A. *thaliana*, Apollo and You:

Collaborative Genome Annotation Editing

Today:

- Review of basics
- Hands on exercises in editing
- What’s next

After this session, you will be able to:

- Perform gene model edits of various types
- Save comments and status

Known issues: Patience appreciated

- Mt and Cp annotations not yet visible
- Some evidence tracks still missing

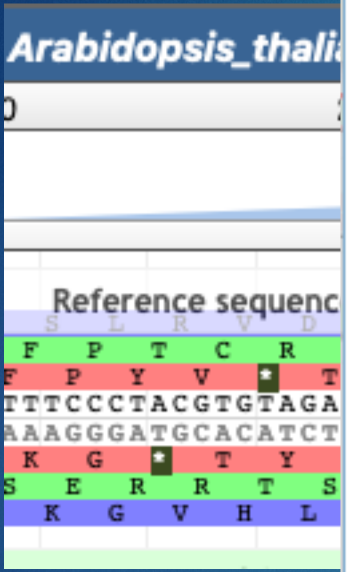
General process of curation

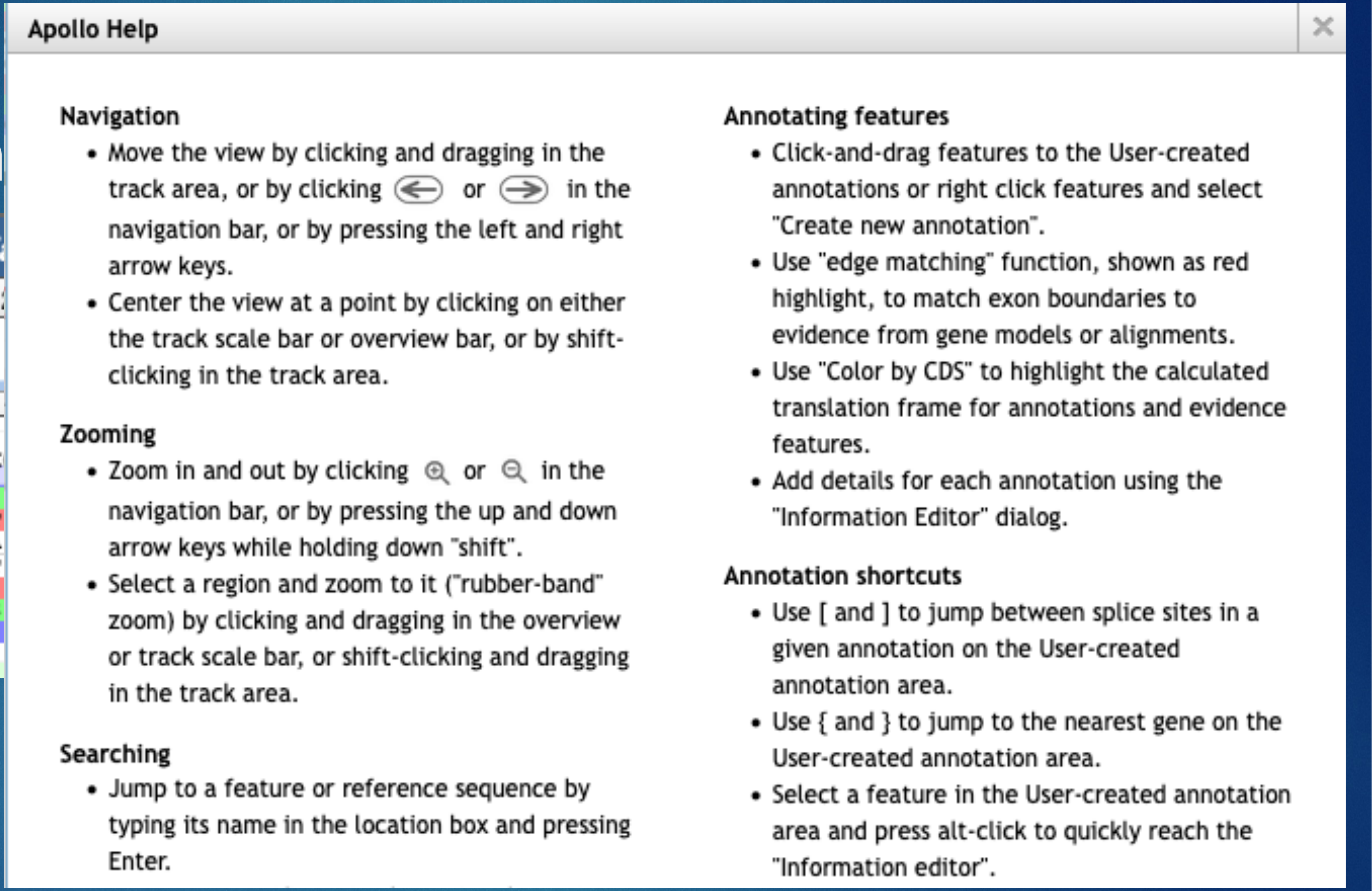
1. Select or find a **region of interest** (e.g., gene or coordinate range).
2. Select appropriate **evidence** tracks to review the genome element to annotate (e.g., gene model).
3. If necessary, **adjust** the gene model.
4. Check your edited gene model for **integrity and accuracy** by comparing it with available homologs.
5. **Comment, change status,** and finish.

Set up for success: Have these tabs ready to go in your browser window

- Apollo
- Apollo user guide
- JBrowse (Araport 11)
- NCBI BLAST

Tips an





Apollo Help

Navigation

- Move the view by clicking and dragging in the track area, or by clicking ⇦ or ⇨ in the navigation bar, or by pressing the left and right arrow keys.
- Center the view at a point by clicking on either the track scale bar or overview bar, or by shift-clicking in the track area.

Zooming

- Zoom in and out by clicking ⊕ or ⊖ in the navigation bar, or by pressing the up and down arrow keys while holding down "shift".
- Select a region and zoom to it ("rubber-band" zoom) by clicking and dragging in the overview or track scale bar, or shift-clicking and dragging in the track area.

Searching

- Jump to a feature or reference sequence by typing its name in the location box and pressing Enter.

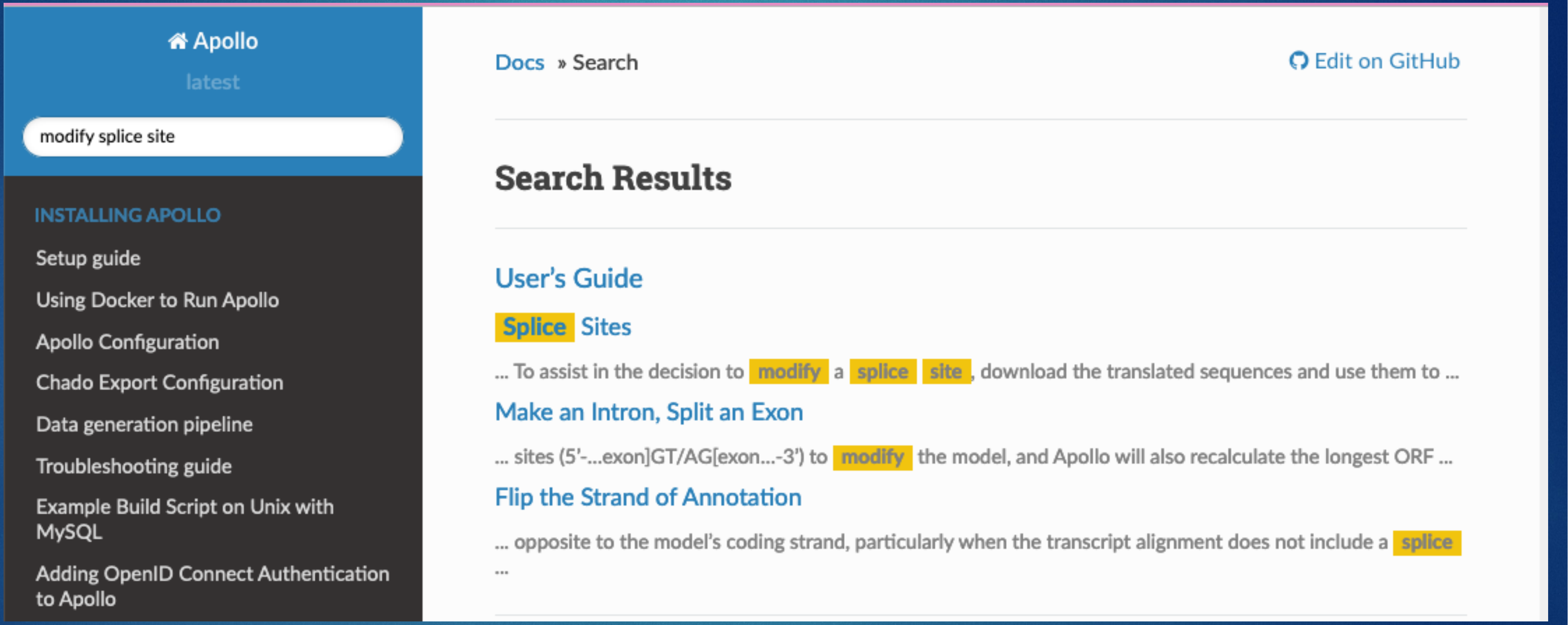
Annotating features

- Click-and-drag features to the User-created annotations or right click features and select "Create new annotation".
- Use "edge matching" function, shown as red highlight, to match exon boundaries to evidence from gene models or alignments.
- Use "Color by CDS" to highlight the calculated translation frame for annotations and evidence features.
- Add details for each annotation using the "Information Editor" dialog.

Annotation shortcuts

- Use [and] to jump between splice sites in a given annotation on the User-created annotation area.
- Use { and } to jump to the nearest gene on the User-created annotation area.
- Select a feature in the User-created annotation area and press alt-click to quickly reach the "Information editor".

Tips and tricks: Apollo Help Docs



INSTALLING APOLLO

- Setup guide
- Using Docker to Run Apollo
- Apollo Configuration
- Chado Export Configuration
- Data generation pipeline
- Troubleshooting guide
- Example Build Script on Unix with MySQL
- Adding OpenID Connect Authentication to Apollo

Search Results

User's Guide

Splice Sites

... To assist in the decision to modify a splice site, download the translated sequences and use them to ...

Make an Intron, Split an Exon

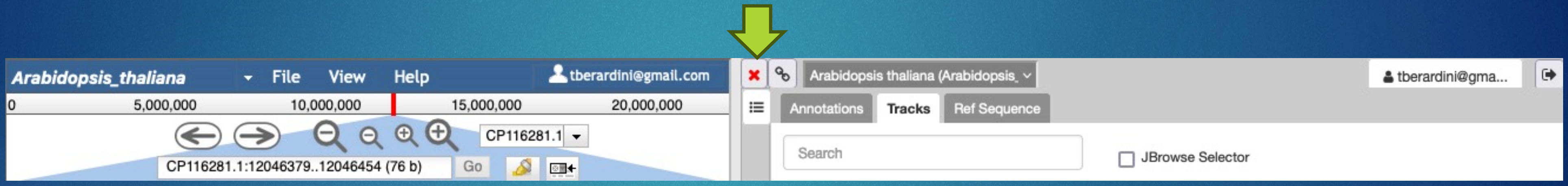
... sites (5'-...exon]GT/AG[exon...-3') to modify the model, and Apollo will also recalculate the longest ORF ...

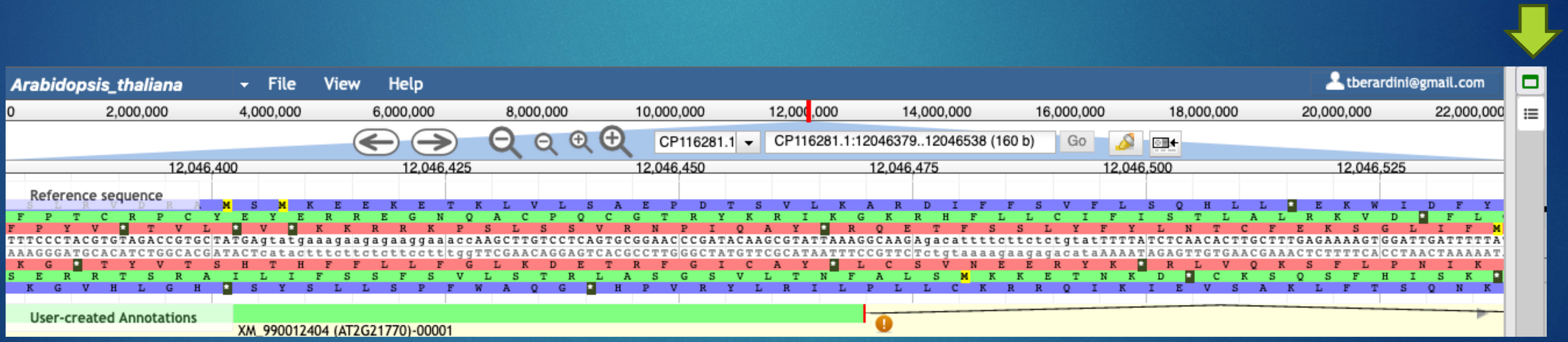
Flip the Strand of Annotation

... opposite to the model's coding strand, particularly when the transcript alignment does not include a splice ...

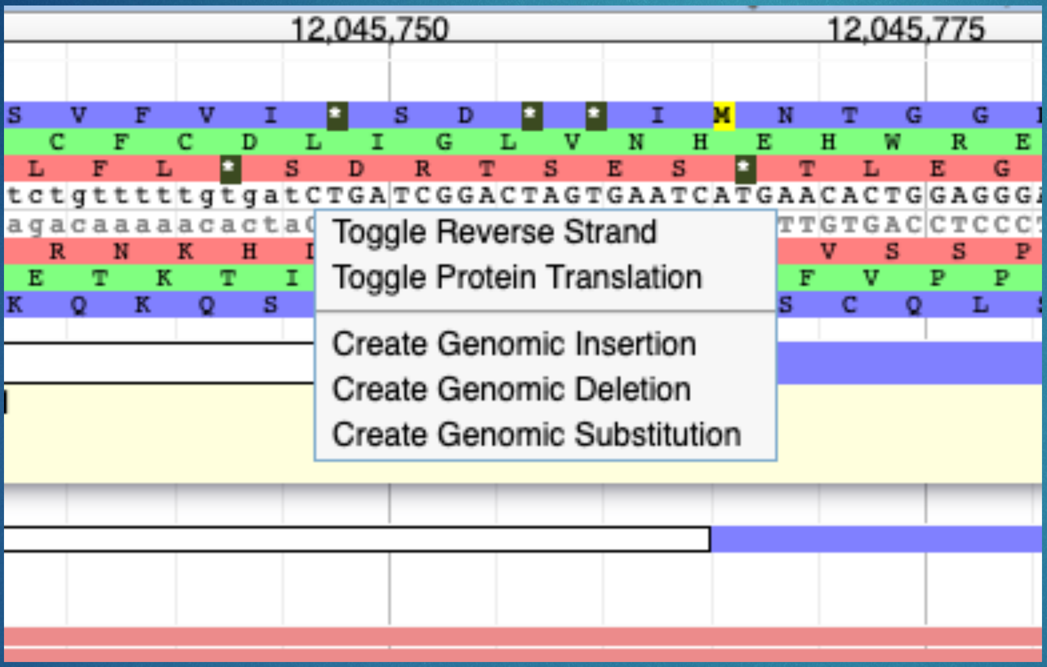
https://genomearchitect.readthedocs.io/en/latest/search.html

Tips and tricks: Show/hide sidebar





Tips and tricks: Toggle sequences



Click in DNA region

Evidence tracks

- Col-CC annotation: end result of pipeline
- Gnomon models: *one* of the inputs into the pipeline
- TSA (transcript shotgun assembly): isoseq contigs + extra isoseq
- Protein alignments: alignments of protein sequences from Genbank records (multi-species) with Col-CC models
- (RNA seq) - A. thaliana
- (Long read RNA) – A. thaliana

Today:

- Review of basics
- Hands on exercises in editing
- What’s next

New server: now with more power

- https://apollo-sandbox.arabidopsis.org/apollo/annotator/index

Types of gene updates

1. deleted
2. split
3. merged
4. novel
5. locus type changed
6. cds changed
7. BUSCO gene disappeared

Exercise: Check a deleted gene

- AT5G07545 in Araport11 is absent in Col-CC prediction
- JBrowse: Chr 1: 16608602..16608817 (AT1G43825)

- Look for neighboring gene
- Check supporting evidence for a gene in that space
  - Absent: Check off the list as verified deleted
  - Present: Reinstate, drag and drop into user-created annotations bar

Exercise: Check a split gene

- Pipeline may inadvertently split two genes, need to be rejoined
- CP116284.1:23073165..23075512
- Check JBrowse
- Check evidence
- Drag both gene models to yellow band
- Click on both (shift click)
- Right click menu: Merge
- Adjust splice sites
- Check protein sequence, NCBI BLAST RefSeq proteins

Exercise: Check a merged gene

- CP116280.1:9679581..9689660
- Look a transcript support
- Check JBrowse
- Check Protein Sequence (UniProt works)
- Highlight flanking exons (shift click), right click menu, split

Exercise: Check a substantial change

- CP116280.1:10025296..10027939
- AT1G28450
- Drag up: observe that CDS got shorter in Col-CC
- Compare to original gene model, fix the start of translation
- Look a transcript support, compare UTRs now
- Look at ‘View details’: Model_evidence: Supporting evidence includes similarity to: 3 Proteins, 6 long SRA reads, and 100% coverage of the annotated genomic feature by RNAseq alignments

Exercise: Check a new gene

- CP116284.1:17415221..17415818
- Look at region around it in JBrowse
- Look at evidence for expression
- WHY???
- Right click Open Annotation
- Rename gene model as DELETE THIS

Exercise: Check a new gene

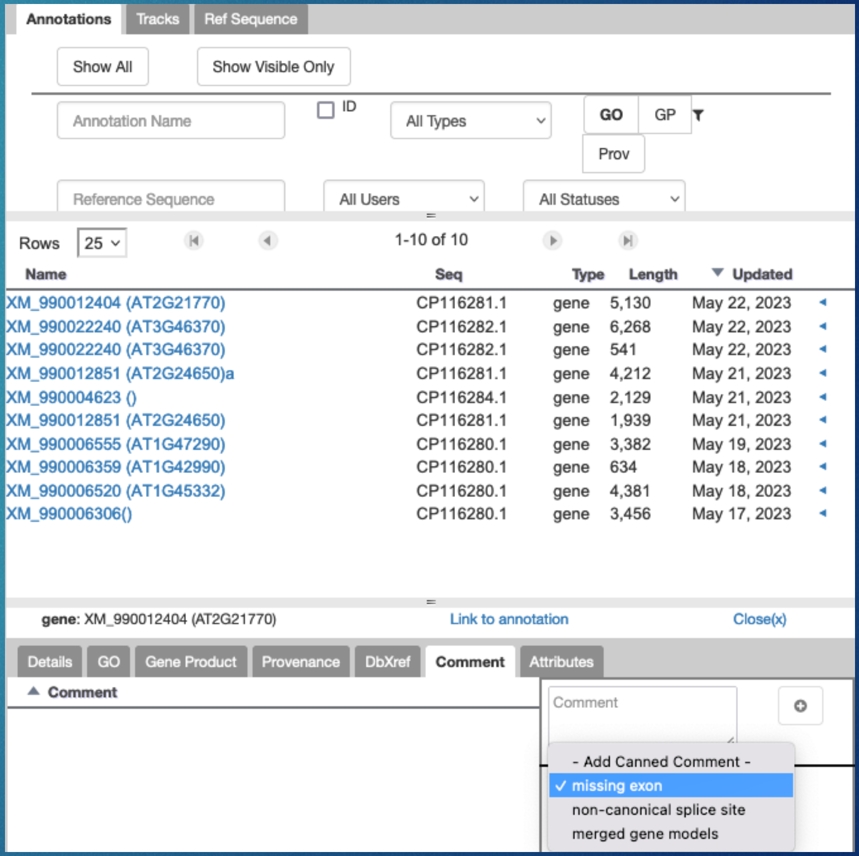
- CP116283.1:12395699..12397962
- Look at region around it in Jbrowse (AT4G15980)
- Look at evidence for expression
- Check sequence, blast
- Keep, model checked

Tips and tricks: Saving comments/status

- Click out of the panel you've changed into another one

2. Click here to save

1. Enter comment



Rows 25 1-10 of 10

Name	Seq	Type	Length	Updated
XM_990012404 (AT2G21770)	CP116281.1	gene	5,130	May 22, 2023
XM_990022240 (AT3G46370)	CP116282.1	gene	6,268	May 22, 2023
XM_990022240 (AT3G46370)	CP116282.1	gene	541	May 22, 2023
XM_990012851 (AT2G24650)a	CP116281.1	gene	4,212	May 21, 2023
XM_990004623 ()	CP116284.1	gene	2,129	May 21, 2023
XM_990012851 (AT2G24650)	CP116281.1	gene	1,939	May 21, 2023
XM_990006555 (AT1G47290)	CP116280.1	gene	3,382	May 19, 2023
XM_990006359 (AT1G42990)	CP116280.1	gene	634	May 18, 2023
XM_990006520 (AT1G45332)	CP116280.1	gene	4,381	May 18, 2023
XM_990006306()	CP116280.1	gene	3,456	May 17, 2023

gene: XM_990012404 (AT2G21770) Link to annotation Close(x)

Today:

- Review of basics
- Hands on exercises in editing
- What's next

What's next?

- ICAR 2023 – June 5 – 9

- Gene set assignment
  1. split
  2. merged
  3. deleted
  4. novel
  5. locus type changed
  6. cds changed
  7. BUSCO gene disappeared
  8. desired gene family (may overlap with 1-7)

What's next? Further out

- Website: tinyurl.com/AthalianaV12
  - Updates, training material, video will be accessible from here
  - Tracking work and review
    - Google Sheet
    - Excel spreadsheet (no Google Drive access)
- Slack channel (#athalianav12-manual-review)
  - Bug reports, asynchronous feedback/questions, paste the link to the region and the issue
- When review starts in earnest: Regular call time: Zoom, Wed 7 – 7:30 am Pacific (proposed)
- Group work? – there are many institutions with >1 person

Thank you!

- **Col-CC Assembly:** Korbinian Schneeberger and lab team
- **NCBI Eukaryotic Genome Pipeline:** Françoise Thibaud-Nissen, Terence Murphy
- **Apollo setup @TAIR:** Shabari Subramaniam, Xingguo Chen, Trilok Prithvi, Chris Childers
- **Training materials:** Moni Munoz Torres, Marcela Tello Ruiz, Monica Poelchau, Jason Williams
- **The wider Arabidopsis community**
- **YOU**