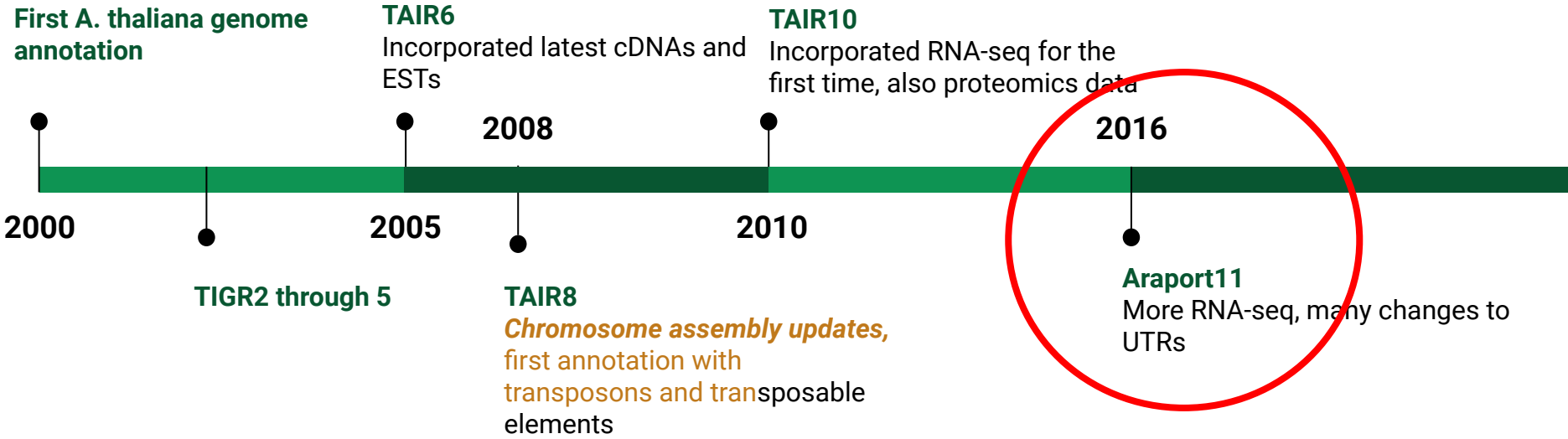
The Arabidopsis Community Effort to Reannotate the Arabidopsis thaliana Genome

Tanya Berardini, TAIR and Phoenix Bioinformatics

tair

Phoenix BIOINFORMATICS

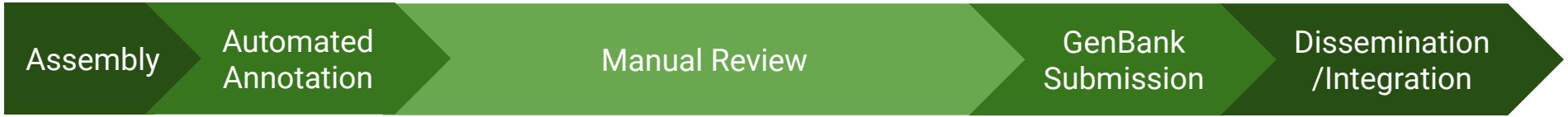
Timeline



tair

Phoenix BIOINFORMATICS

Creating a new genome annotation

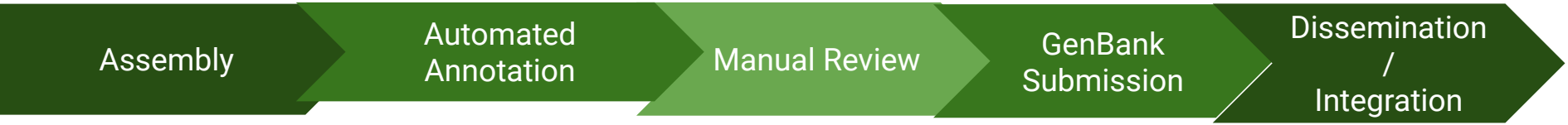


V12: Community effort, contributions of expertise and computing resources

tair

Phoenix BIOINFORMATICS

Aspirational timeline: V12 released by Jan. 2024



Dec 2022 (DONE, 12/23/22)

Currently in queue for assembly review

Jan - March 2023?

April - Sept 2023?

Oct - Nov 2023?

Start: Jan 2024?

tair

Phoenix BIOINFORMATICS

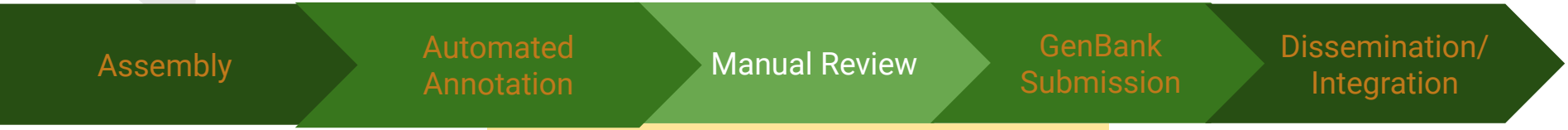
Where we are

Assembly	Automated Annotation	Manual Review	GenBank Submission	Dissemination/ Integration
Who: Schneeberger lab	**Who:** NCBI	**Who:** Community experts for review, TAIR for coordination	**Who:** TAIR + NCBI	**Who:** BAR, TAIR, EnsemblPlants, NCGR GCV, AtPeptide Atlas, many more
Input: 13 Col-CC sequences	**Input:** Col-CC Assembly	**Input:** Automated genome annotation	**Input:** Reviewed genome annotation	**Input:** Updated RefSeq record for A. thaliana
Output: Col-CC Assembly v.1	**Output:** Genome annotation with groups of genes for manual review	**Output:** Reviewed genome annotation	**Output:** Updated RefSeq record for A. thaliana	**Output:** updated tools and viewers for A. thaliana genome data
+NOR2/4 v.2		**Tool:** Apollo		

tair

Phoenix BIOINFORMATICS

Manual Review



- Review of the automated annotation pipeline results
- Annotation of specific types of genome elements
  - Transposable elements
  - Tandem repeat elements
  - 5SrRNAs
  - Long non-coding RNAs

tair

PhoenixBIOINFORMATICS

Organizing the community

- Manual review and team review
  - Slack
  - Email
  - X
  - Website: tinyurl.com/Athalianav12
  - Zoom
    - Office hours
    - Team meetings

tair

Phoenix BIOINFORMATICS

Resources

- TAIR/Phoenix
  - Project manager
  - Bioinformatician
  - Systems admin
    - Apollo server: 1 TB

tair

Phoenix BIOINFORMATICS

Automated Annotation Results: overview

- Fewer (!) total protein coding genes
- Fewer splice variants
- More rRNA genes

tair

Phoenix BIOINFORMATICS

Distributing work

- By class of change to review (merge, split, removed, significantly changed)
- By type of gene (protein coding, rRNA, transposable element, type of TE)
- By gene family

tair

Phoenix BIOINFORMATICS

Table 2: Current Counts of New, Deleted, Split, Merged Protein-Coding Genes in V12

	CP116280.1	CP116281.1	CP116282.1	CP116283.1	CP116284.1	Totals
New or Reinstated Gene Models	857	549	593	465	684	3148
Deleted Genes	130	98	92	79	66	465
Deleted Overlapping Genes	275	125	126	128	163	817
Split/Split Recommendations	25	11	8	14	16	74
Merge/Merge Recommendations	29	6	23	8	22	88

tair

Phoenix BIOINFORMATICS

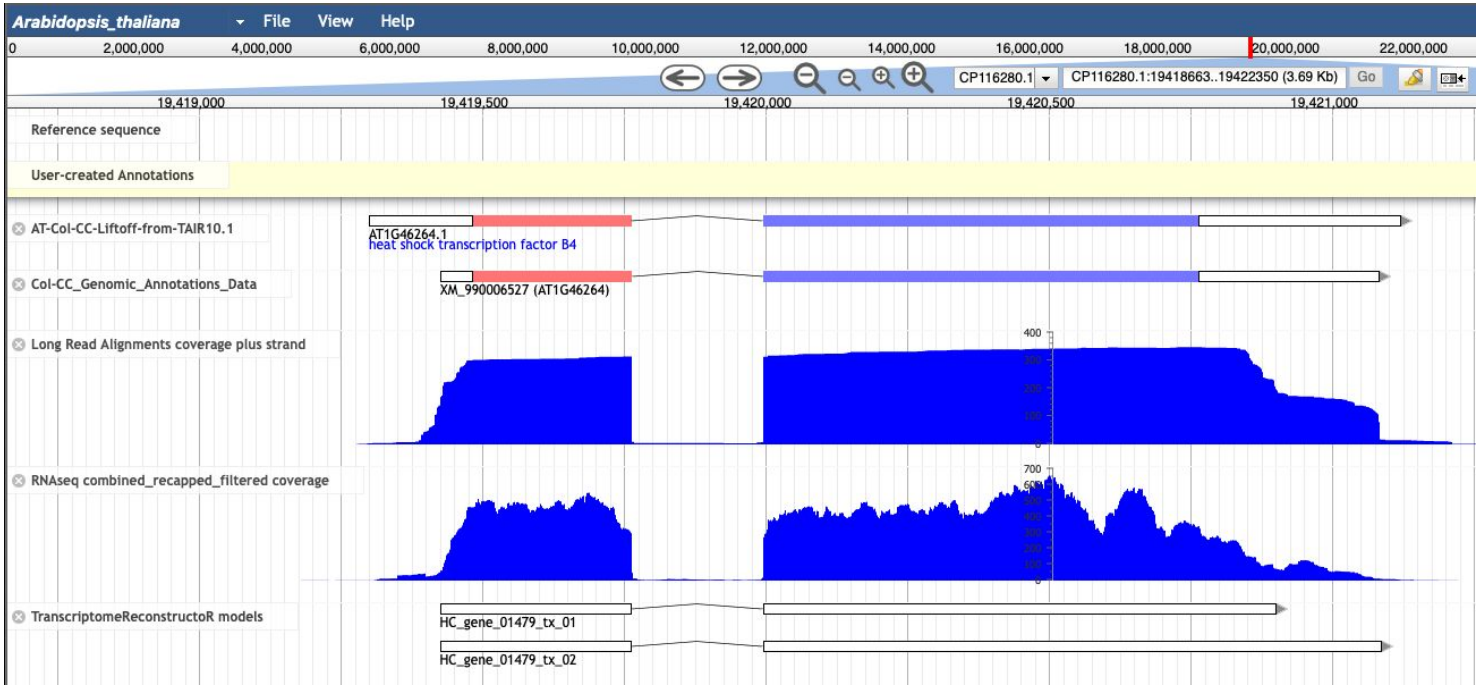
Manual review: Apollo status, as of 1/14/24

Type	Number reviewed
Updated, no secondary review needed	2180
Secondary review complete, accepted	709
Under secondary review	4
Updated, secondary review requested	7
Unable to update	16
For discussion	64
No update needed	72
Under primary review	1
Total	3053

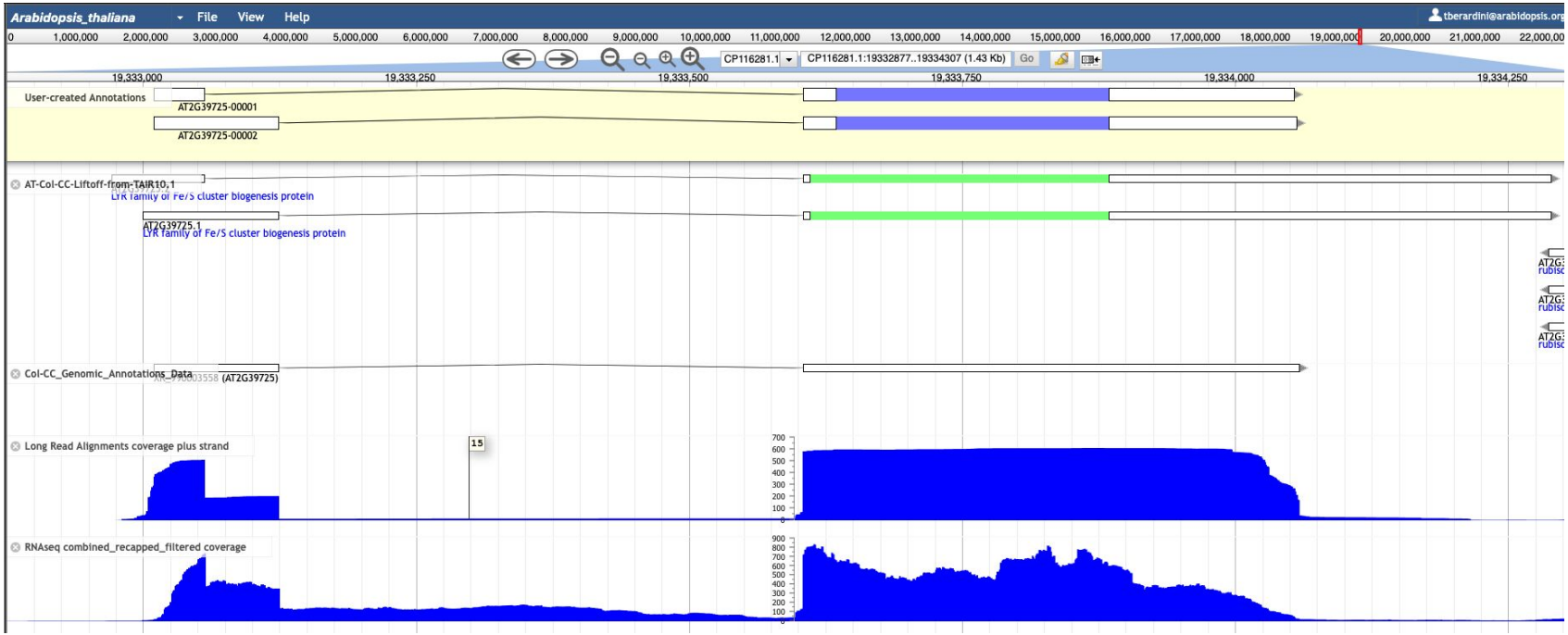
tair

Phoenix BIOINFORMATICS

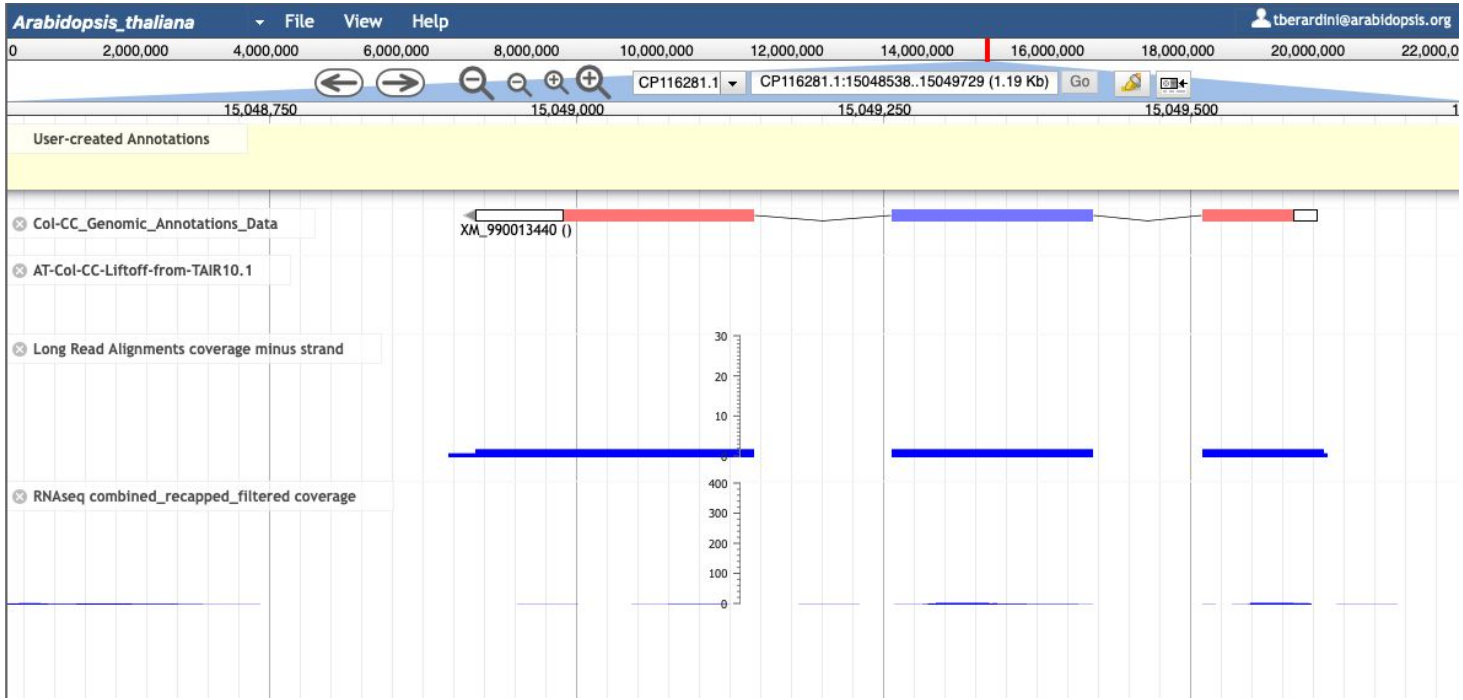
Col-CC prediction cleans up V11 model



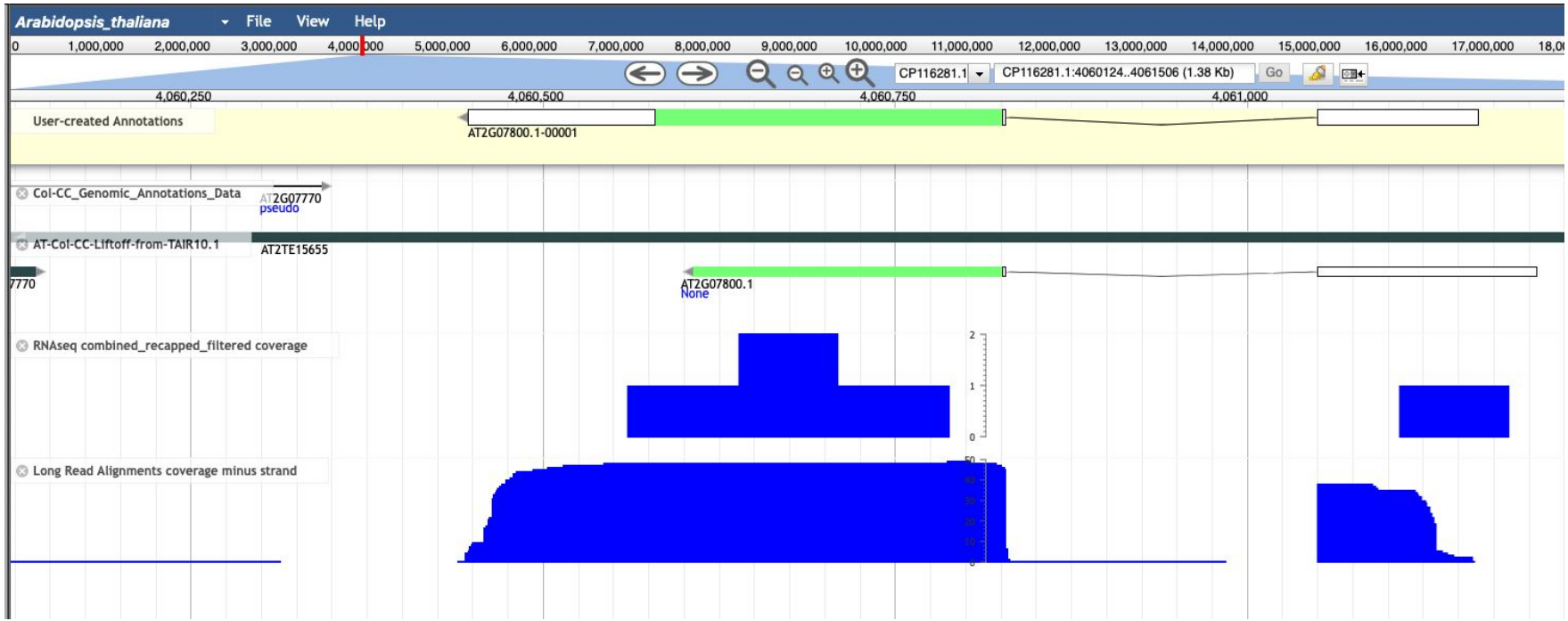
V11 plus transcript evidence improves Col-CC prediction



A real New gene



Reinstated falsely deleted gene



To do

- Secondary review on all new/split/merged
- Complete various group-based (lncRNA, TE, etc) review and QC
- Groups to deliver GFF file to TAIR team
- TAIR team to integrate all data into single file to submit to Genbank

tair

Phoenix BIOINFORMATICS

Deadlines

GenBank submission

- January 15, 2024
  - all external groups to provide sample GFF3 file to TAIR
- Feb. 29, 2024
  - all external groups to provide final GFF3 file to TAIR
- April 15, 2024
  - For TAIR to submit to Genbank

tair

Phoenix BIOINFORMATICS

The TAIR Team

Tanya Berardini
TAIR Director

Leonore Reiser

Erica Bakker

Shabari
Subramaniam

Alyssa Proia

Trilok Prithvi

Swapnil Sawant

Xingguo Chen

tair

Since 2013, supported by the community through subscriptions

Phoenix BIOINFORMATICS

Contact: curator@arabidopsis.org

tair

Phoenix BIOINFORMATICS