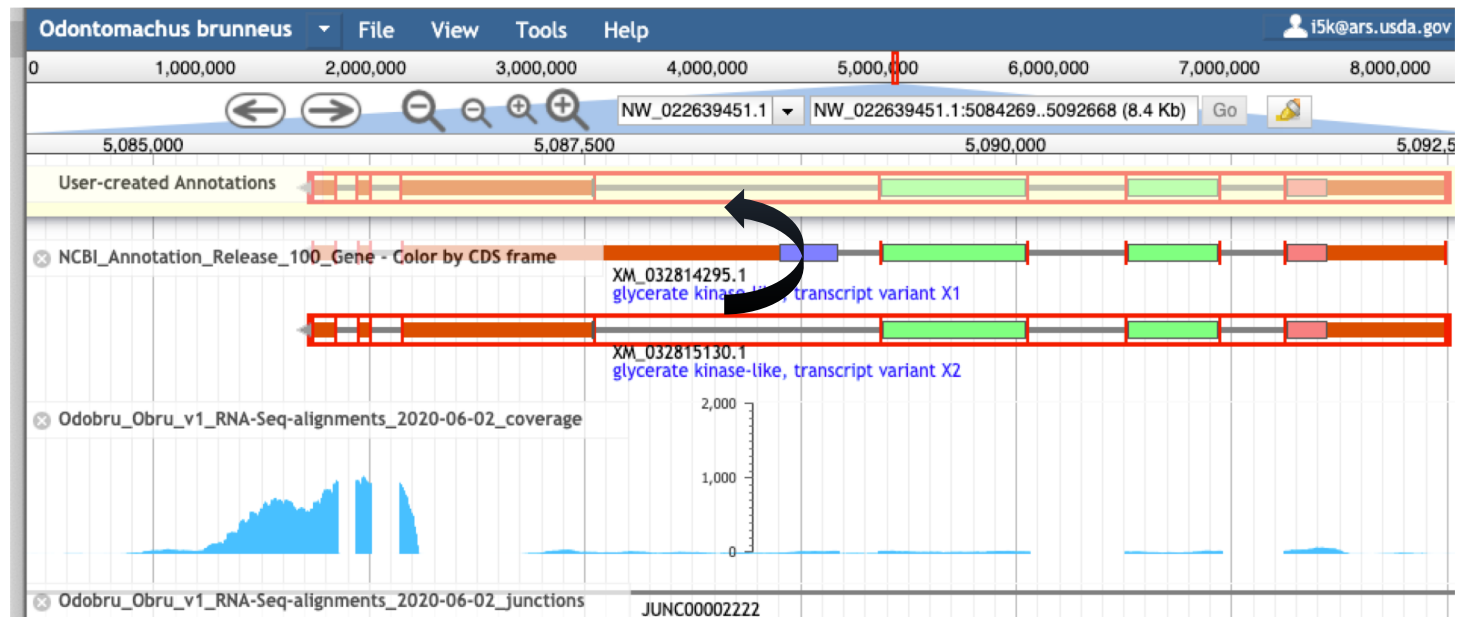
RNA-Seq evaluation



Junction reads don't support
connection between the two
expressed regions

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

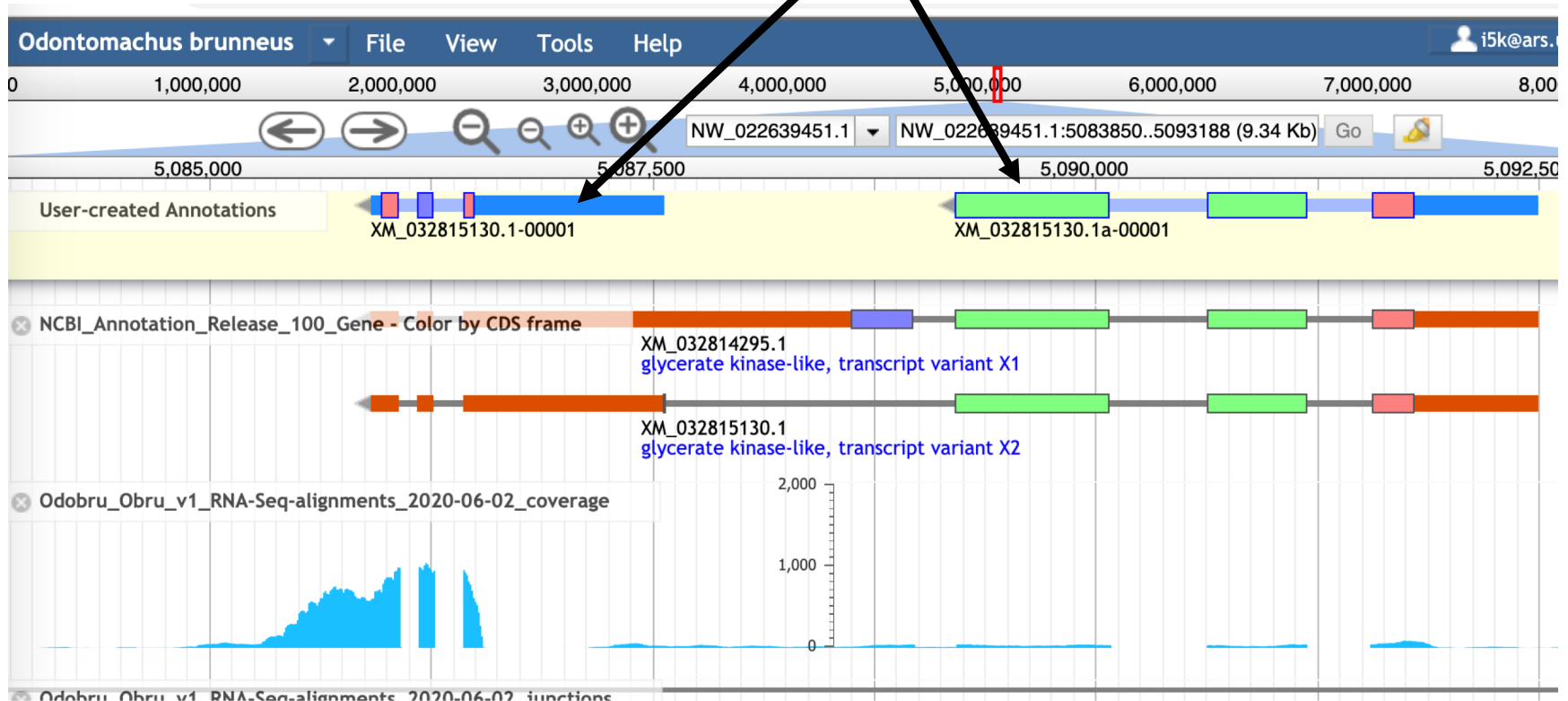
The screenshot displays the genome browser interface for *Odontomachus brunneus*. The top navigation bar includes 'File', 'View', 'Tools', and 'Help'. The main track shows 'User-created Annotations' with two exons highlighted in red boxes. A right-click context menu is open over the second exon, listing various actions. The 'Split' option is highlighted in blue. Below the menu, the 'NCBI_Annotation_Release_100_Gene - Color by CDS frame' track shows two transcripts: XM_032814295.1 (glycerate kinase-like, transcript variant X1) and XM_032815130.1 (glycerate kinase-like, transcript variant X2). The bottom track shows 'Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02_coverage' with a blue histogram.

- 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

Select 'split'

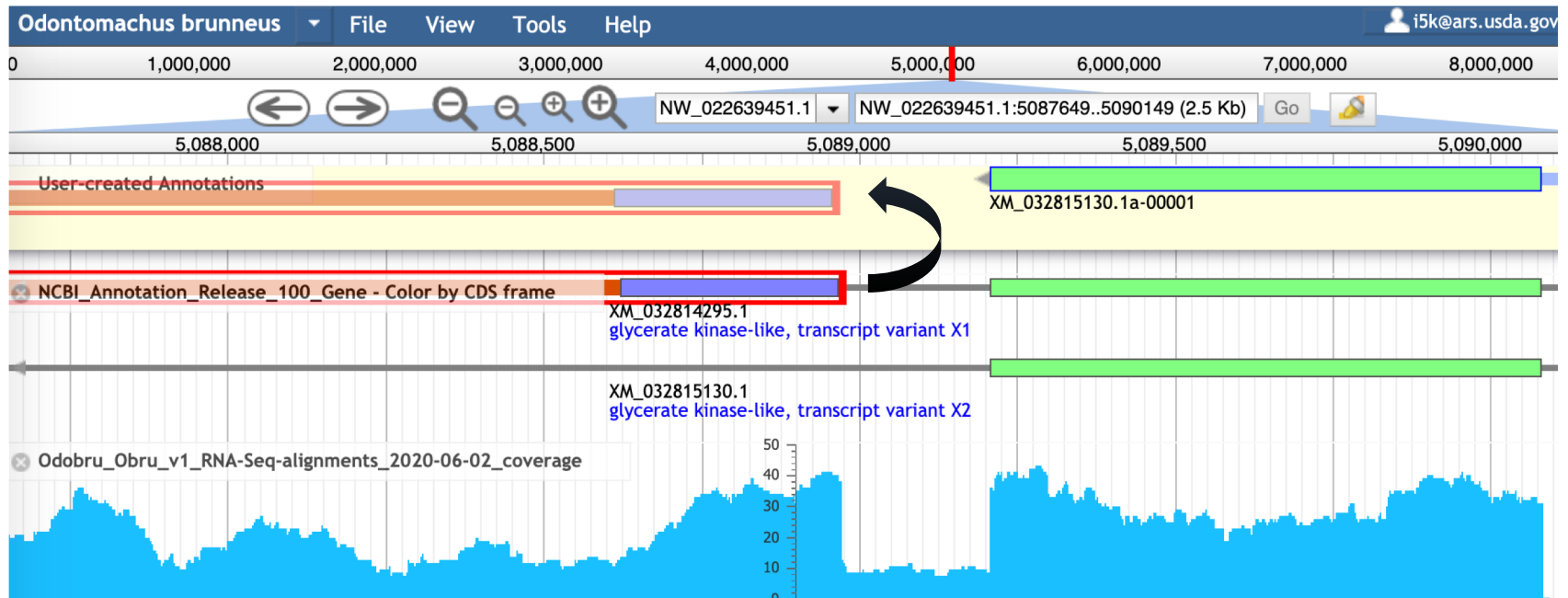
Split model

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 Genious genome browser interface for the species *Odontomachus brunneus*. The top navigation bar includes 'File', 'View', 'Tools', and 'Help' menus, along with a user profile icon and email address 'i5k@ars.usda.gov'. The main view shows a genomic track with coordinates from 1,000,000 to 8,000,000. A zoomed-in section shows coordinates from 5,088,000 to 5,090,000. The track includes several annotations: 'User-created Annotations' (yellow background), 'NCBI_Annotation_Release_100_Gene - Color by CDS frame' (orange background), and 'Odobru_Obru_v1_RNA-Seq-alignments_2020-06-02_coverage' (blue histogram). Two exons are highlighted with red boxes: 'XM_032814295.1-00001' (blue bar) and 'XM_032815130.1' (green bar). A context menu is open over the second exon, listing various actions. The 'Merge' option is highlighted in blue. The menu options are: 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.

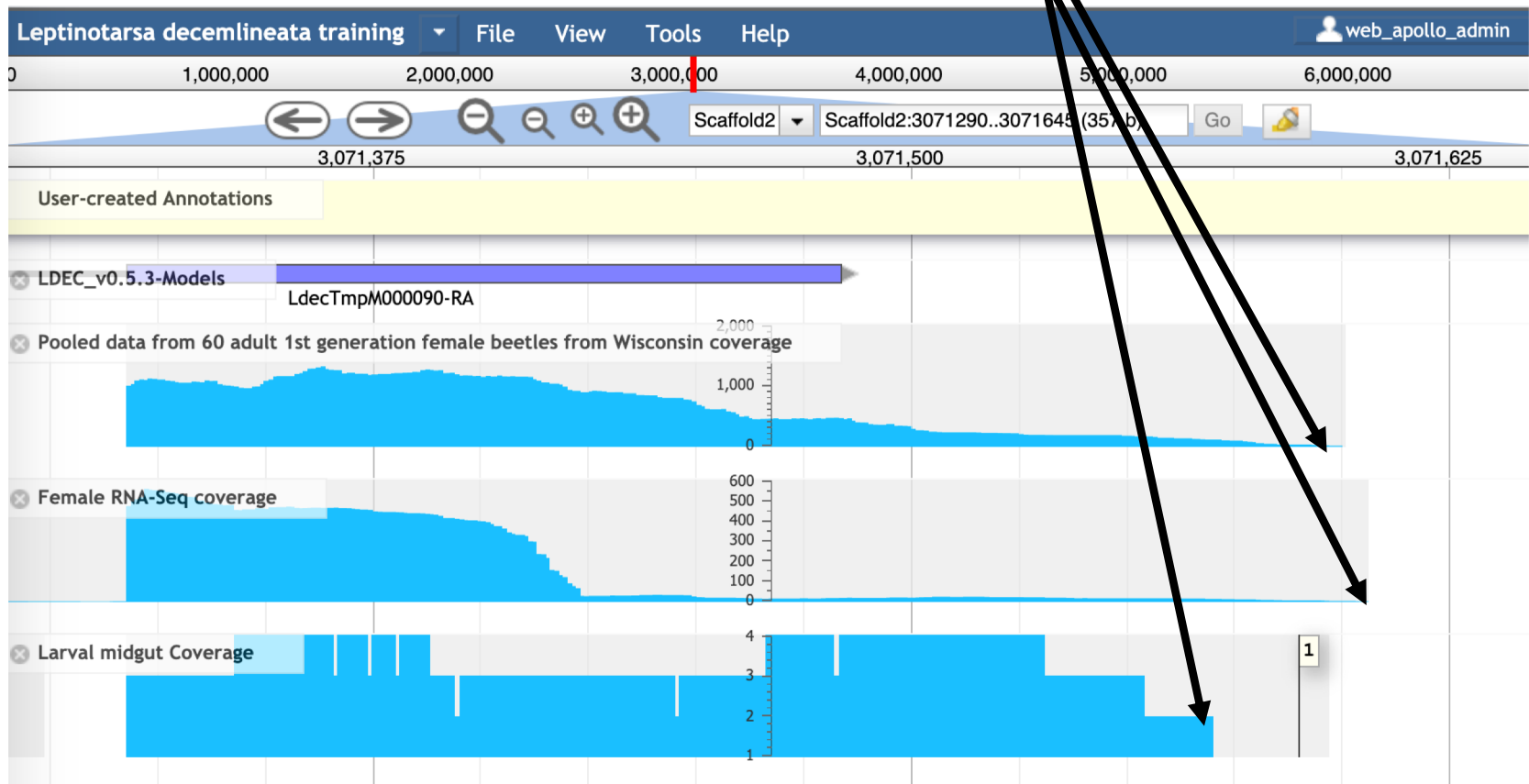
UTRs – how and when to
add or adjust

Adding or adjusting UTR boundaries

- When should you add or change UTRs?
 - Only if you have RNA-Seq evidence with sufficient coverage (e.g. > 50 reads)
 - 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

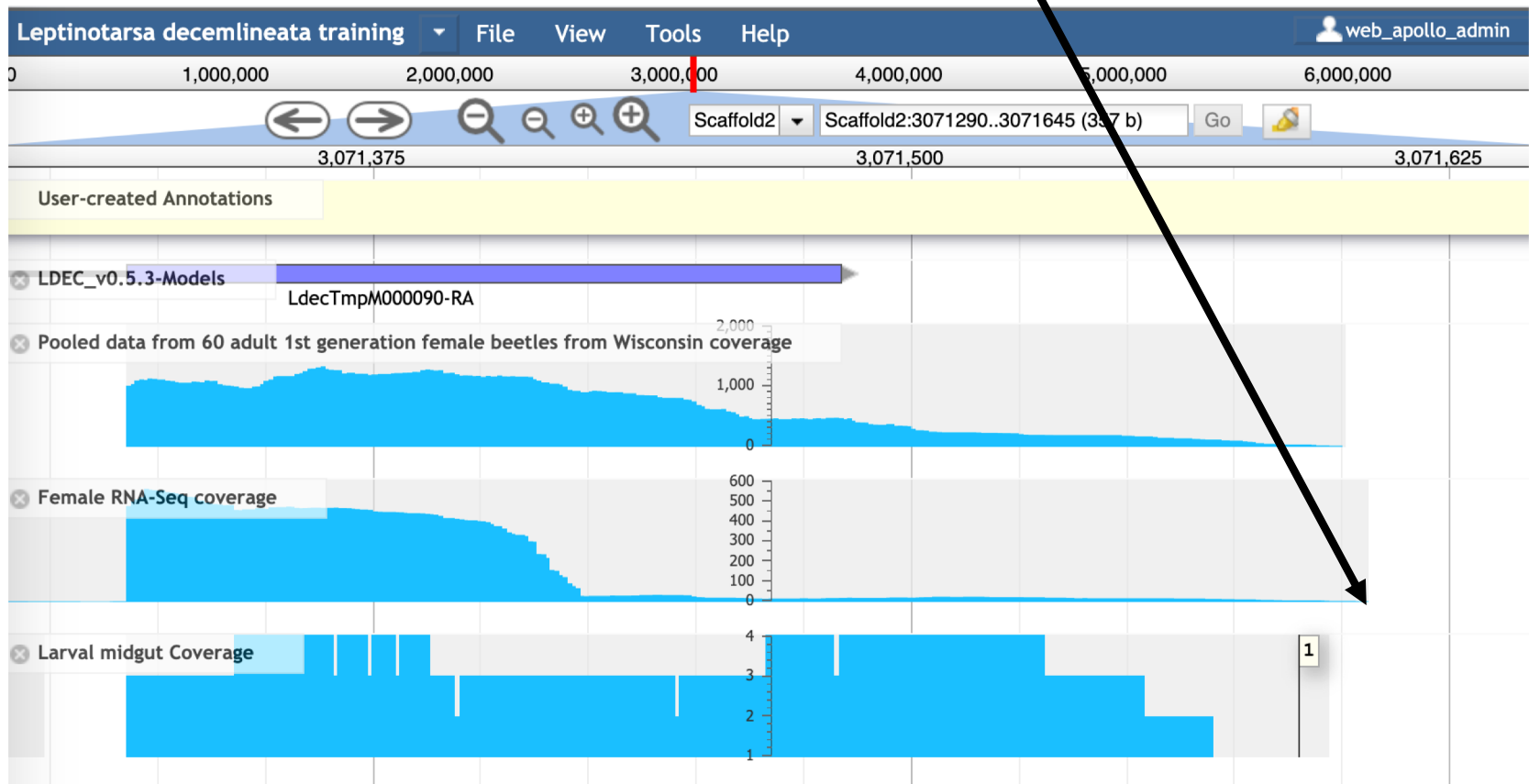
Adjusting gene boundaries

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

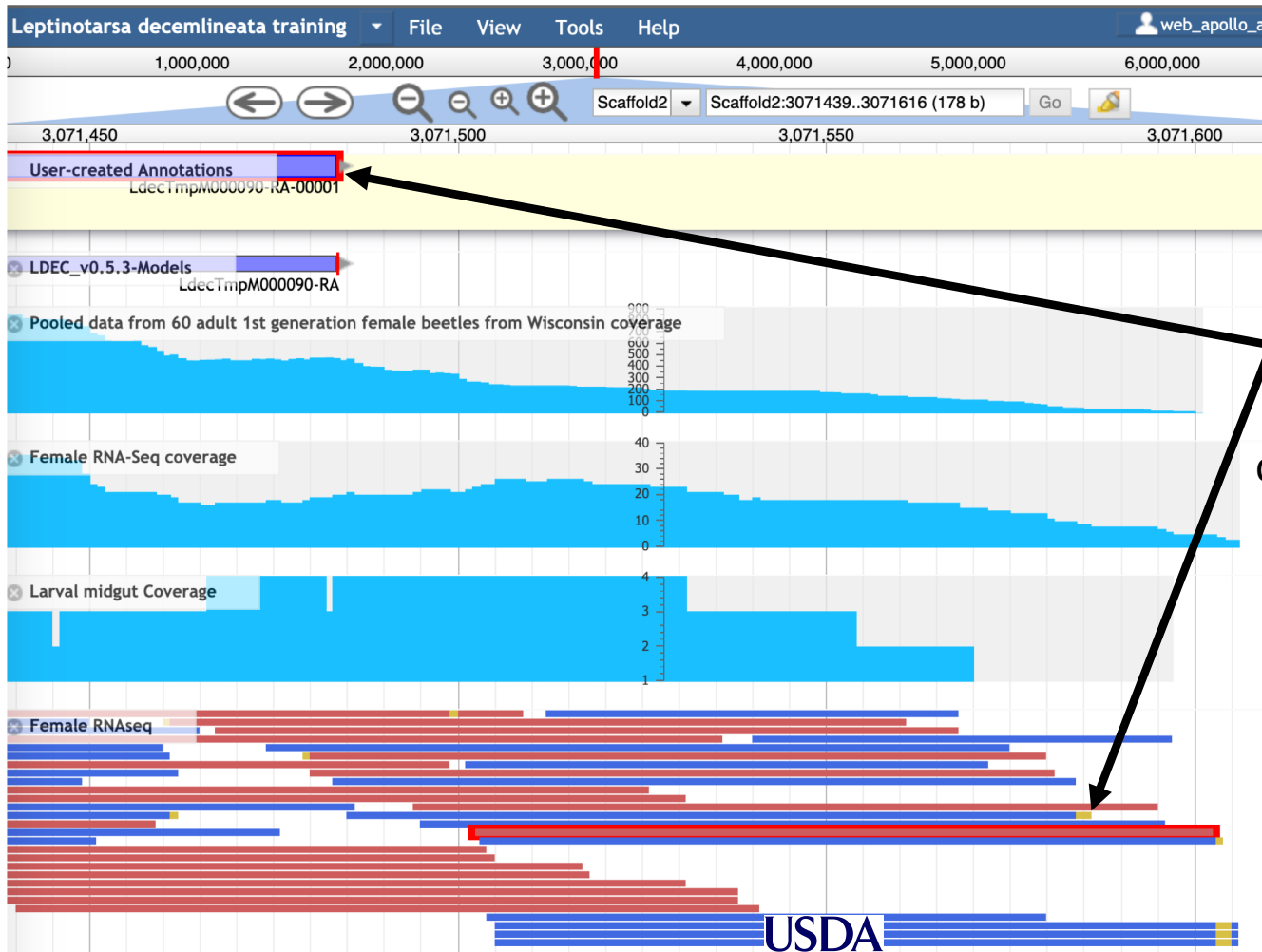


Adjusting gene boundaries

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

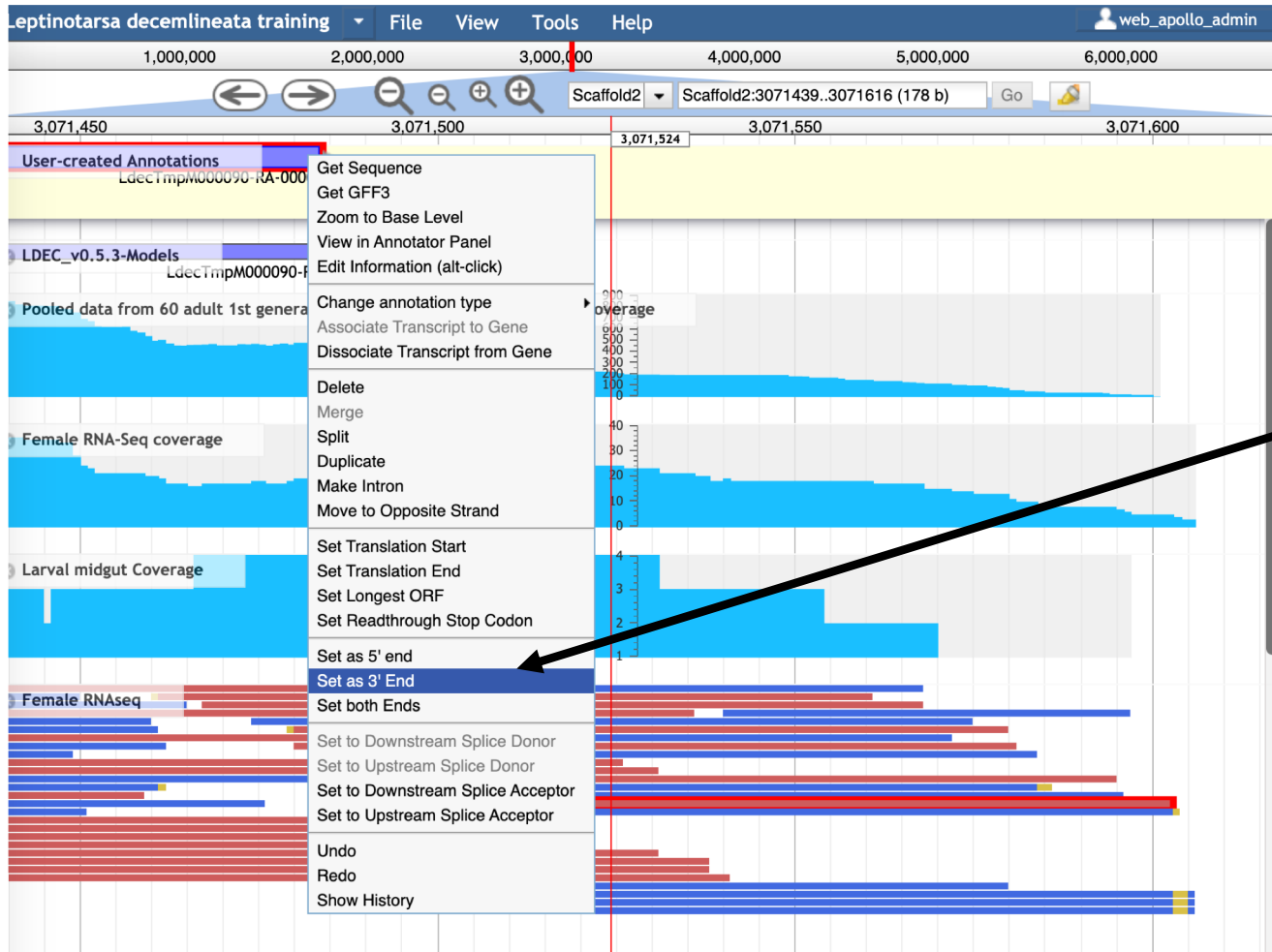


Adjusting gene boundaries



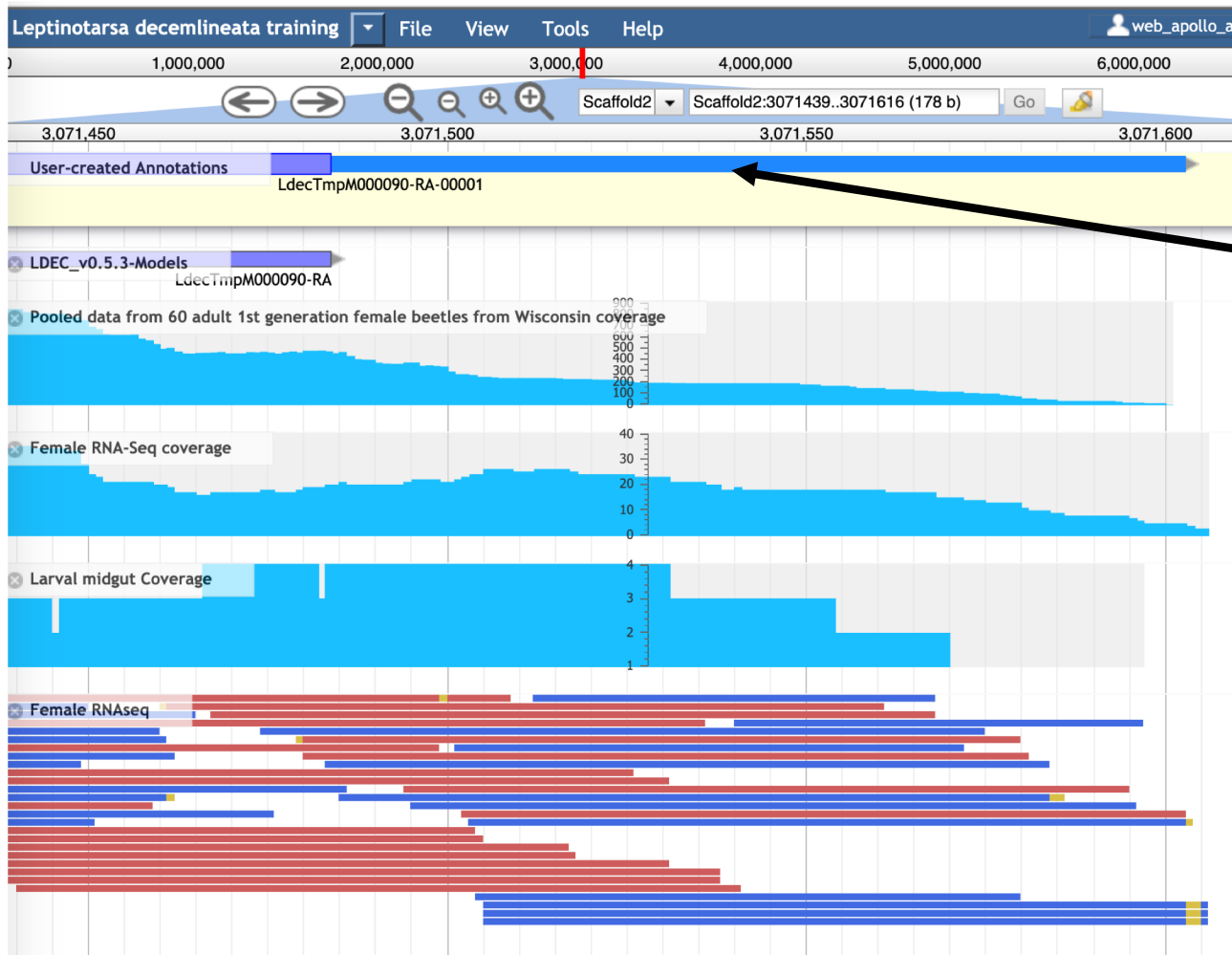
One way to change the boundary: find a mapped read on the same strand as the model; hold shift and click on the read and the model to highlight them both

Adjusting gene boundaries



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

Leptinotarsa decemlineata training | File | View | Tools | Help | web_apollo_admin

Information Editor

Tag	Value
-----	-------

Add Delete

PubMed IDs

Add Delete

Gene Ontology IDs

Add Delete

Comments

Added 3' UTR based on Female RNA-Seq

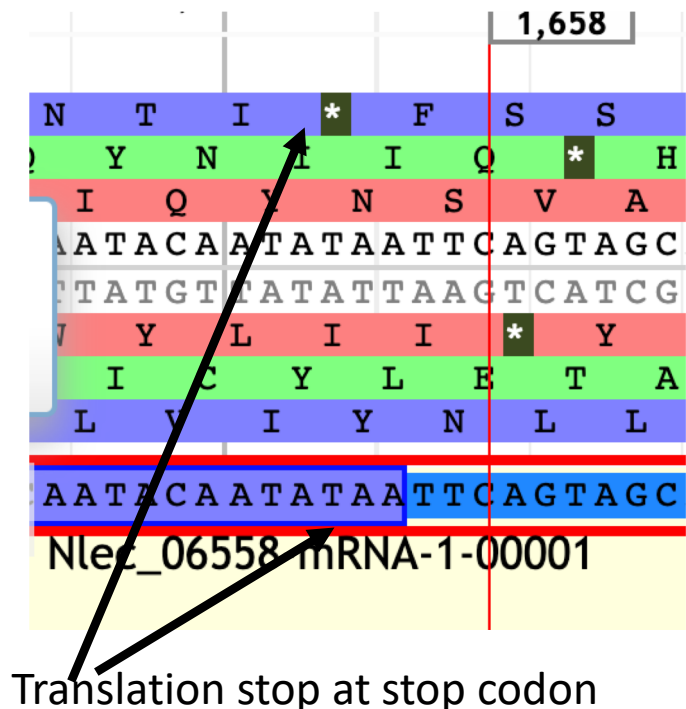
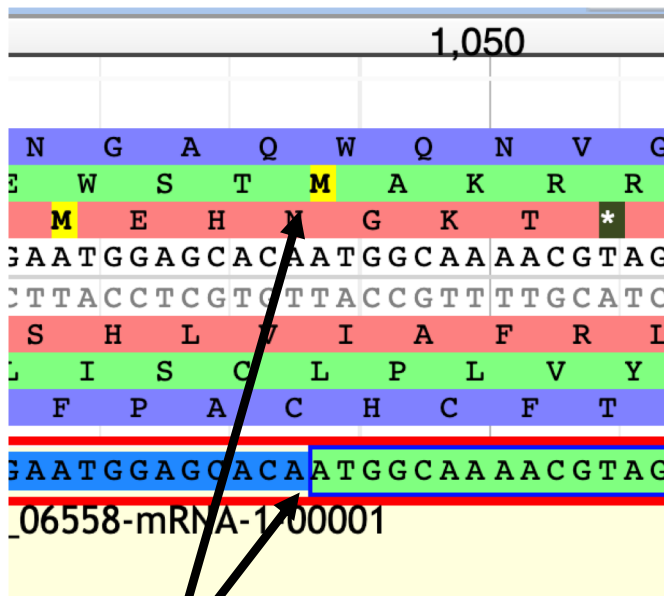
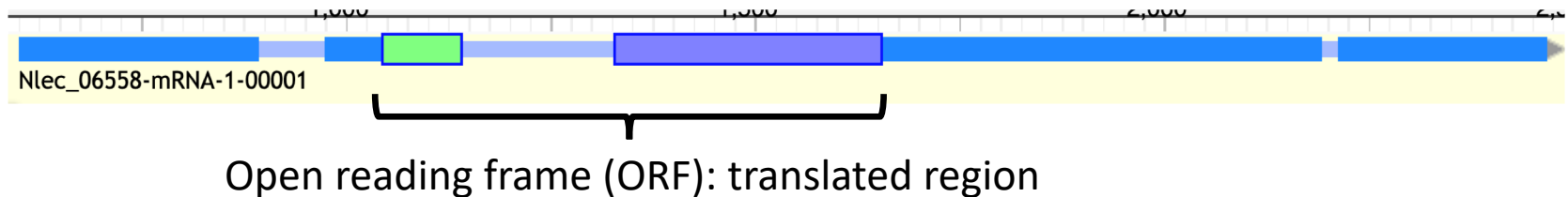
Add comment explaining
UTR addition

Starts, stops, open
reading frames

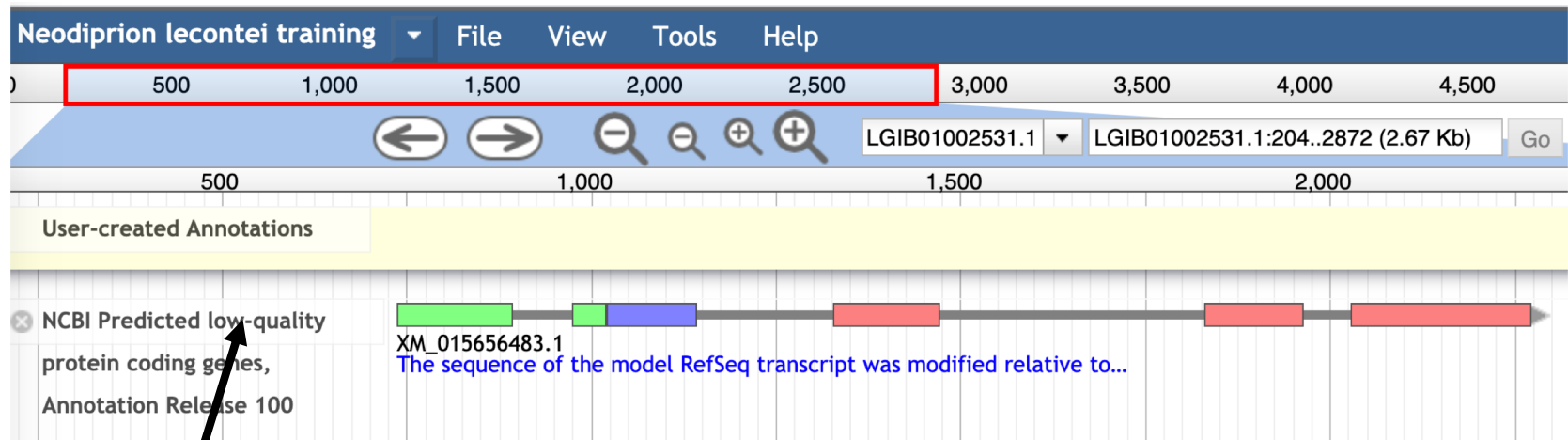
Setting the sequence start, stop, and open reading frame (ORF)

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

Starts, stops, ORFs

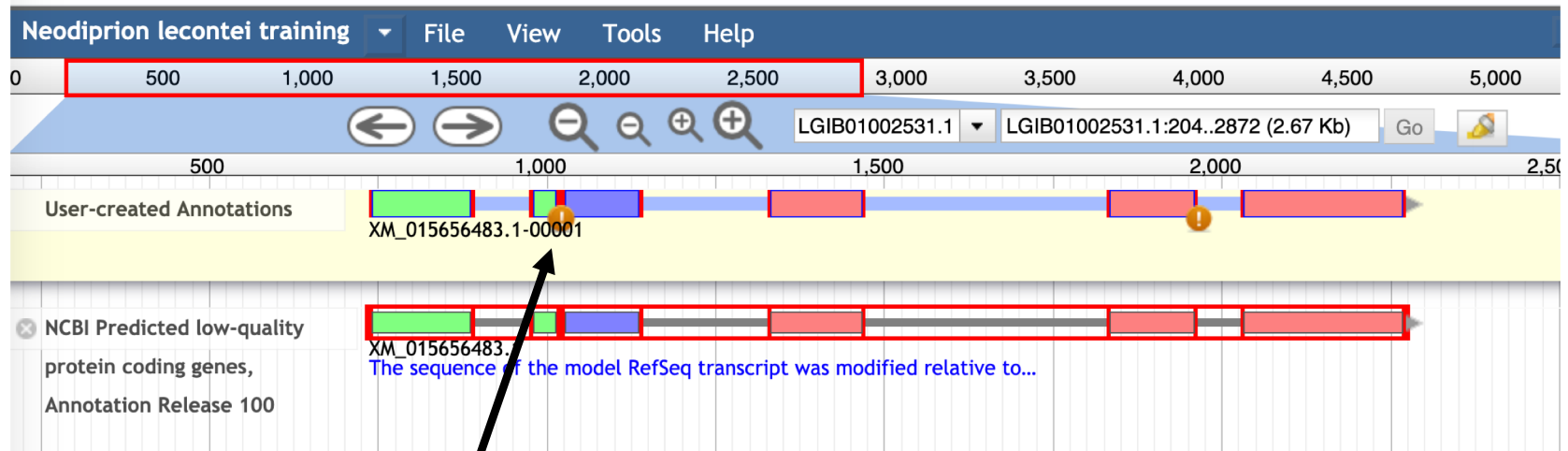


Starts, stops, ORFs



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 * 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

GTACGCCTGAATTGATATTATCCACGCTATCTCAGATACGGATGCCAACTACTGCGATACGCAATTGAGGCCCGGAGAGCTGTATAAGGTTGCTAGAAATATTGACTGCCTCTGATTCATGGAAAATGT?

CATGCGGACTTAAGTATAATAGGTGCGATAGAGTCTATGCCTACGGTTATGACGCTATGCGTAACTCCGGGCCTCTCGACATATTCCAACGATCTTATAACTGACGGAGACTAAGTACCTTTTACAT

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 Y 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

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

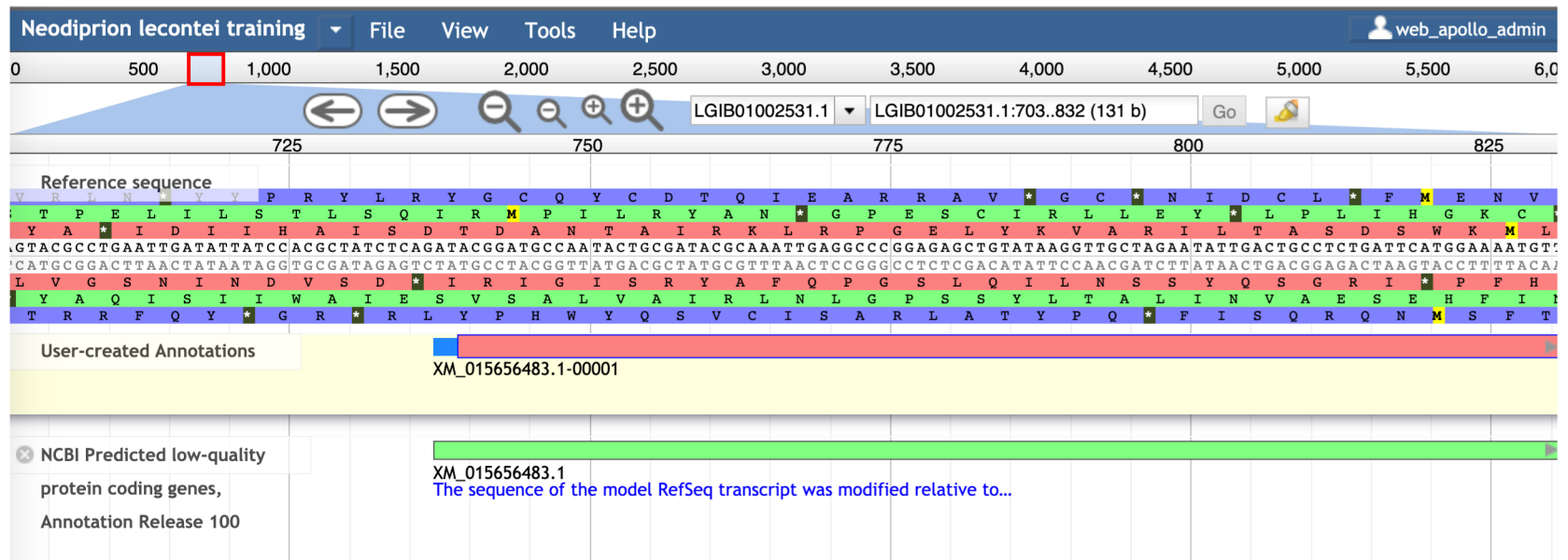
☒ Peptide sequence
☐ cDNA sequence
☐ CDS sequence
☐ Genomic sequence
☐ Genomic sequence +/- 500 bases

Starts, stops, ORFs

The screenshot shows the 'Neodiprion lecontei training' interface. At the top, there's a menu bar with 'File', 'View', 'Tools', and 'Help'. Below it is a scale from 0 to 3,000 with a red box highlighting the 1,000 mark. A reference sequence is displayed with nucleotides color-coded by reading frame: blue (1st), green (2nd), and pink (3rd). The sequence is: V R L N Y Y P R Y L R Y G C Q Y C D T Q I. The 3rd reading frame (pink) is highlighted, and a red box is drawn around the 'UcA' (TCA) codon at position 739. A right-click context menu is open over this codon, with the option 'Set Translation Start' highlighted in blue. Other options in the menu 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 End', 'Set Longest ORF', 'Set Readthrough Stop Codon', and 'Set as 5' end'.

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



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

Starts, stops, ORFs

Sequence

>f1107d76-0b75-45c4-a0bb-d082c7dc52bd (sequence:mRNA) 282 residues [LGIB01002531.1:737-2273 + strand] [peptide]

TDANTAIRKL R P G E L Y K V A R I L T A S D S W K M L M A I V P K D G I E N V P K F S T E H F K L I E Q A S R Q Q R K A A E I F L E
E W S T M A K R R P T L Q S T L I L L V K V H L L K A A D Y I A V D L L H G Q P P Q R L V S G P A A P V I I S D K E I E M L L D E T A S E H
G E Q L T Y R I L E P P E N L N L E L K S S E L S V V V A K L G N G I R T M Y S E L P Q V V A E I G K R E S P N P T E L P S S S N S N T H A
M N T Y D N Y T N D N I L I C L N A L L R S L E N Y E L I T A E L P Q I V A D L G R 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

That looks better.

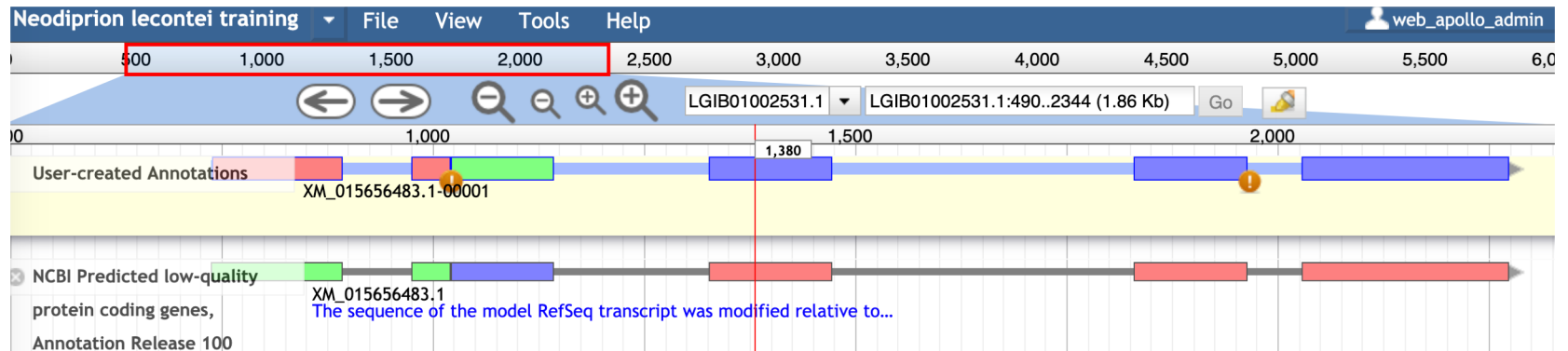
☒ Peptide sequence
☐ cDNA sequence
☐ CDS sequence
☐ Genomic sequence
☐ Genomic sequence +/- bases

Starts, stops, ORFs

The screenshot displays the 'Neodiprion lecontei training' software interface. At the top, a menu bar includes 'File', 'View', 'Tools', and 'Help'. Below this is a genomic track with a scale from 0 to 3,500. A red box highlights the scale from 500 to 2,000. A context menu is open over a green annotation bar labeled 'XM_015656'. The menu options are: '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' (highlighted in blue), and 'Set Readthrough Stop Codon'. The track also shows 'User-created Annotations' and 'NCBI Predicted low-quality protein coding genes, Annotation Release 100'.

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

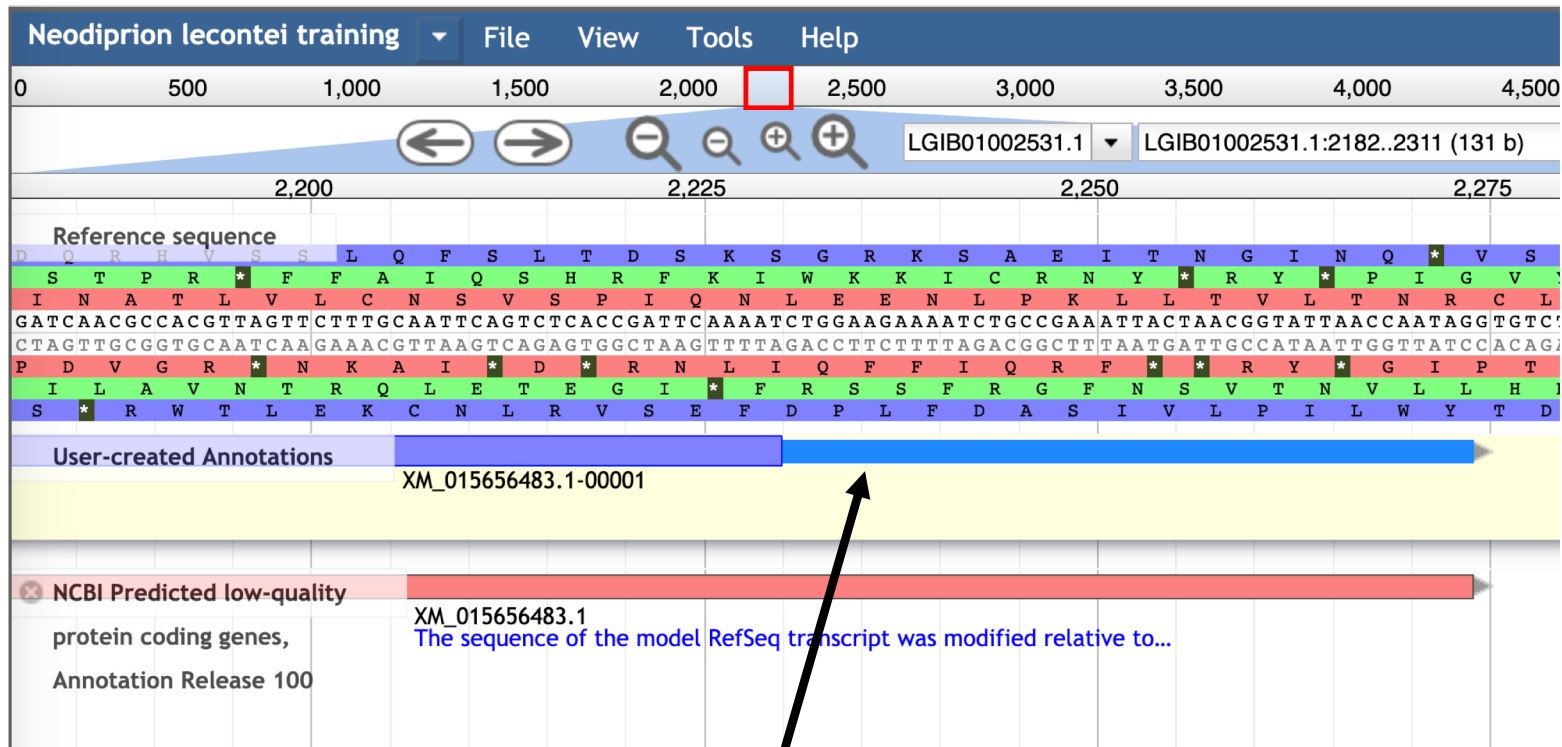
```
>9190e15d-dcee-45c8-9236-5d7babfca448 (sequence:mRNA) 282 residues [LGIB0100253 strand] [peptide]
TDANTAIRKLRPGELYKVARILTASDSWKMLMAIVPKDGIENVPKFSTEHFKLIEQASRQQRKAAEIFLE
EWSTMAKRRPTLQSTLILLVKVHLLKAADYIAVDLLHGQPPQRLVSGPAAPVIISDKEIEMLLDETASEH
GEQLTYRILEPPENLNLELKSSSELSVVVAKLGNIGRTMYSELPQVVAEIGKRESPNPTELPSSSNSNTHA
MNTYDNYTNDNILICLNALLRSLNELYELITAELPQIVADLGRDQRHVSSLQFSLTDSKSGRKS AEITNGI
NQ
```

Starts, stops, ORFs

The screenshot shows the Neodiprion lecontei training interface. The top menu bar includes 'File', 'View', 'Tools', and 'Help'. Below the menu is a scale from 0 to 4,000. The main area displays a reference sequence with nucleotide positions 2,200, 2,225, 2,229, 2,250, and 2,275. The sequence is color-coded by codon: green for start codons (ATG), red for stop codons (TAA, TAG, TGA), and blue for other codons. A user-created annotation is shown below the reference sequence, with the text 'XM_015656483.1-00001'. A right-click context menu is open over the sequence, with the following options: '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', and 'Set Largest ORF'. The 'Set Translation End' option is highlighted in blue.

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



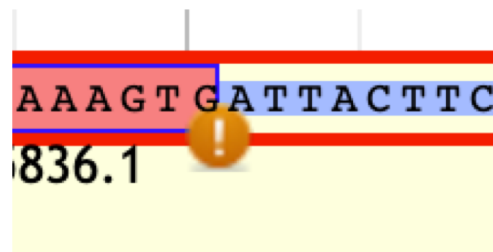
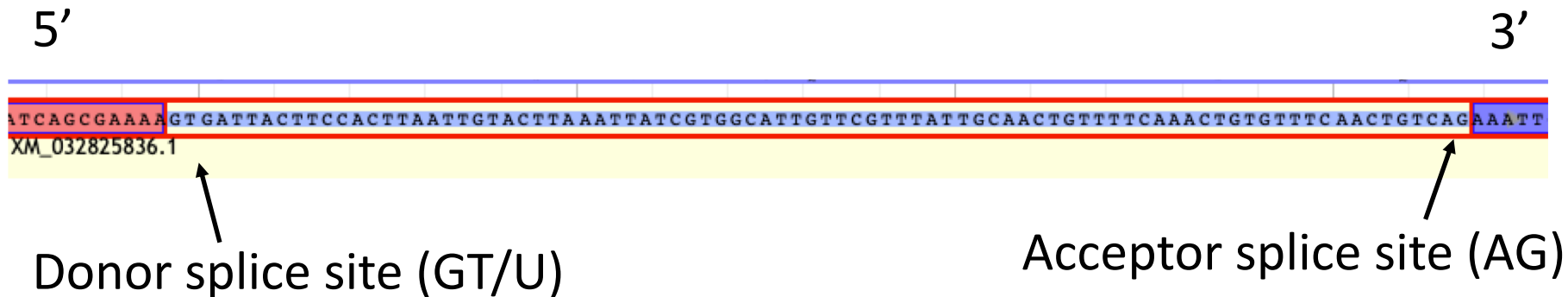
Now the sequence after the translation end is 3' UTR.

Non-canonical splice sites

Splice sites

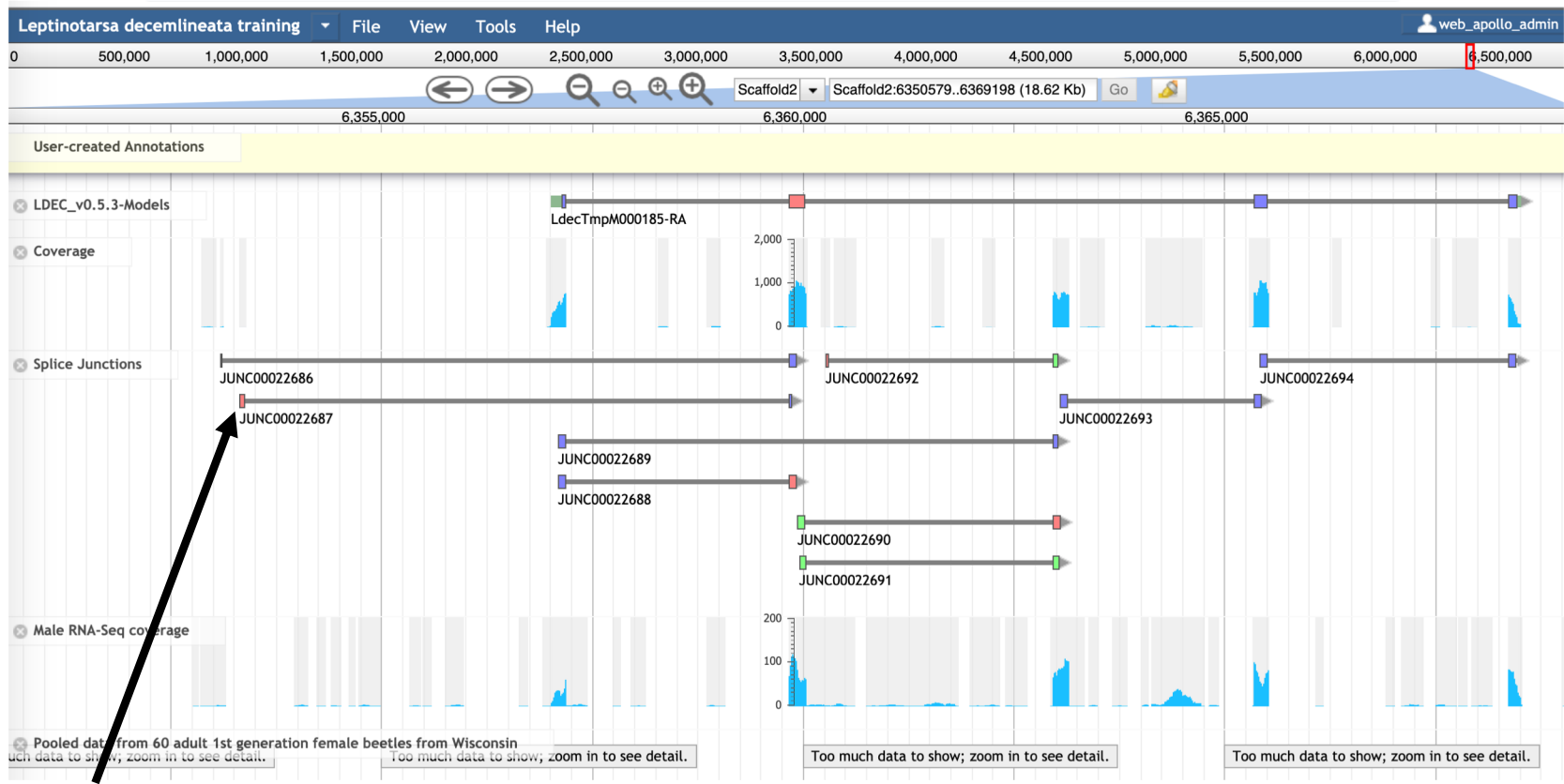
Introns are removed from primary transcripts by cleavage at conserved sequences called **splice sites**. These sites are found at the 5' and 3' ends of introns.

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



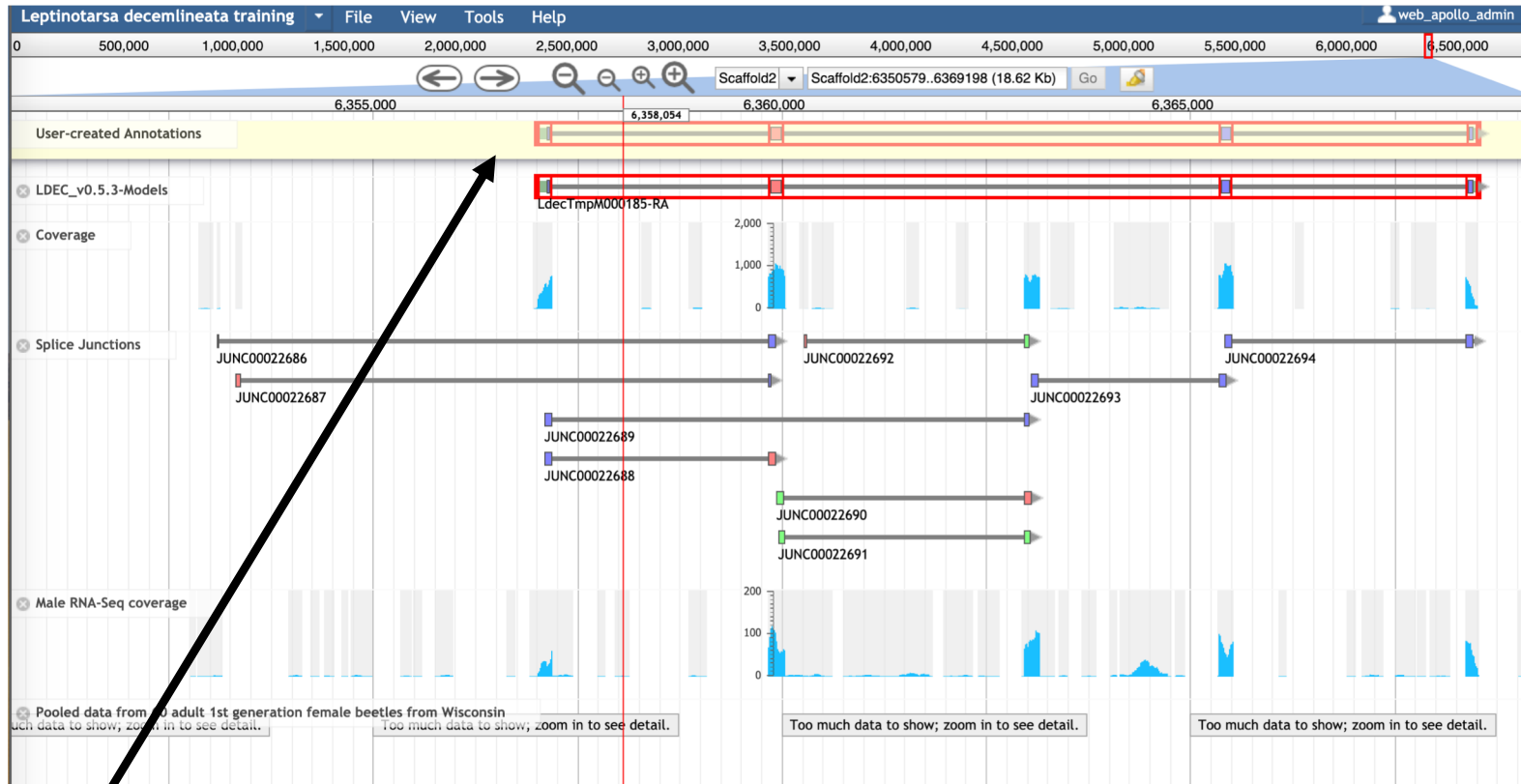
'Non-canonical' splice sites – non-conserved, and possibly erroneous sites – are marked by an exclamation point in Apollo.

Fixing non-canonical splice sites



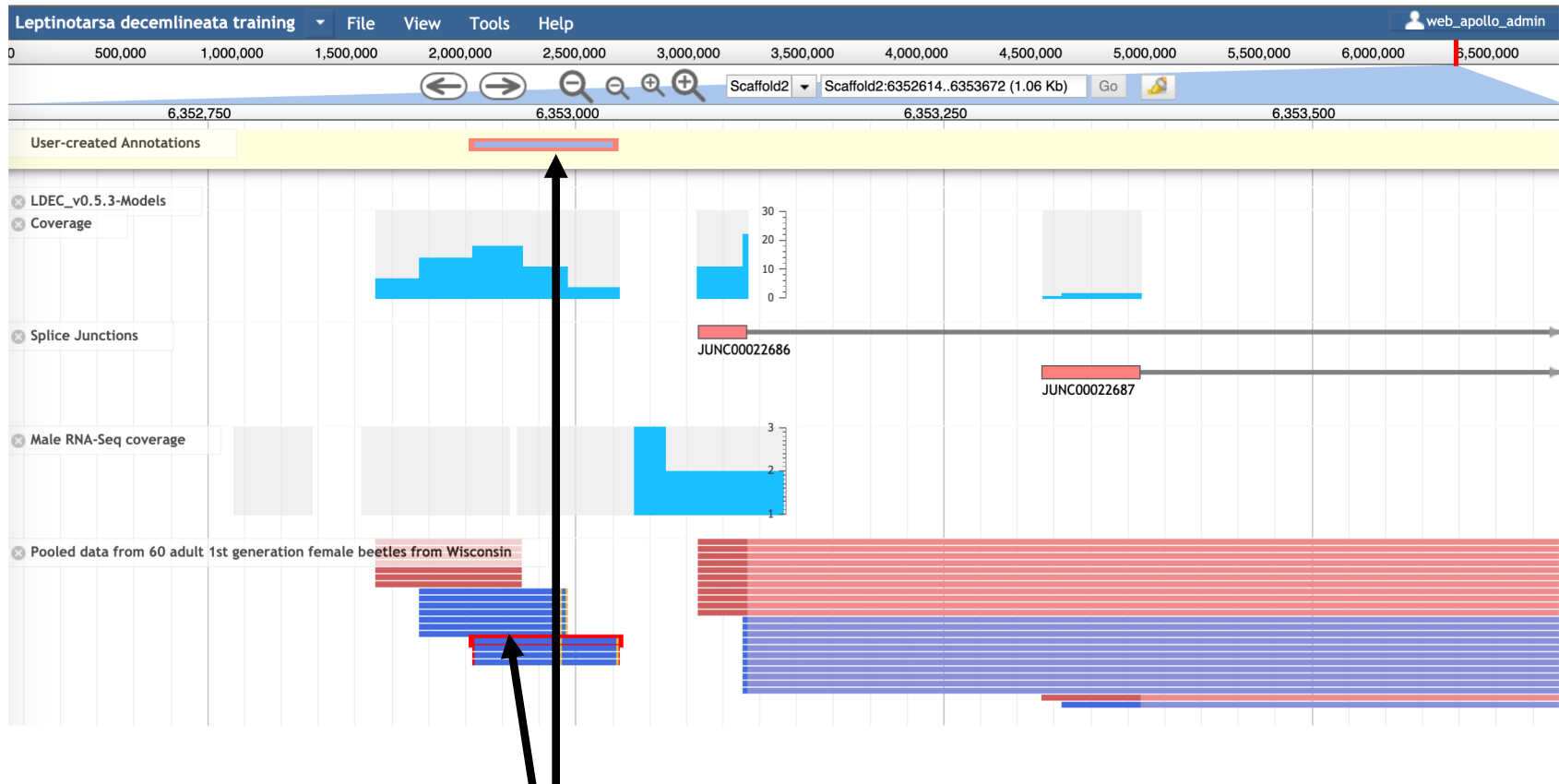
Looks like there's another isoform here – let's add an exon

Fixing non-canonical splice sites



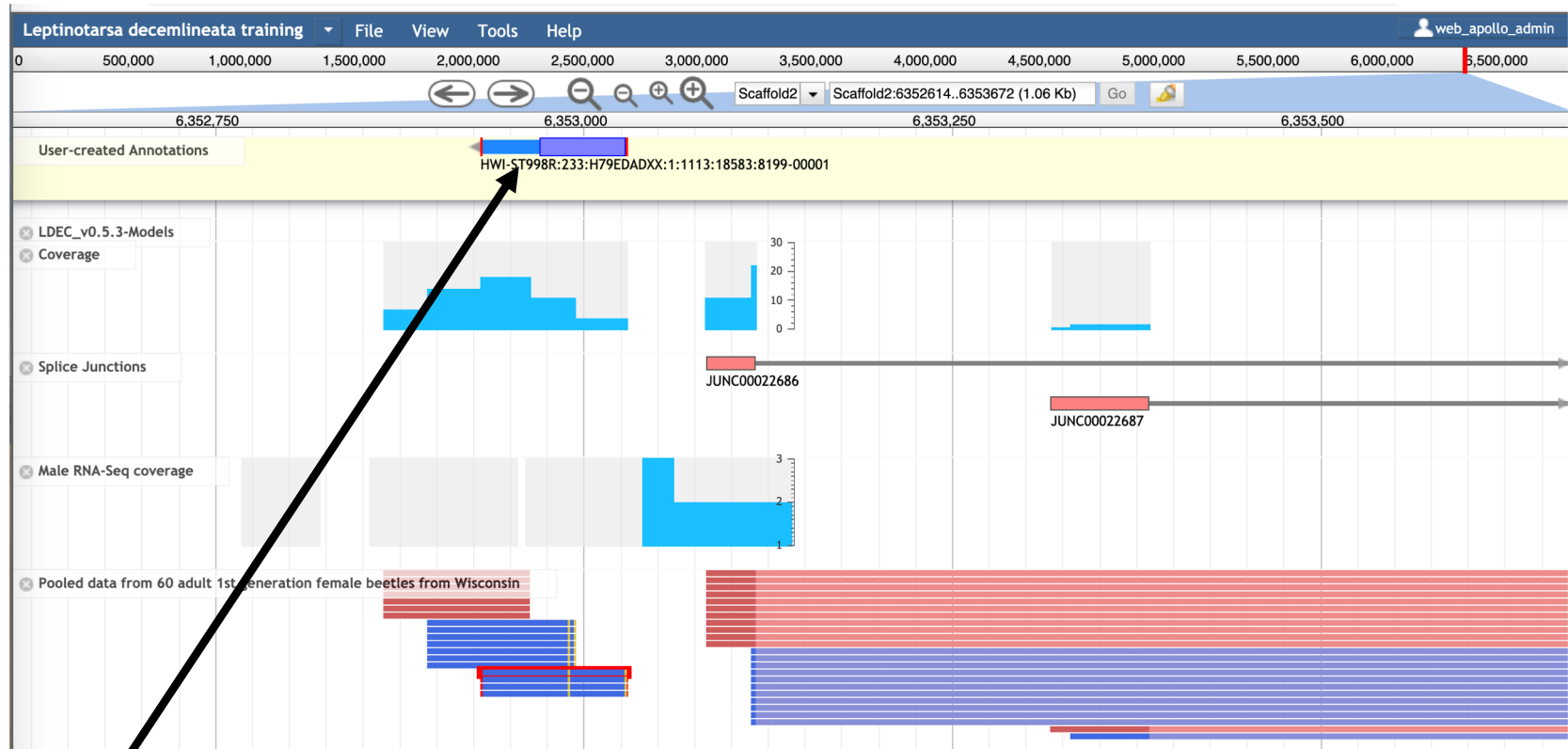
Drag model to UcaA track

Fixing non-canonical splice sites



Drag evidence to Uca track to add 5' exon

Fixing non-canonical splice sites



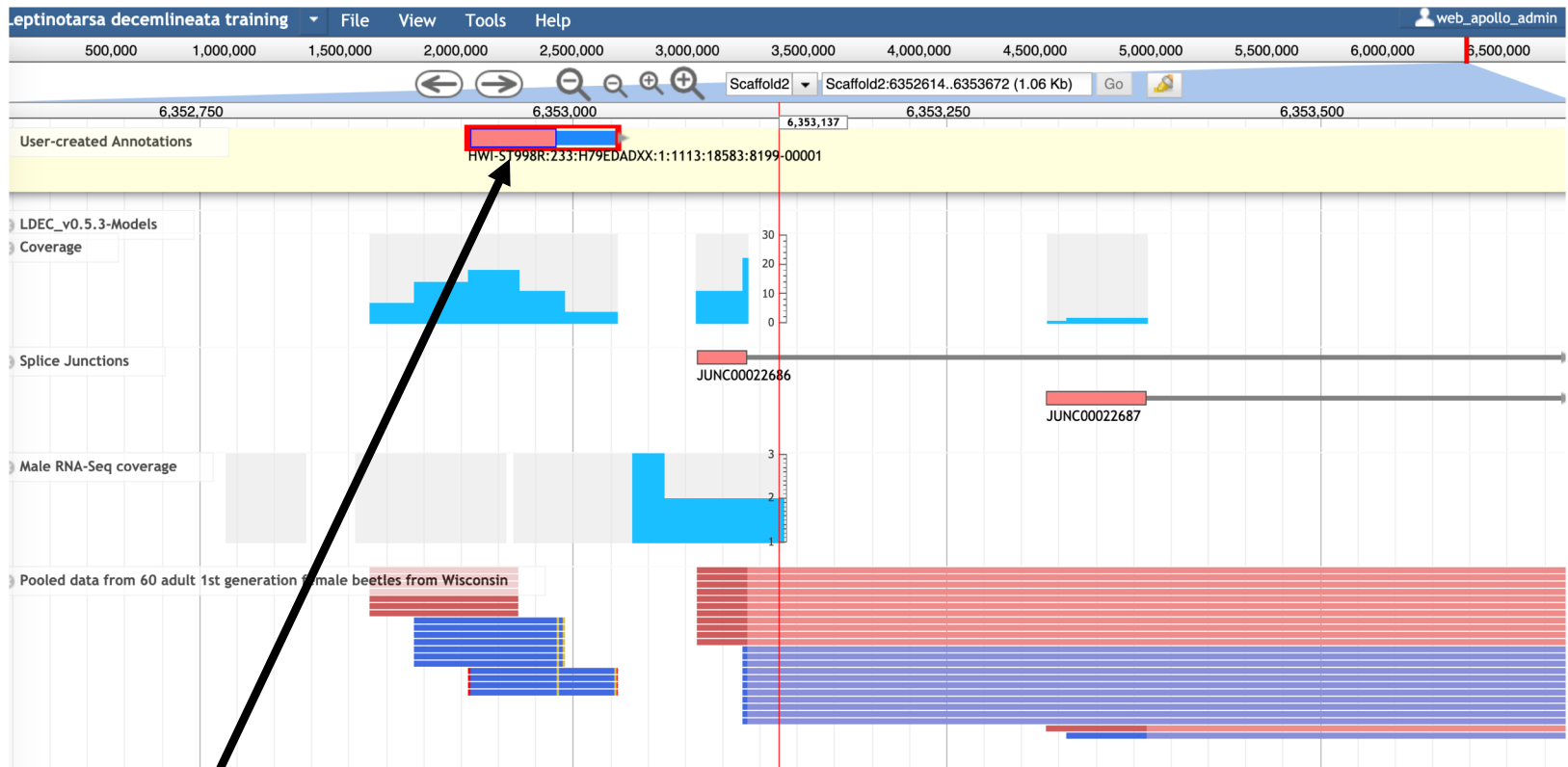
Oops, wrong strand – our model is on the forward strand

Fixing non-canonical splice sites

The screenshot shows the Leptinotarsa decemlineata training interface. The top menu bar includes File, View, Tools, and Help. The main window displays a genomic track with various annotations. A context menu is open over a splice junction, showing 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, 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. A black arrow points from the text 'Let's flip it to the opposite strand' to the 'Move to Opposite Strand' option.

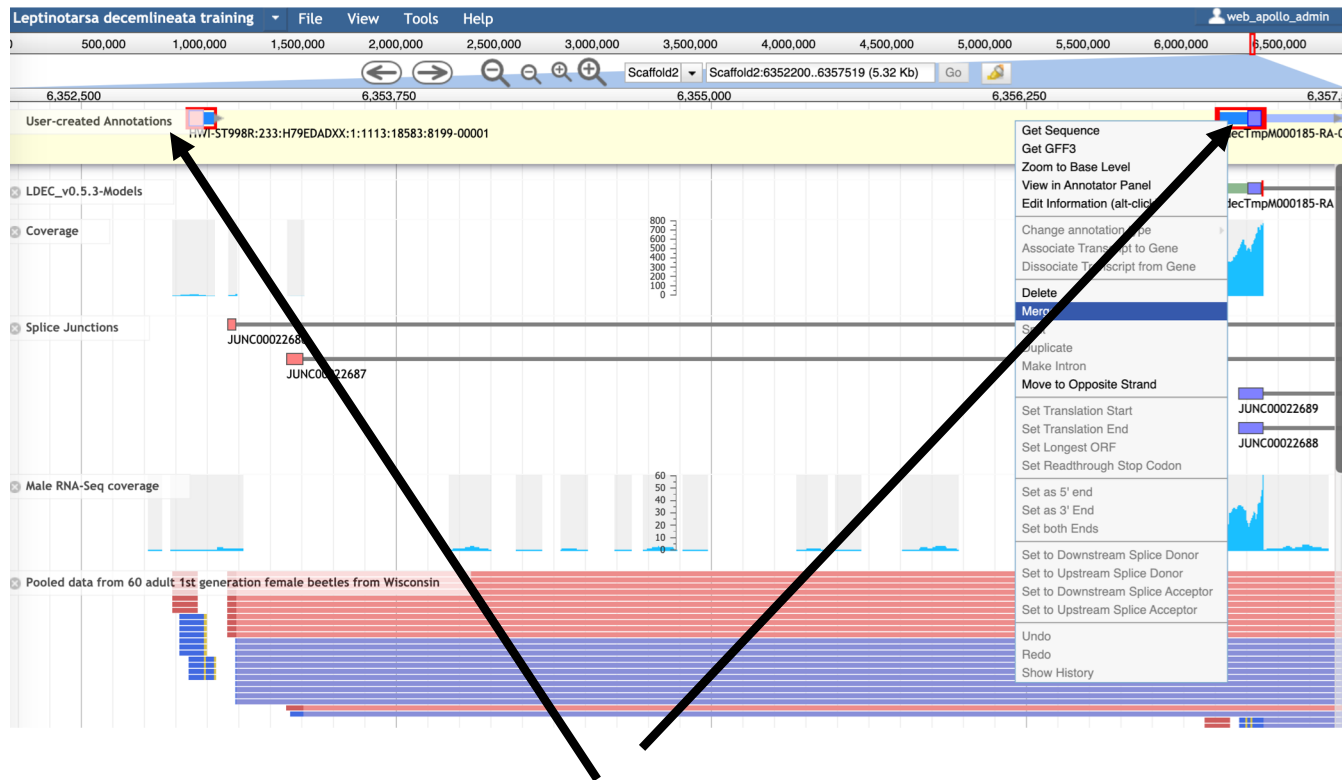
Let's flip it to the opposite strand

Fixing non-canonical splice sites



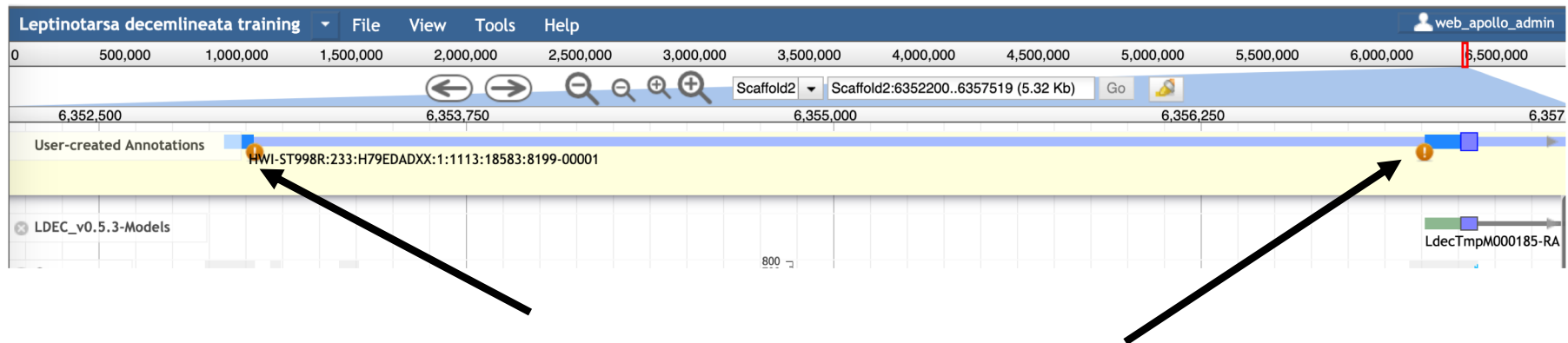
That's better.

Fixing non-canonical splice sites



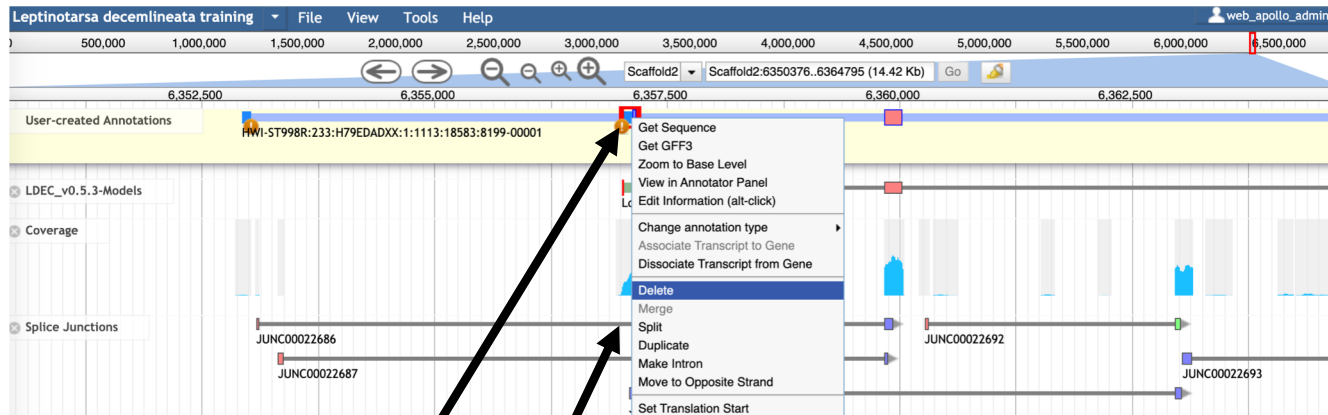
Merge the new exon to the gene model

Fixing non-canonical splice sites



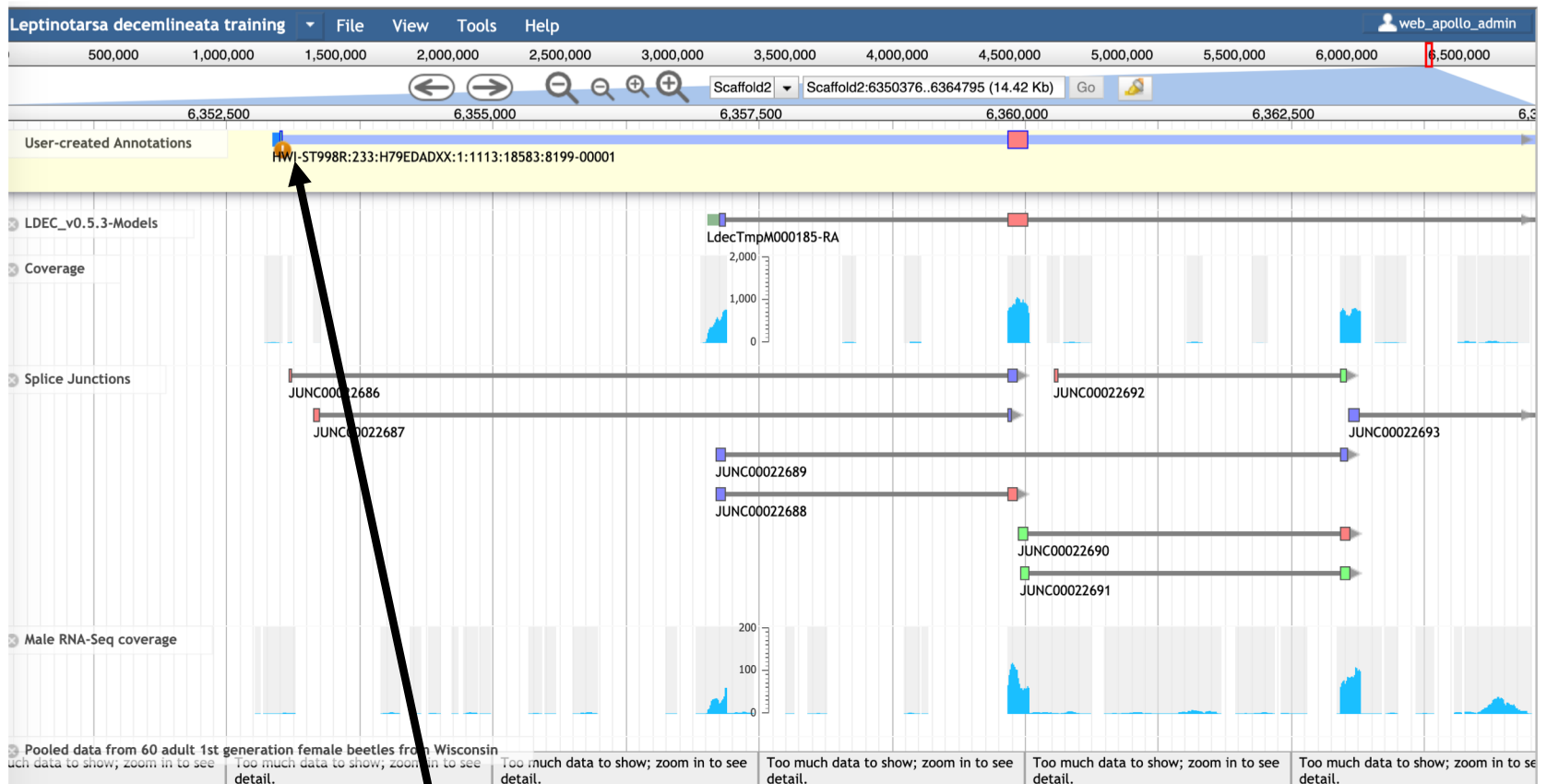
Non-canonical splice sites in merged model

Fixing non-canonical splice sites



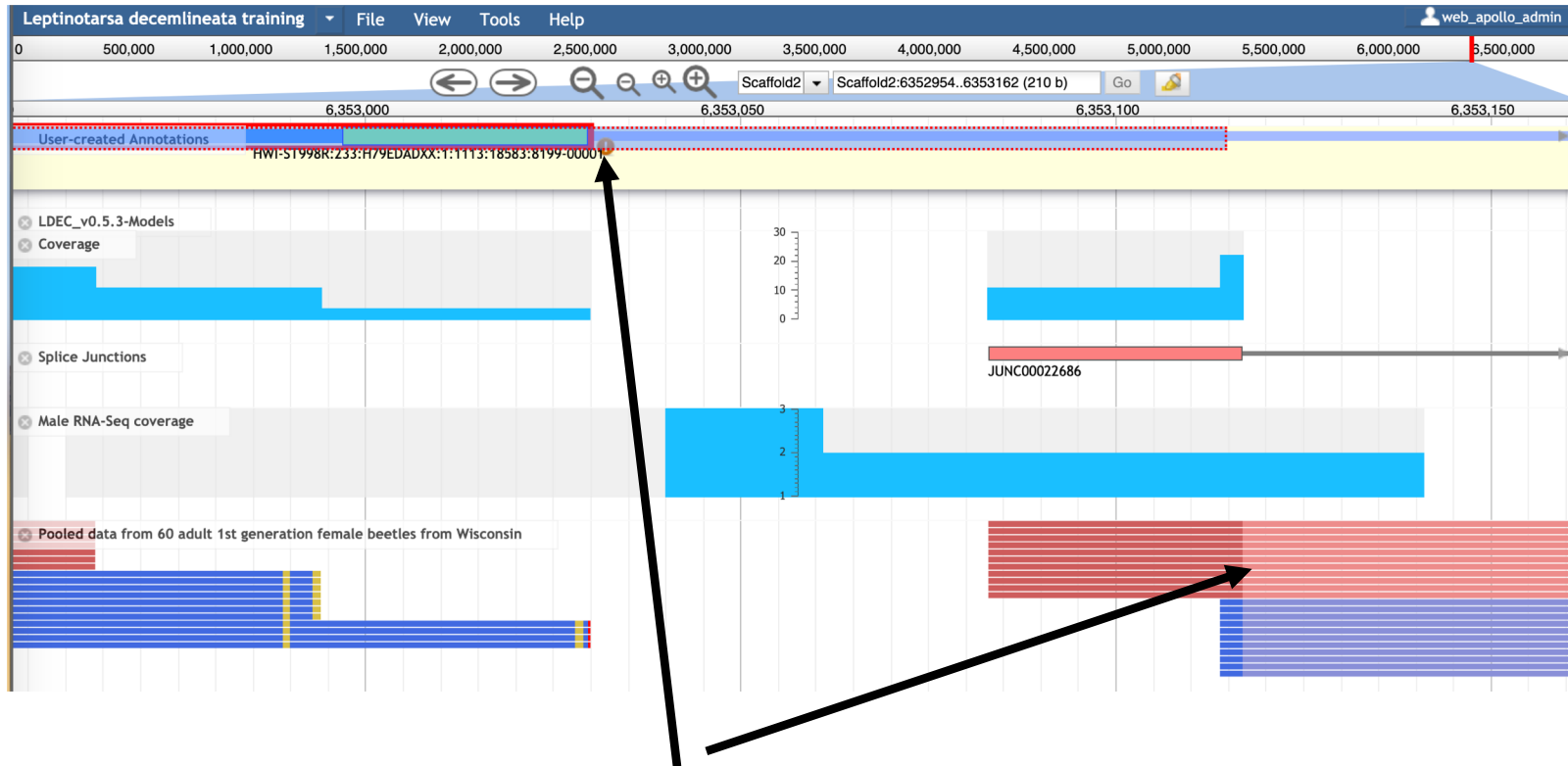
The splice junction reads don't support the 2nd exon with the new 5' exon, so let's remove it

Fixing non-canonical splice sites



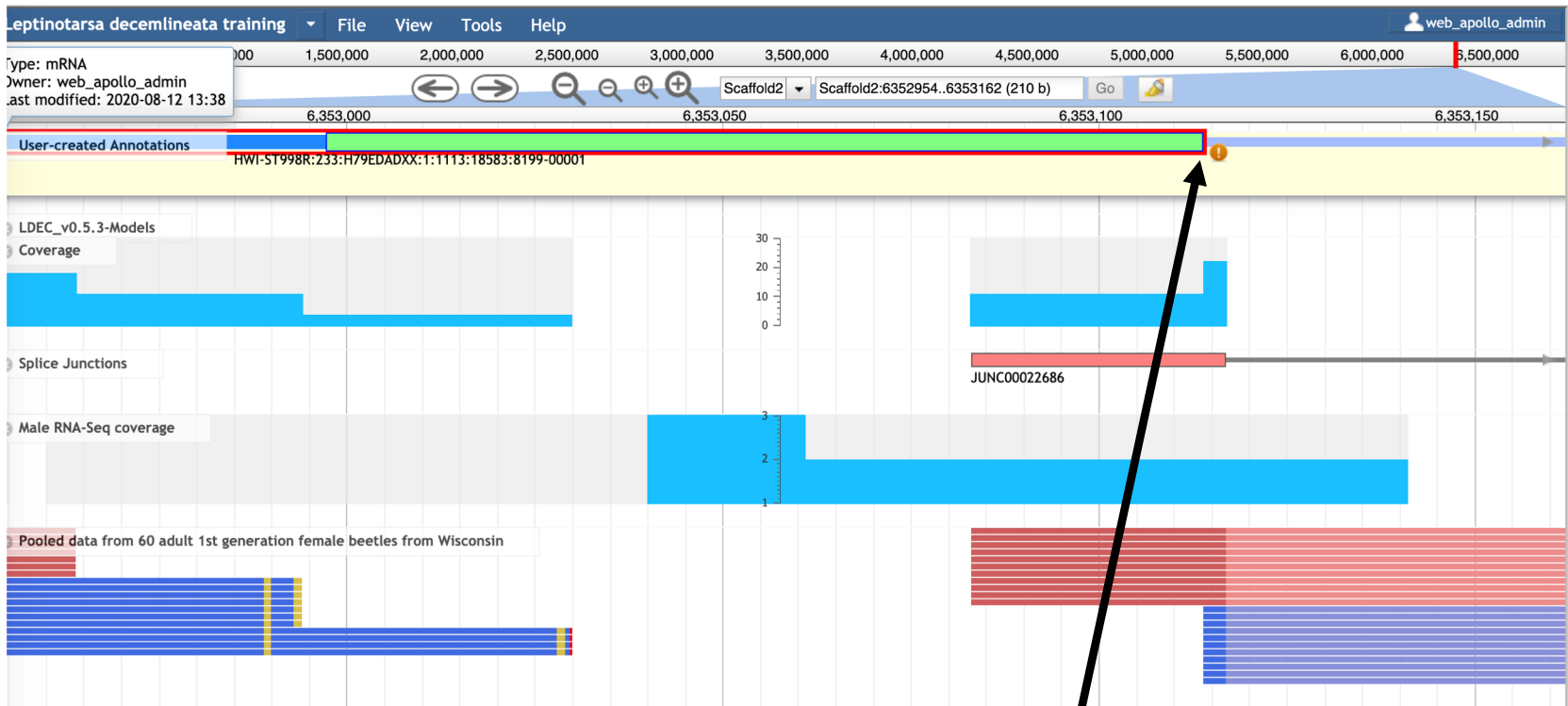
Only 1 non-canonical splice site left to fix

Fixing non-canonical splice sites



Extend exon to RNA-Seq boundary

Fixing non-canonical splice sites

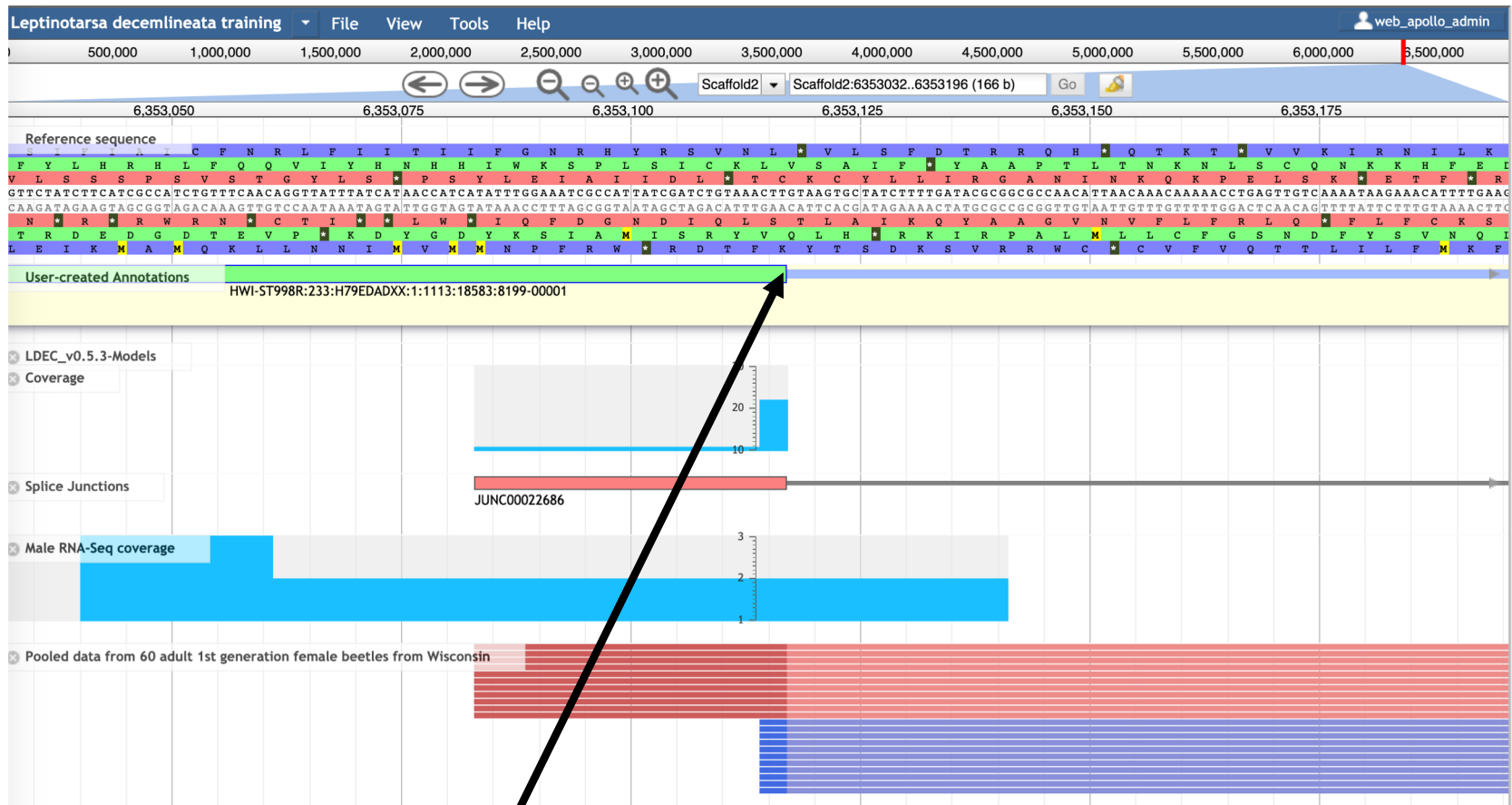


Still not quite right

The screenshot displays the Apollo genome browser interface. At the top, a coordinate bar shows positions from 0 to 5,500,000. Below this, the reference sequence for Scaffold2 is shown, with positions 6,353,050 to 6,353,175 highlighted. A context menu is open over a feature labeled 'User-created Annotations' (HWI-ST998R:Z33:H79EDADXX:1:1113). 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 submenu for Associate Transcript to Gene and 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 both Ends, Set to Downstream Splice Donor (highlighted by a black arrow), Set to Upstream Splice Donor, Set to Downstream Splice Acceptor, and Set to Upstream Splice Acceptor. The background shows various tracks including LDEC_v0.5.3-Models, Coverage, Splice Junctions, Male RNA-Seq coverage, and Pooled data from 60 adult 1st generation female beetles from Wis.



Fixing non-canonical splice sites



Fixed!

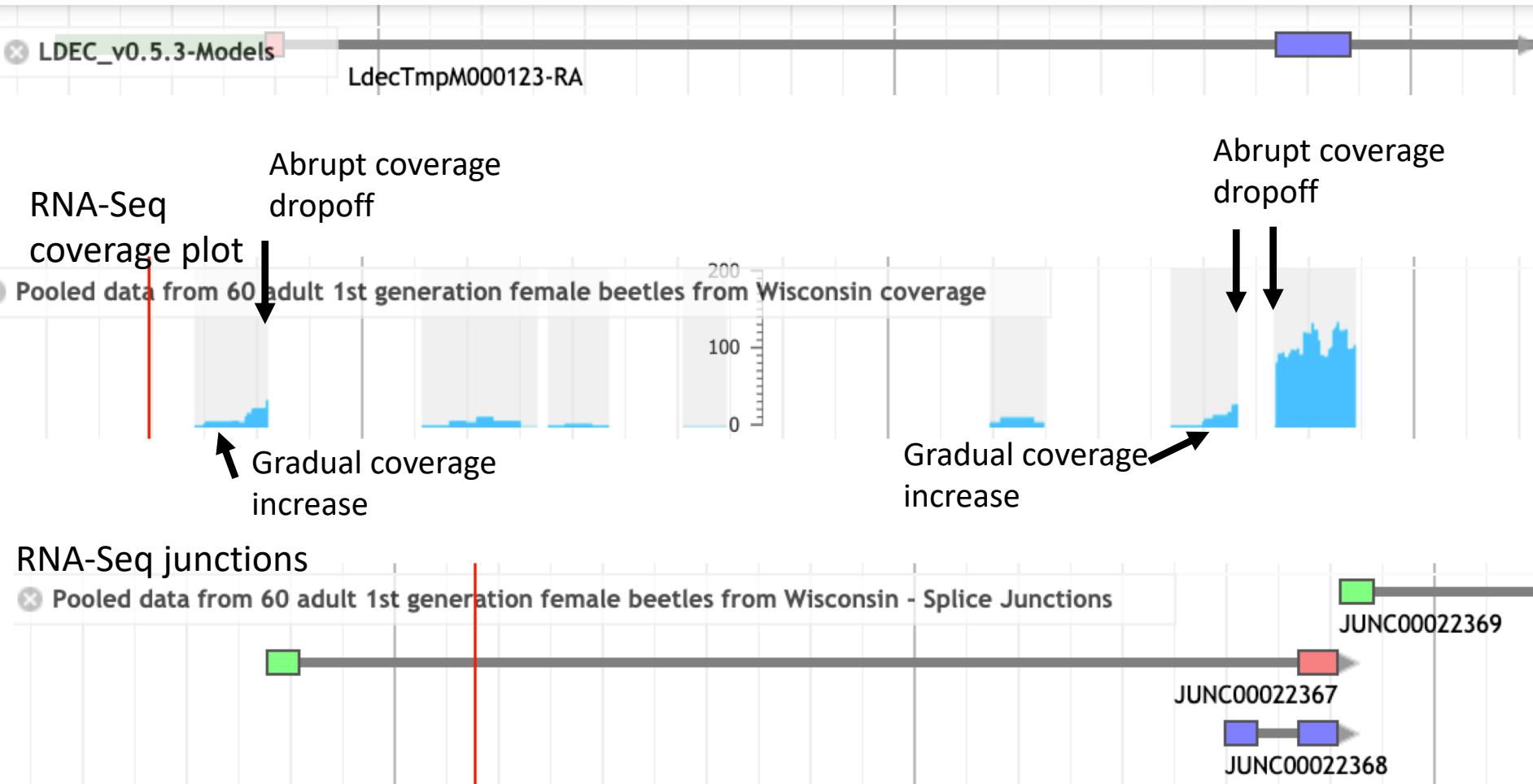
Annotating isoforms

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

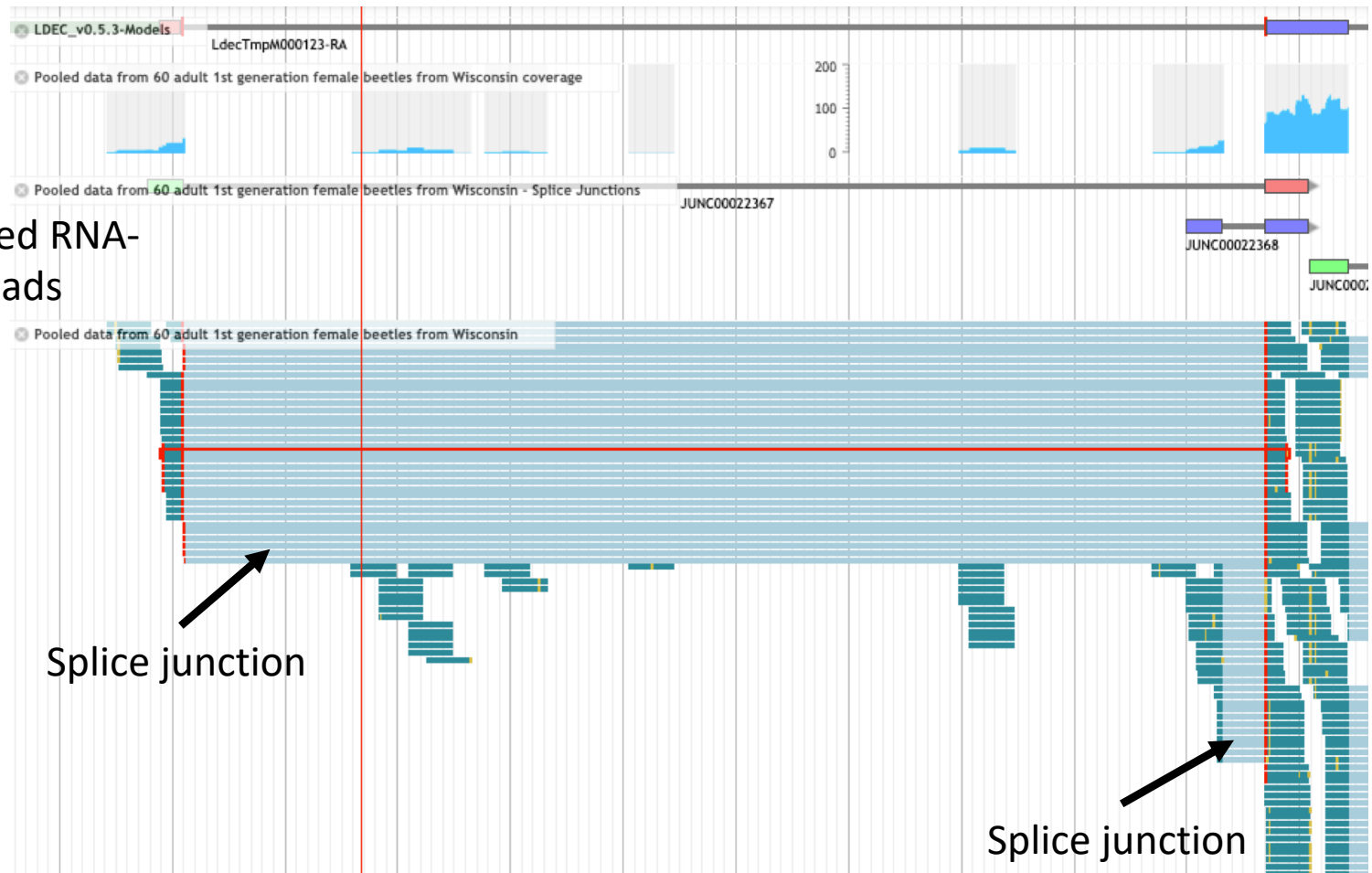
Isoform annotation example

5' end of MAKER tyrosine
protein kinase gene prediction



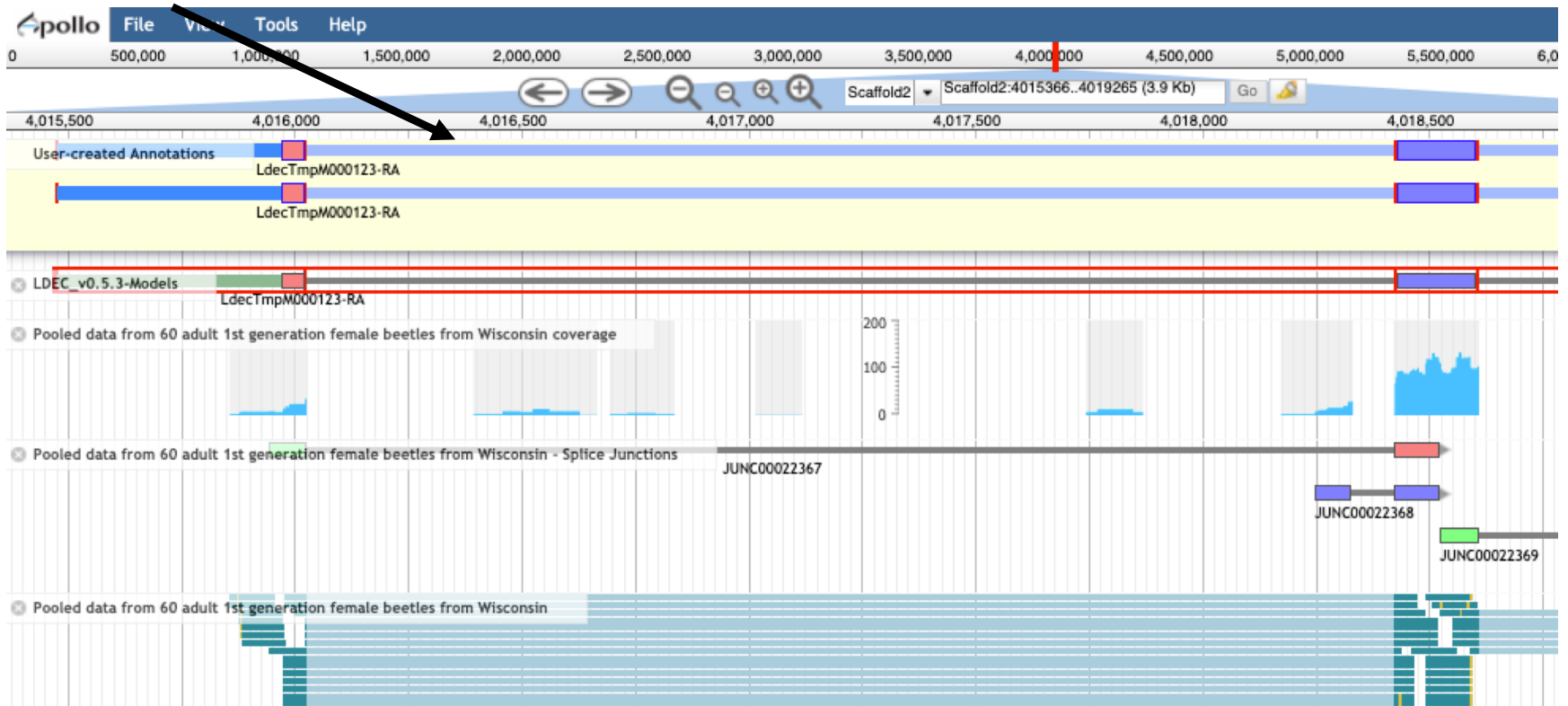
Isoform annotation example

Mapped RNA-Seq reads



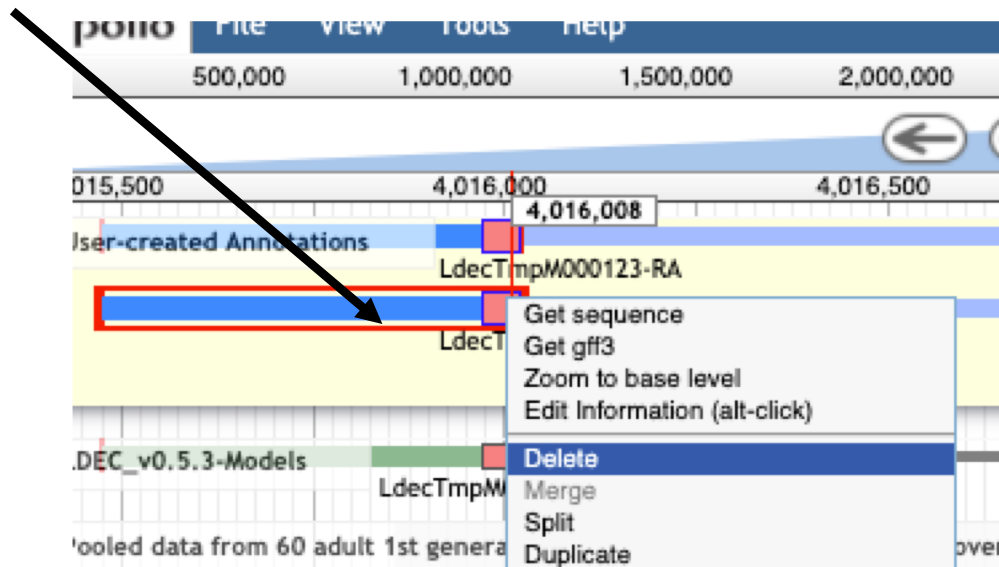
Isoform annotation example

Create 2 isoforms
from Maker model



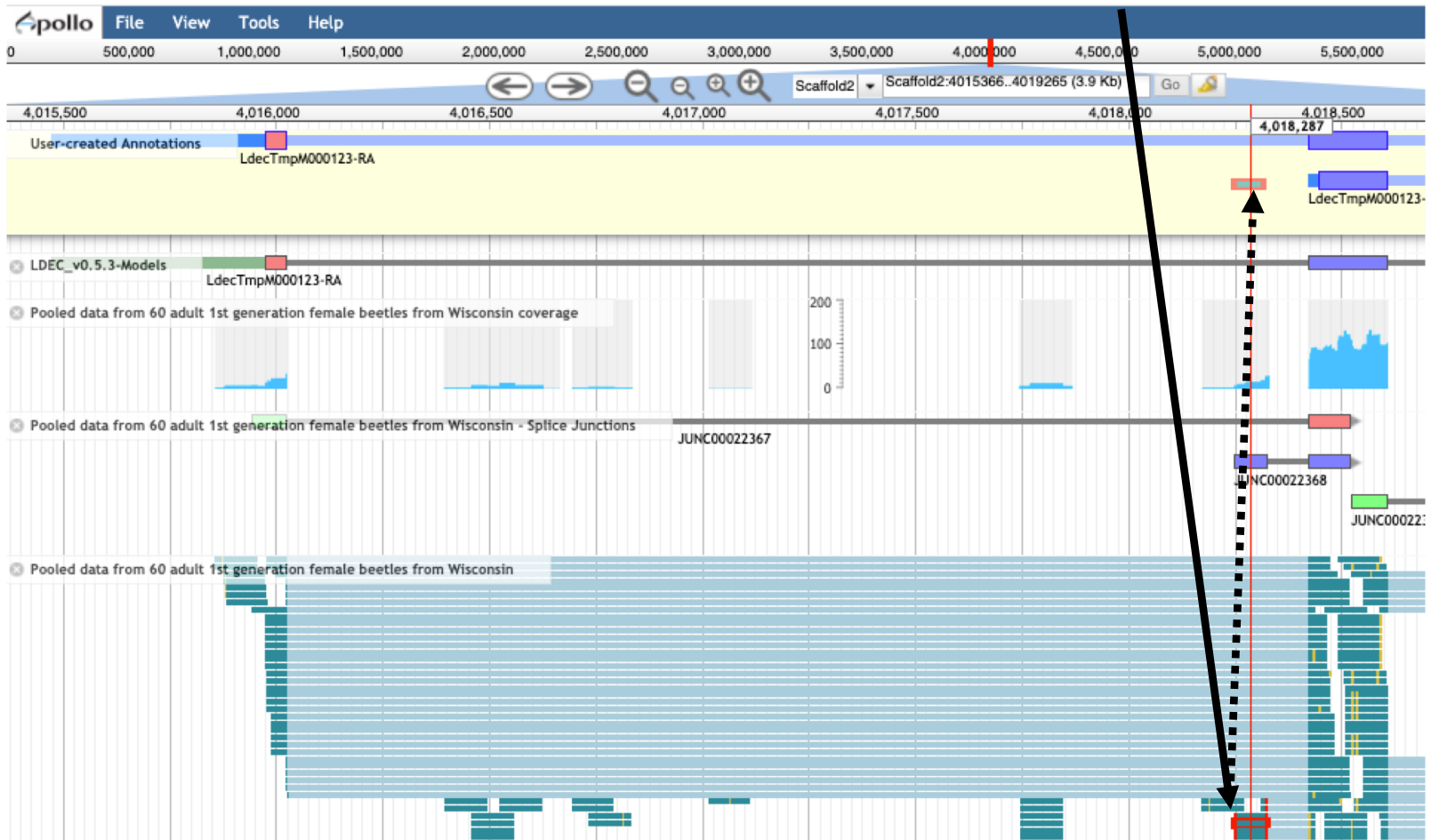
Isoform annotation example

Select and delete 5'
exon from one of
the isoforms



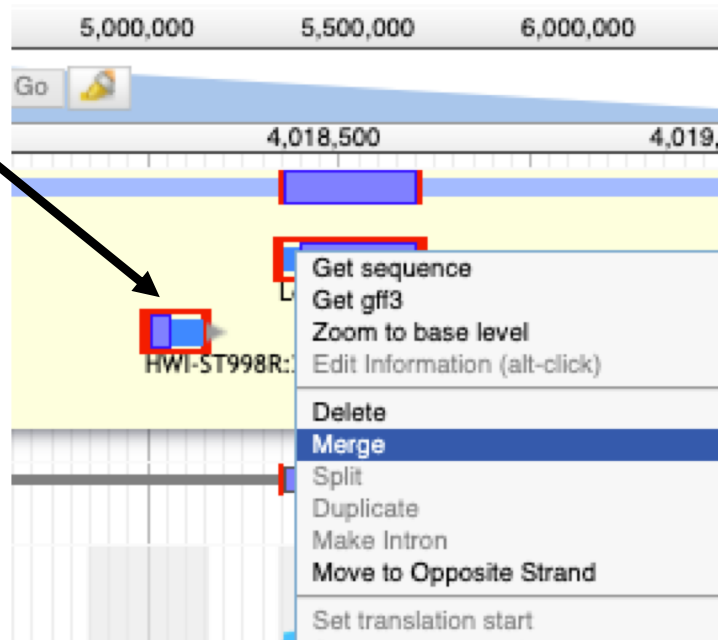
Isoform annotation example

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



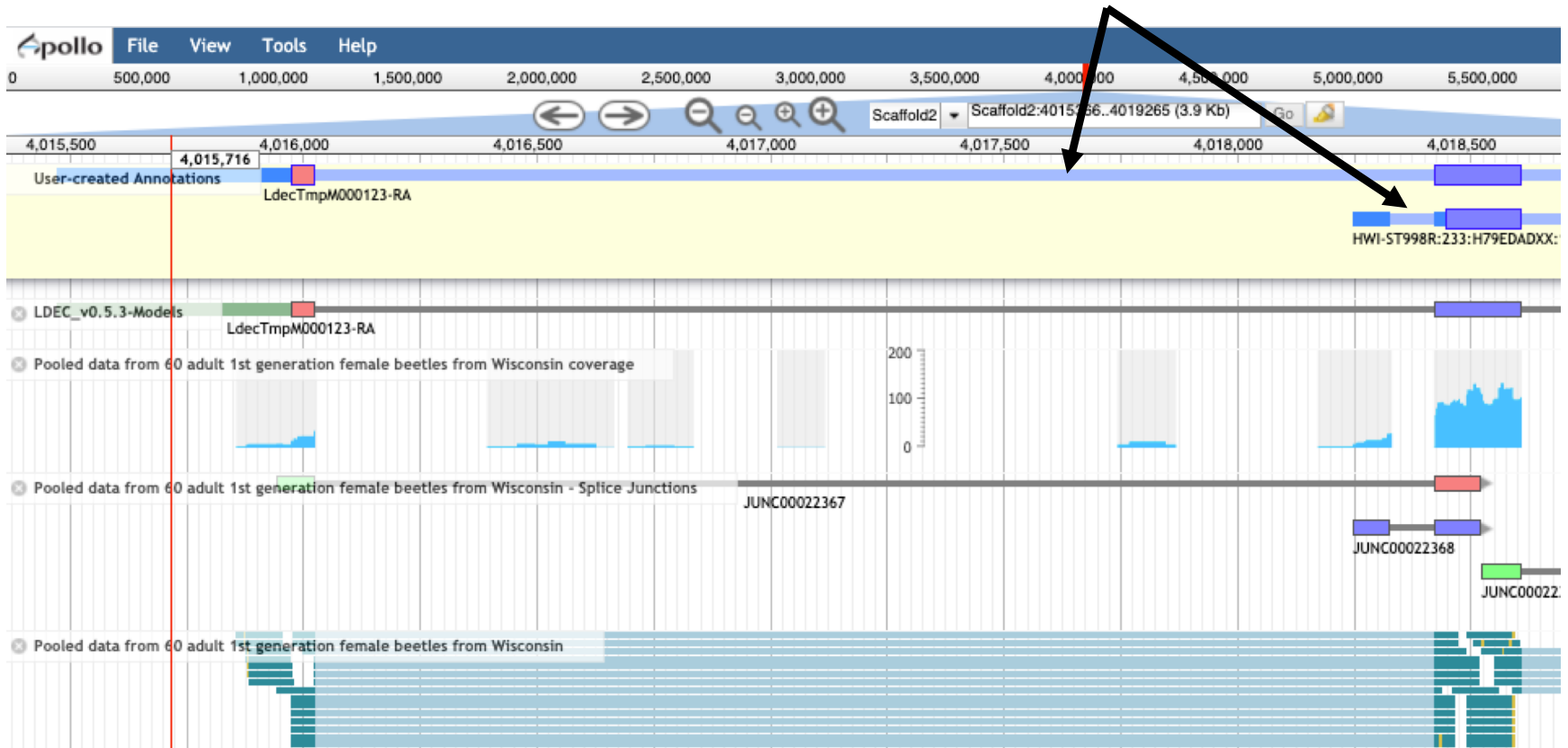
Isoform annotation example

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



Isoform annotation example

2 isoforms supported by RNA-Seq evidence



Sequence alterations and stop-codon readthroughs

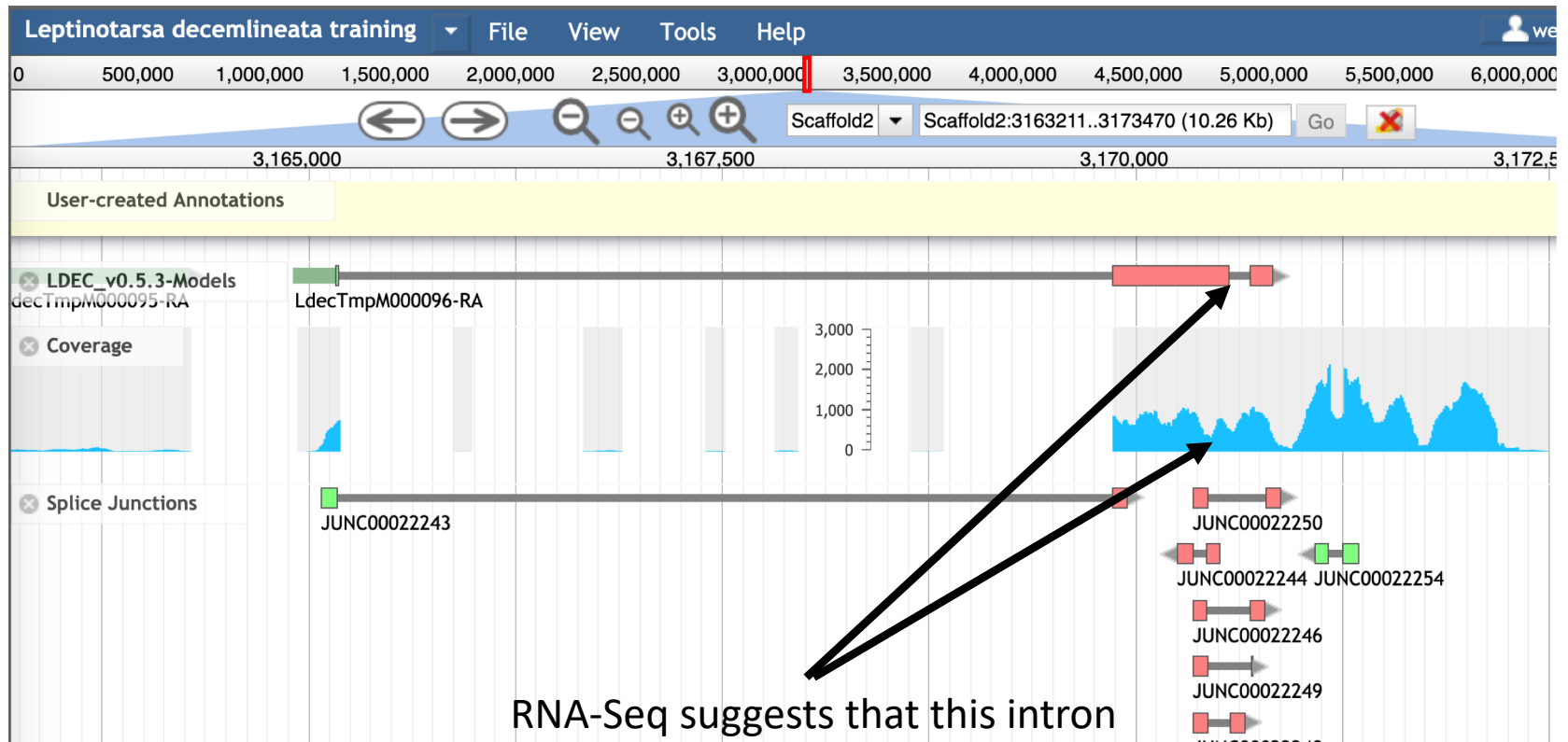
Sequence alterations

- Apollo supports annotating the genome sequence with insertions, deletions, and substitutions
- Note that this will not change the genome fasta in the sequence export – but it will allow Apollo to recalculate a gene model's sequence
- Only add a sequence alteration in Apollo if there is evidence for it – e.g. SNPs in mapped RNA-Seq

Stop-codon readthroughs

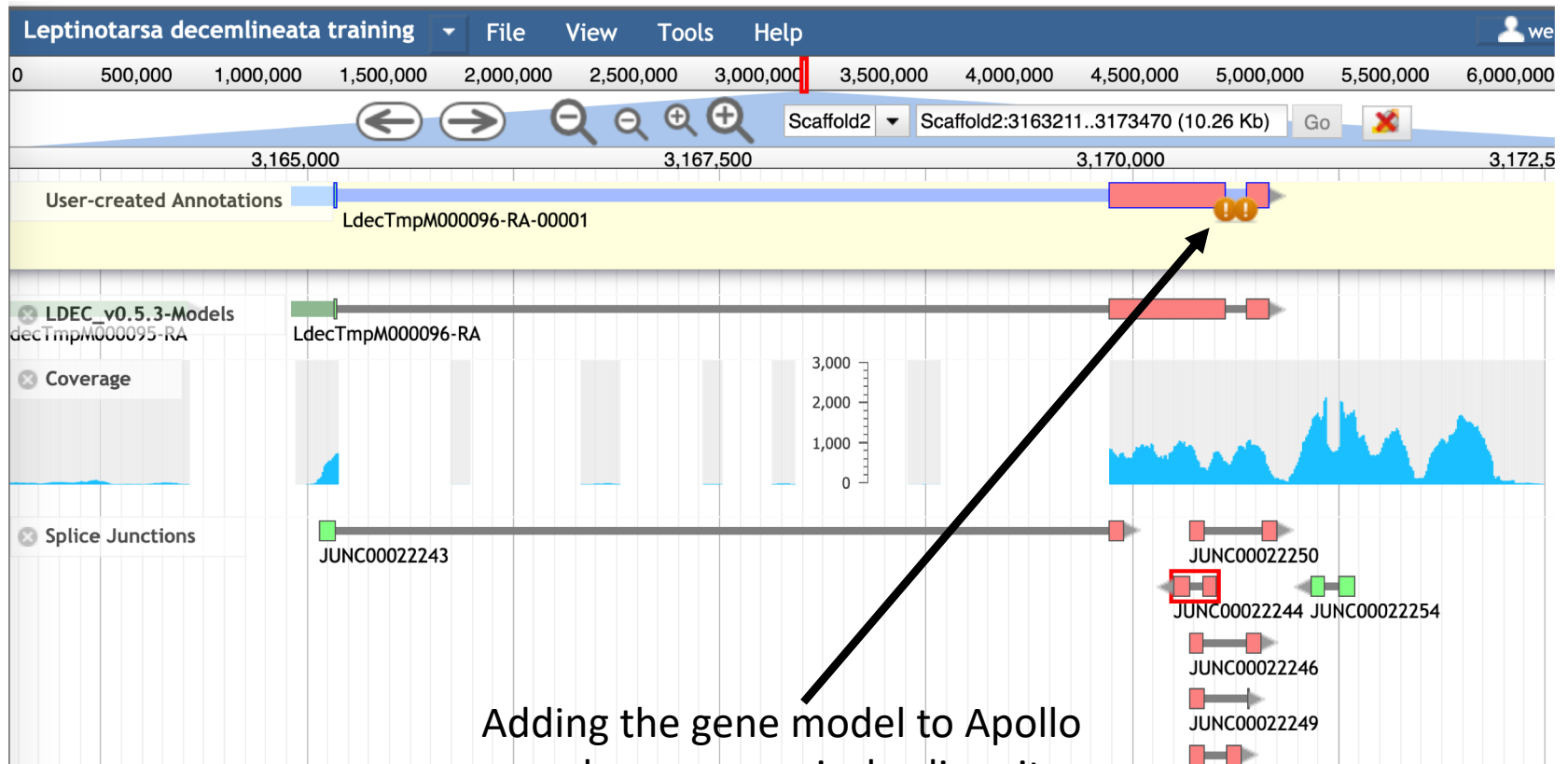
- Apollo allows you to annotate stop-codon readthrough features on the coding sequence of a gene model
- This is a special case for selenocysteine-containing proteins.
- This feature can be used in other cases – e.g. if you have evidence of errors in the genome assembly - but we don't recommend it

Sequence alterations

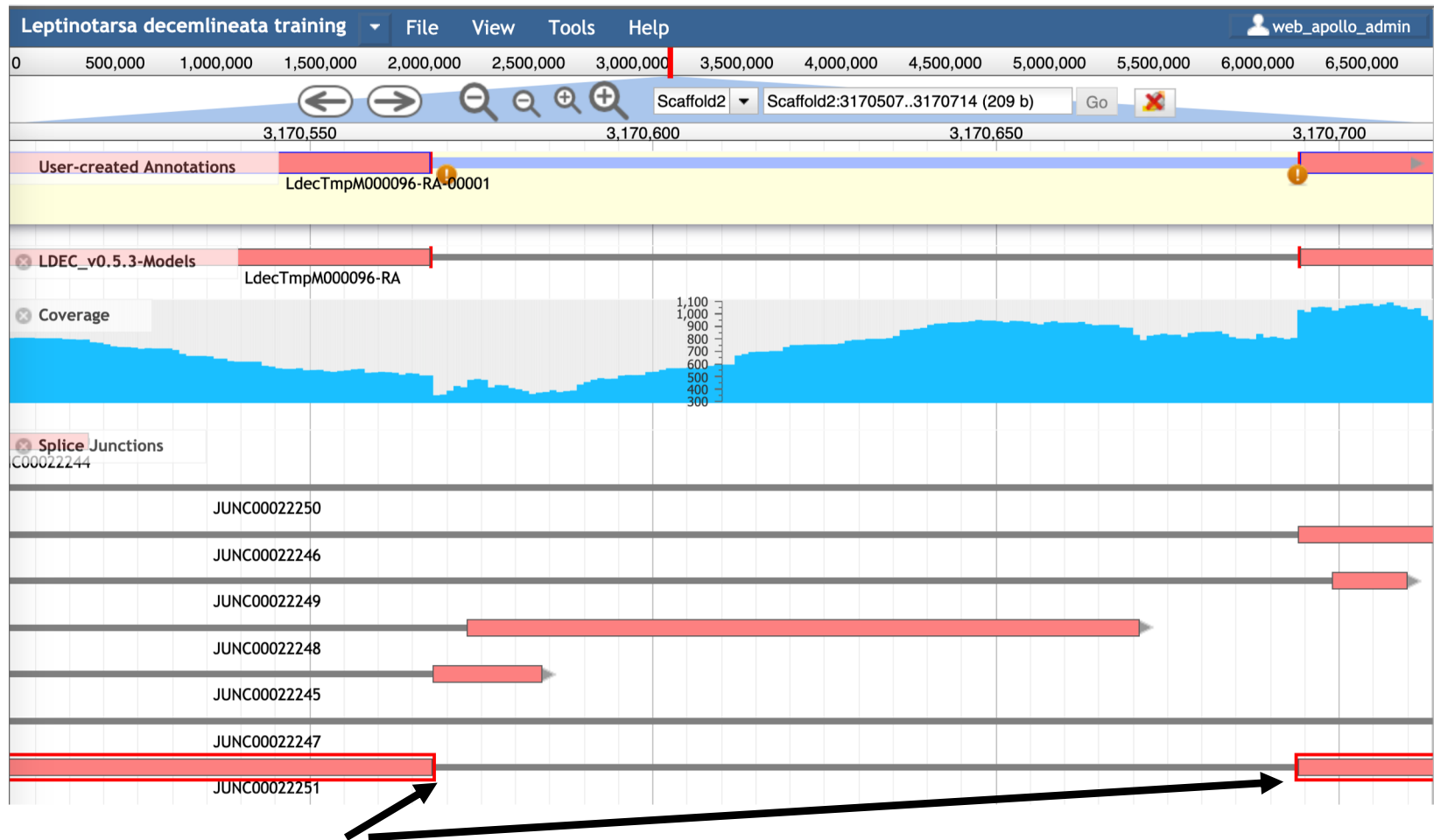


RNA-Seq suggests that this intron
may not exist

Sequence alterations

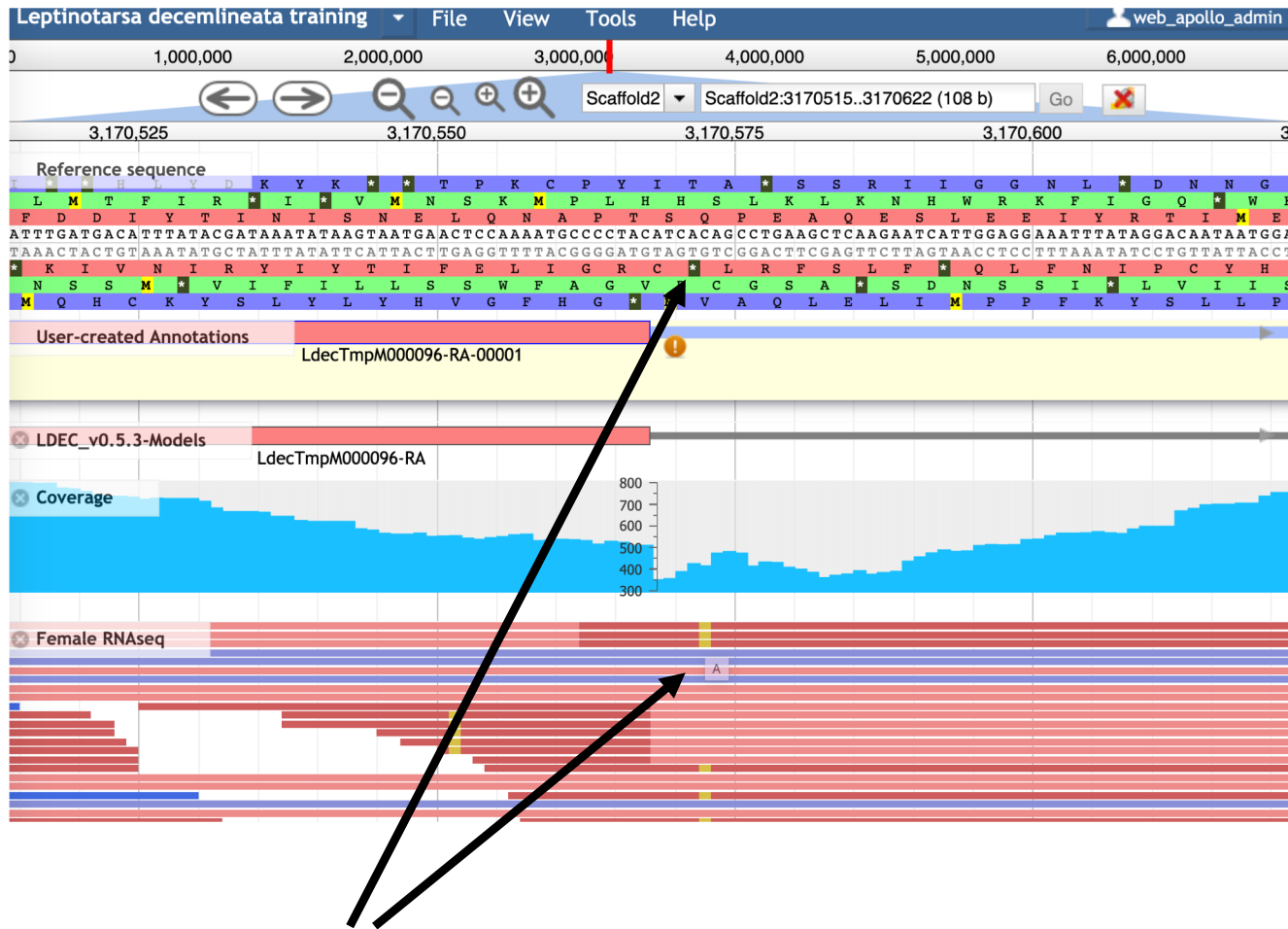


Sequence alterations



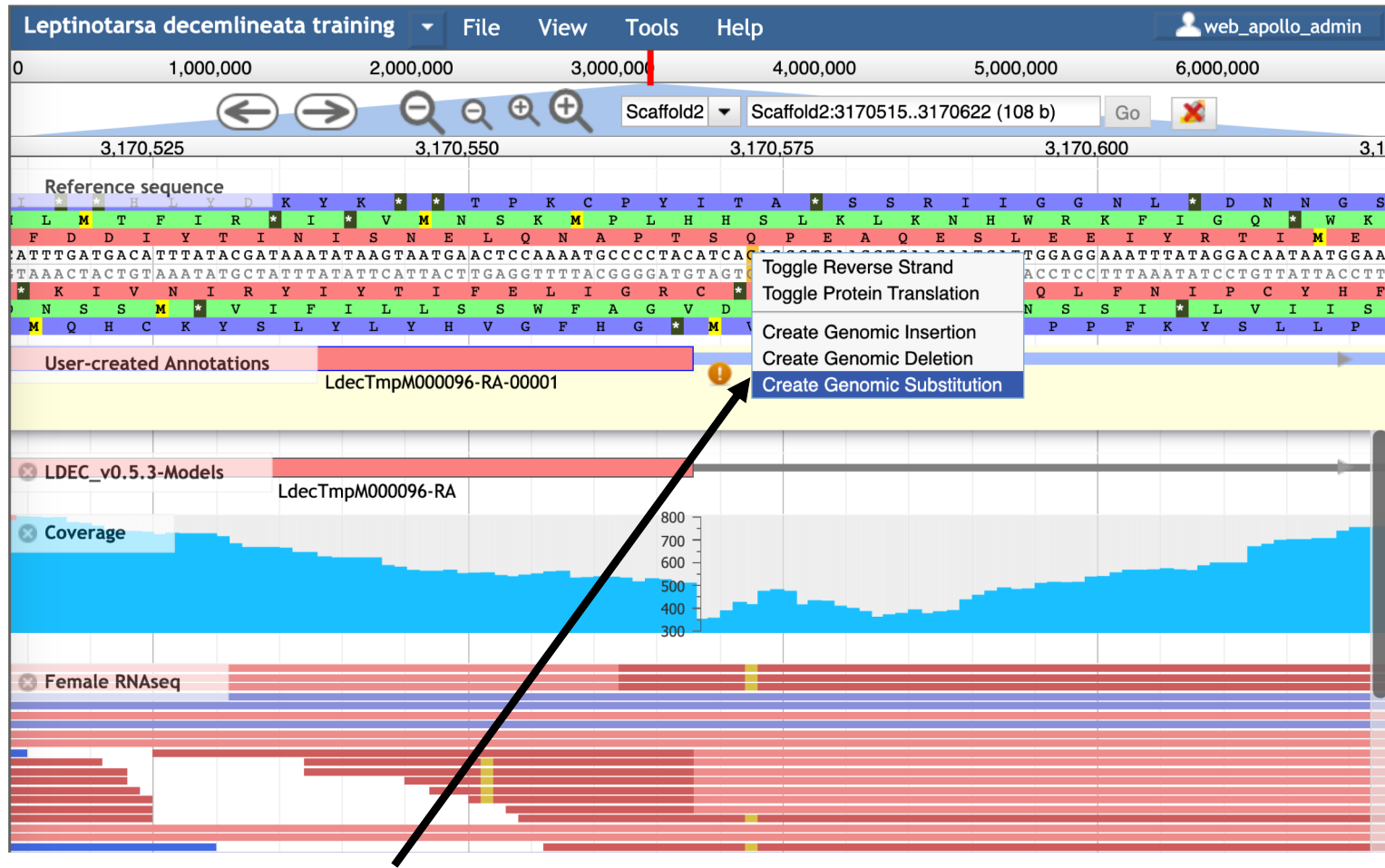
There is some support for a spliced isoform, but the RNA-Seq also suggests contiguous coding sequence

Sequence alterations



Zooming in, we see a stop codon in the 'pink' frame on the reverse strand, but SNPs in all the RNA-Seq reads

Sequence alterations



Right-click on the corresponding nucleotide in the *genome assembly* and select 'Create Genomic Substitution'

Sequence alterations

Leptinotarsa decemlineata training

File View Tools Help

web_apollo_admin

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

Scaffold2 Scaffold2:3170515..3170622 (108 b) Go

3,170,525 3,170,550 3,170,575 3,170,600

Reference sequence

K Y K * * T P K C P Y I T A * S S R I I G G N L * D N N G

L M T F I R * I * V M N S K M P L H H S L K L K N H W R K F I G Q * W K

F D D I Y T I N I S N E L Q N A P T S Q P E A Q E S L E E I Y R T I M E

ATTTGATGACATTTATACGATAAAATATAAGTAATGAACCTCCAAAATGCCCTACATCAAGCCTGAAGCTCAAGAATCATTGGAGGAAATTTATAGGACAATAATGGA

TAAACTACTGTAAATATGCTATTTATATTCTTACTTGAGGTTTTACGGGGATGTAGTGTCTGGACTTCGAGTCTTAGTAACCTCCTTTAAATATCCTGTTATTACCT

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

N S S M * V I F I L L S S W F A G V D C G S A * S D N S S I * L V I I S

M Q H C K Y S L Y L Y H V G F H G * M V A Q L E L I M P P F K Y S L L P

User-created Annotations

LdecTmpM000096-RA-00001

LDEC_v0.5.3-Models

LdecTmpM000096-RA

Coverage

Female RNAseq

Add Substitution

+ strand A

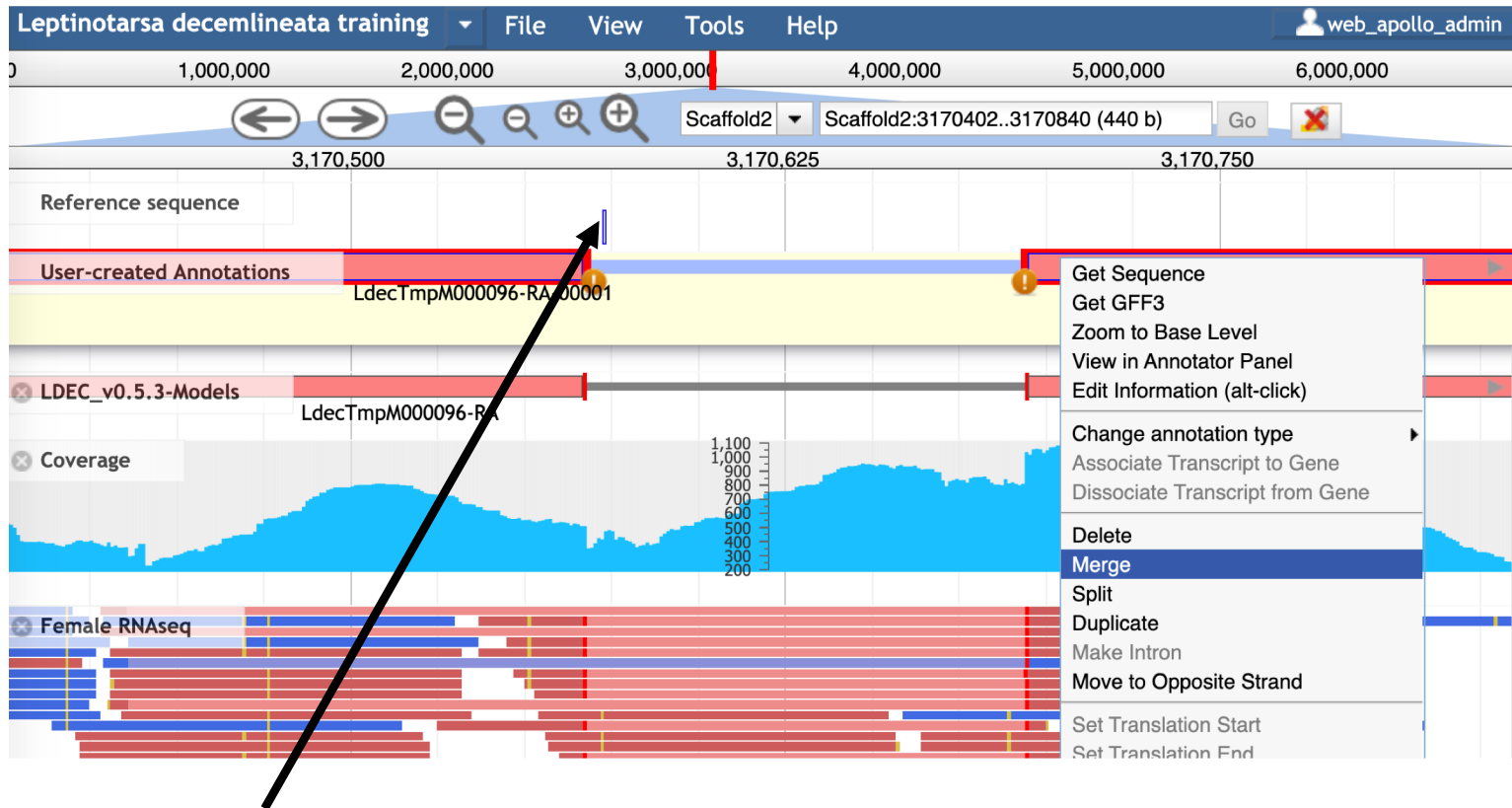
- strand T

Comment All available RN

Add

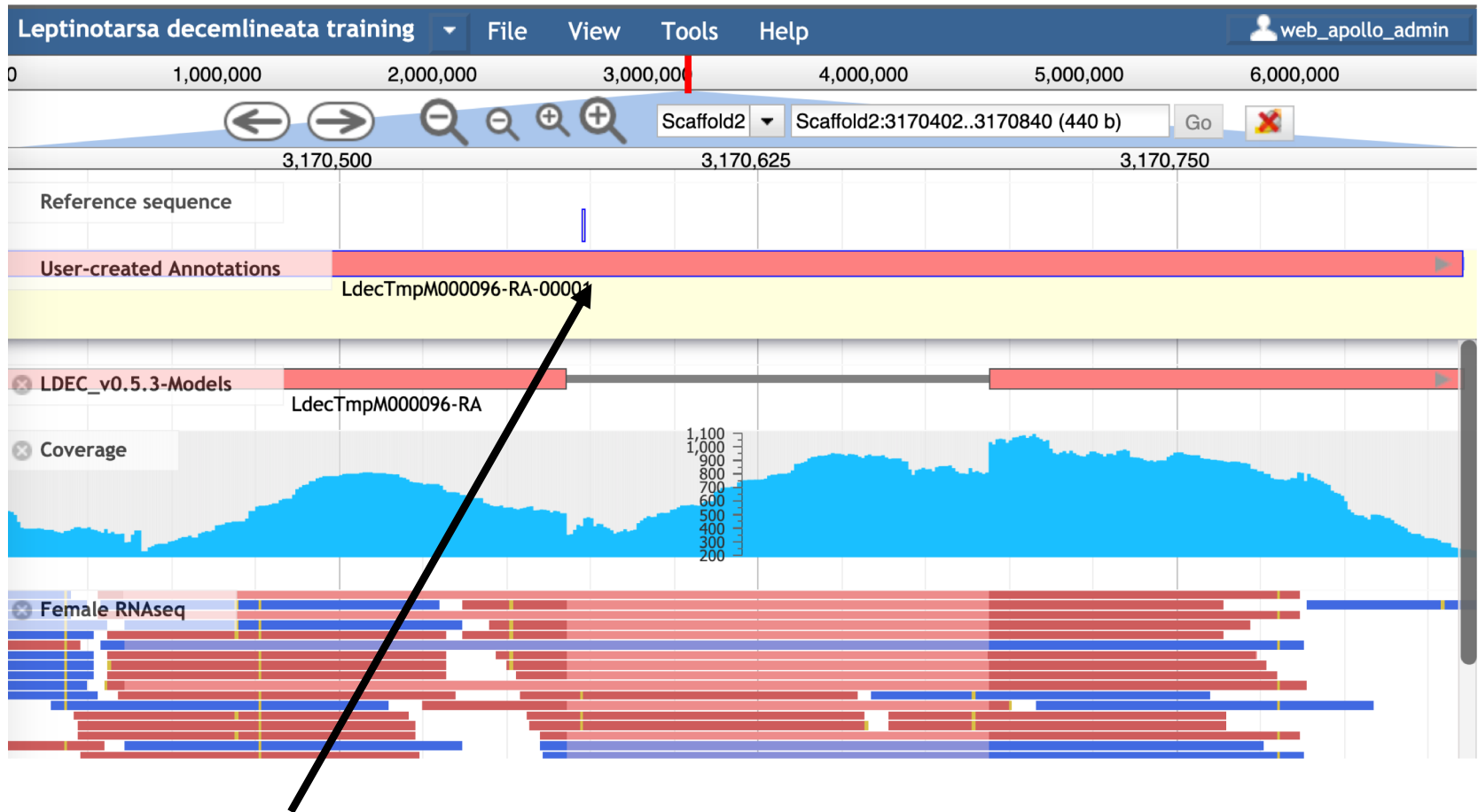
Add the substitution, and a (required) justification

Sequence alterations



Apollo added the substitution! Now, let's merge the CDS regions

Sequence alterations



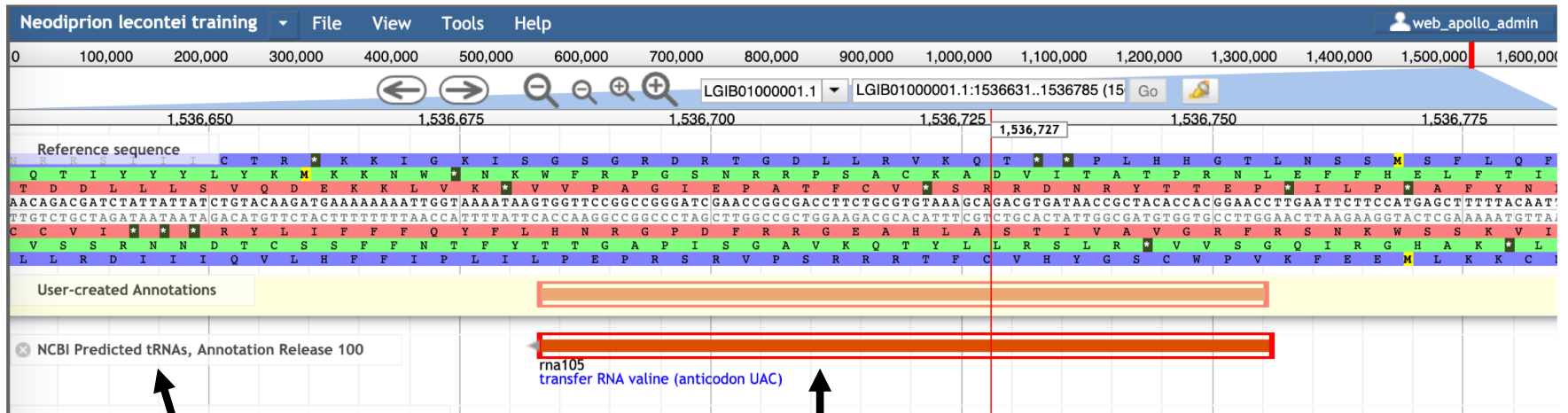
The sequence merged!

Non-coding features

Non-coding features

- Apollo supports protein-coding and non-coding features
- By default, Apollo will create protein-coding features
- For non-coding features, you can set the feature type before or after setting up the model

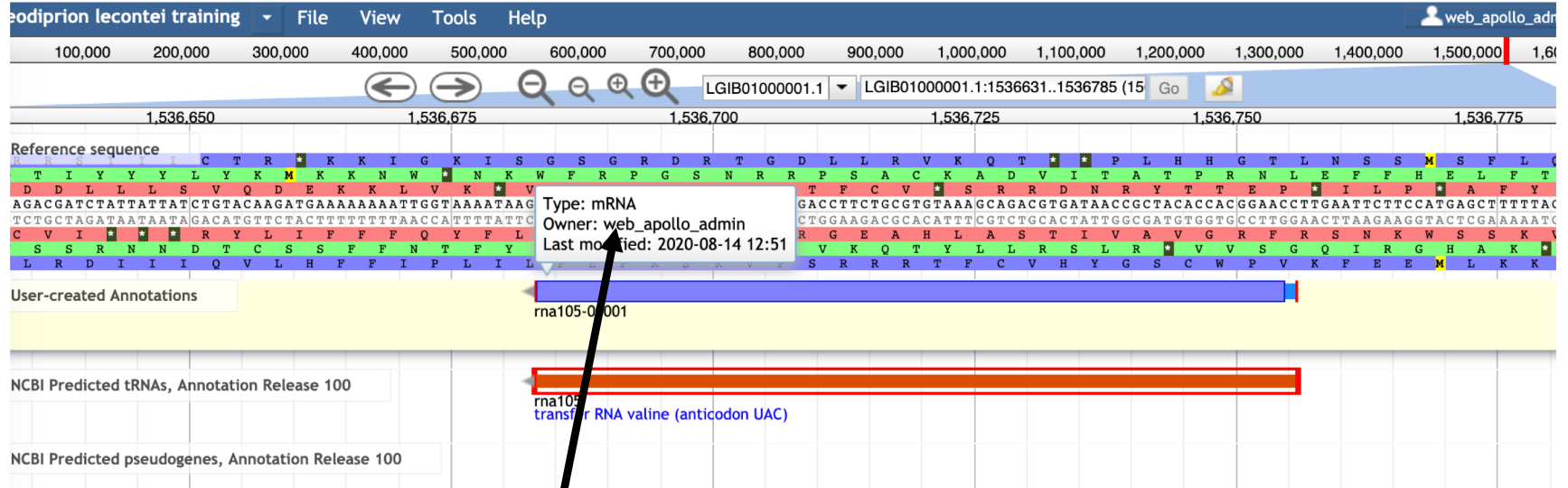
Non-coding features



This feature is a
tRNA

Let's drag it to
the UcaA track to
modify it

Non-coding features



Apollo turned it
into an mRNA –
let's fix that

Non-coding features

The screenshot displays the Neodiprion lecontei genome browser interface. The top navigation bar includes 'Neodiprion lecontei training', 'File', 'View', 'Tools', 'Help', and a user profile 'web_apollo_admin'. A genomic scale from 100,000 to 1,600,000 is shown at the top. Below the scale, navigation controls (back, forward, zoom) and a search bar containing 'LGIB01000001.1' are visible. The main track shows the 'Reference sequence' with nucleotide bases (A, C, G, T) and a 'User-created Annotations' track. A specific annotation, 'rna105-00001', is highlighted in a yellow box. A context menu is open over this annotation, listing various actions such as 'Get Sequence', 'Zoom to Base Level', 'Change annotation type', and 'Set Translation Start'. The 'Change annotation type' option is currently selected, showing a sub-menu with options like 'gene', 'pseudogene', 'rRNA', 'snRNA', 'tRNA', 'ncRNA', 'miRNA', 'repeat_region', and 'transposable_element'. The 'tRNA' option is highlighted in blue. Below the main track, there are tracks for 'NCBI Predicted tRNAs, Annotation Release 100' and 'NCBI Predicted pseudogenes, Annotation Release 100'. The 'rna105' tRNA is also visible in this track, with a note 'transfer RNA valine (anticodon UAC)'.

Right-click on feature, select
'Change annotation type', then
'tRNA'

Non-coding features

The screenshot displays the Neodiprion lecontei training interface. The top menu bar includes 'File', 'View', 'Tools', and 'Help'. The user is logged in as 'web_apollo_admin'. The genomic track shows a reference sequence with coordinates from 1,536,650 to 1,536,775. A tRNA annotation is visible, labeled 'rna105-000001', with a tooltip showing 'Type: tRNA', 'Owner: web_apollo_admin', and 'Last modified: 2020-08-14 12:52'. Below the track, there are checkboxes for 'NCBI Predicted tRNAs, Annotation Release 100' and 'NCBI Predicted pseudogenes, Annotation Release 100'. A black arrow points from the text 'Now we have the correct feature type' to the tRNA annotation.

Neodiprion lecontei training

File View Tools Help

web_apollo_admin

100,000 200,000 300,000 400,000 500,000 600,000 700,000 800,000 900,000 1,000,000 1,100,000 1,200,000 1,300,000 1,400,000 1,500,000 1,600,000

1,536,650 1,536,675 1,536,700 1,536,725 1,536,750 1,536,775

Reference sequence

1,536,689

Type: tRNA
Owner: web_apollo_admin
Last modified: 2020-08-14 12:52

User-created Annotations

rna105-000001

NCBI Predicted tRNAs, Annotation Release 100

NCBI Predicted pseudogenes, Annotation Release 100

rna105
transfer RNA valine (anticodon UAC)

Now we have the correct feature type

Non-coding features

The screenshot shows the Neodiprion lecontei training interface. The top bar includes a menu (File, View, Tools, Help) and a user profile (web_apollo_admin). Below the menu is a scale from 100,000 to 1,600,000. A search bar contains 'LGIB01000001.1' and 'LGIB01000001.1:1536631..1536785 (15)'. The reference sequence is displayed below the scale, with a red bar highlighting a region. A context menu is open over the red bar, showing options: 'View details', 'Highlight this tRNA', and 'Create new annotation'. The 'Create new annotation' menu is expanded, showing a list of feature types: gene, pseudogene, tRNA, snRNA, snoRNA, ncRNA, rRNA, miRNA, repeat_region, and transposable_element. An arrow points from the text below to the 'tRNA' option in the menu.

Reference sequence

User-created Annotations

NCBI Predicted tRNAs, Annotation Release 100

NCBI Predicted pseudogenes, Annotation Release 100

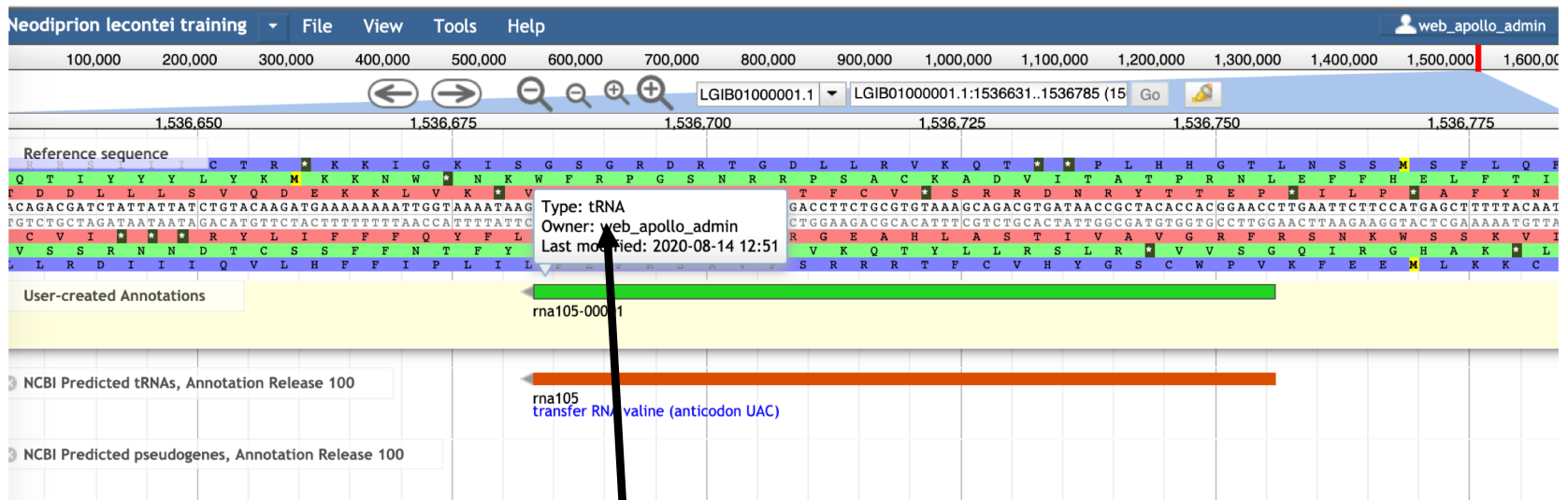
rna105
transfer RNA valine (anticodon UAC)

View details
Highlight this tRNA
Create new annotation

- gene
- pseudogene
- tRNA
- snRNA
- snoRNA
- ncRNA
- rRNA
- miRNA
- repeat_region
- transposable_element

Another way – right-click on model *before* adding it to the Uca track, select 'Create new annotation', then select 'tRNA'

Non-coding features



Now it's a tRNA!

Thank you!

The NAL Team

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

- 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.Childers@usda.gov

