

I5k Workspace updates: What's new with Apollo2

March 29, 2022

Chris Childers

Monica Poelchau

Vern Chapman

i5k@usda.gov

Updates at the i5k Workspace

- Introduction
- Apollo2.1.1 vs Apollo2.6.5
 - New look
 - New interfaces
- Apollo1.0.4 Vs Apollo2.6.5
 - LOTS of changes
- An example annotation workflow

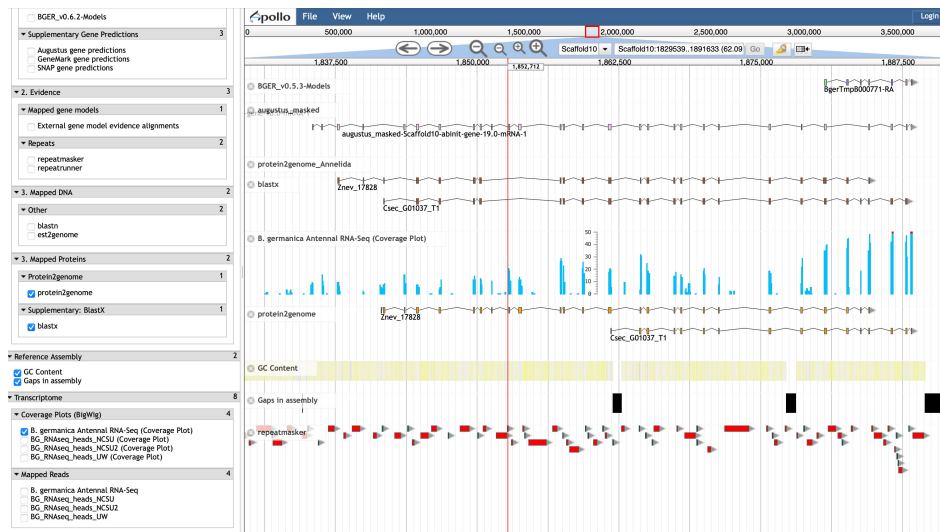
This has been a long time coming...

- Updating Apollo to this version has taken a lot longer than anticipated
 - A large number of updates on the back-end
 - Reworking the track styling system
 - Manually checking each evidence track for issues
 - Creating a new feature type for showing BLAST hits
-
- If you see any issues, please let us know
 - i5k@usda.gov

What is Apollo?

- Apollo is a web-based visual gene curation tool
- Apollo is an extension for the JBrowse genome browser

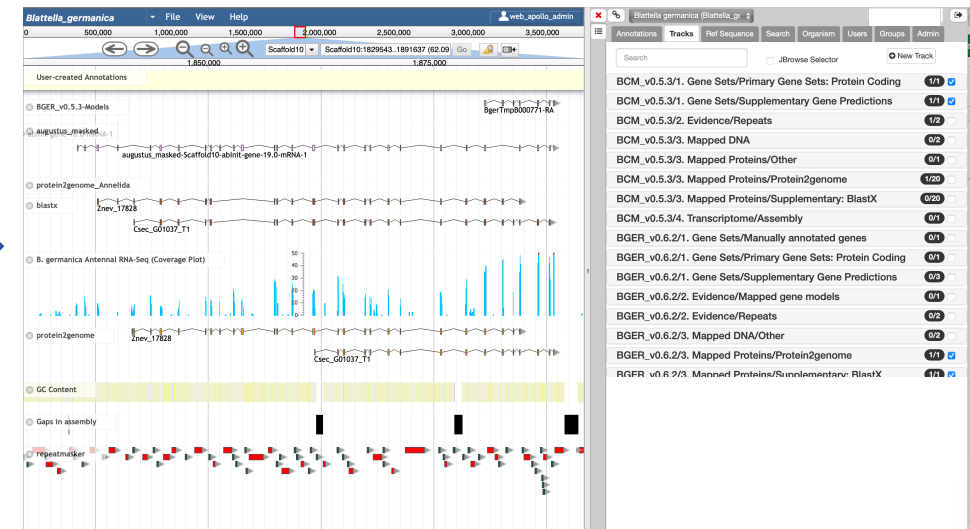
JBrowse



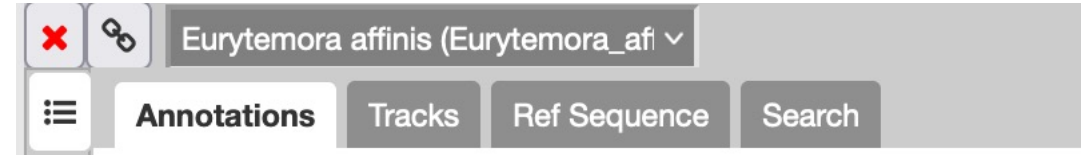
Login



Apollo



Apollo Basics



- Many of the functions are located in the panel on the right
- Your account can have access to multiple organisms
- Tabs allow quick access to the tools
- Annotations – show annotations created by users
- Tracks – shows the groups of tracks, can be turned on/off by group
- Ref Sequences – view and navigate to any scaffold
- Search – BLAT sequence search

Example Gene annotation workflow

- Obtain sequence of interest / map to assembly using BLAST
- Use link in BLAST results to see the matches in JBrowse
- Turn on other evidence tracks (gene predictions, Transcriptome, etc)
- Login to Apollo -> create and refine the gene model

1. Find sequence/search the assembly

BLAST Databases

Organisms

- ☐ *Drosophila rhopaloe*
- ☐ *Drosophila takahashii*
- ☐ *Dufourea novaeangliae*
- ☐ *Ephemera danica*
- ☐ *Eufriesea mexicana*
- ☐ *Euglossa dilemma*
- ☒ *Eurytemora affinis*
- ☐ *Fopius arisanus*
- ☐ *Frankliniella occidentalis*
- ☐ *Galleria mellonella*
- ☐ *Gerris buenoi*
- ☐ *Habropoda laboriosa*
- ☐ *Halyomorpha halys*

Eurytemora affinis

Nucleotide

- ☒ Genome Assembly - Eaff_11172013.genome_new_ids.fa
- ☒ Transcript - EAFF_new_ids.fna

Peptide

- ☐ Protein - EAFF_new_ids.faa

Query Sequence

Your sequence is detected as nucleotide:

```
>CLEC010822-RA:cDNA, Heat shock  
protein 70-2  
TGGAATTTAAATTTTCGATTGGCGCGC  
CTTTAAGCCGGCGCCAATCGGTTTCGGA  
ACGTATTGTCAGTCAGCCGACCAATCAAC  
GCCGTCACAGATTCCCGACTCTCCCGTC
```

(load from example: nucleotide or peptide)

Or load it from disk:

Browse... No file selected.

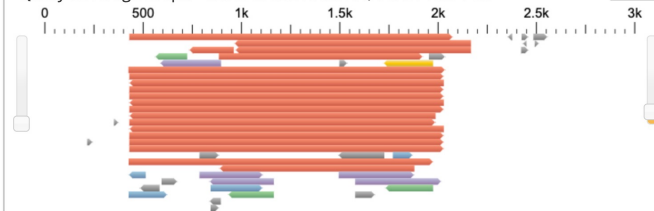
Program

☒ blastn ☐ tblastn ☐ tblastx ☐ blastp ☐ blastx

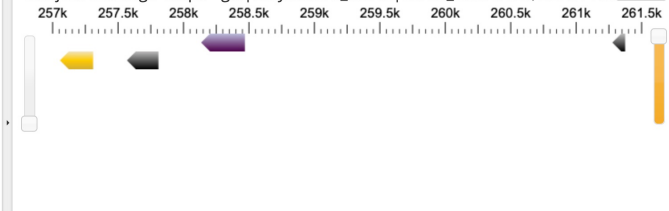
blastn - Nucleotide vs. Nucleotide

[Tutorial](#)

Query Coverage Graph - CLEC010822-RA:cDNA, BLAST Hits 1-50



Subject Coverage Graph - gnl|Eurytemora affinis|euraff_Scaffold26, BLAST Hits 1-50



Showing 35 to 46 of 148 entries

Find in table

blastdb	qseqid	sseqid	pident	length	mismatch	gapopen
euraff	CLEC010822-RA:cDNA	Scaffold26	72.941	255	67	2
euraff	CLEC010822-RA:cDNA	Scaffold26	66.116	242	76	3
euraff	CLEC010822-RA:cDNA	Scaffold26	70.588	102	27	1
euraff	CLEC010822-RA:cDNA	Scaffold26	74.667	75	15	2
euraff	CLEC010822-RA:cDNA	Scaffold26	84.615	39	2	2
euraff	CLEC010822-RA:cDNA	Scaffold26	95.652	23	1	0
euraff	CLEC010822-RA:cDNA	EAFF012888-RA	68.861	395	111	5
euraff	CLEC010822-RA:cDNA	EAFF012888-RA	70.854	199	53	3
euraff	CLEC010822-RA:cDNA	EAFF027773-RA	68.243	444	128	7
euraff	CLEC010822-RA:cDNA	EAFF027773-RA	68.773	269	75	3
euraff	CLEC010822-RA:cDNA	EAFF003903-RA	71.138	246	63	3

BLAST Report FASTA

2475	Sbjct	258169	GTGAACGAGTTCAAGCGCAAGCAAG---AAGGATAT	258135
2476				
2477				
2478				
2479				
2480				
2481				
2482				
2483				
2484				
2485				
2486				
2487				
2488				
2489				
2490				
2491				
2492				
2493				
2494				
2495				
2496				
2497				
2498				
2499				
2500				
2501				
2502				
2503				
2504				
2505				

Download

2022 - National Agricultural Library

3. Use BLAST linkouts to see hits in the assembly

15k@NAL Tools About Us Contact

Query Coverage Graph - CLEC010822-RA:cDNA, BLAST Hits 1-50

Subject Coverage Graph - gnl|Eurytemora_affinis|euraff_Scaffold26, BLAST Hits

Showing 35 to 46 of 148 entries

Find in table

blastdb	qseqid	sseqid	pid	length	mismatch	gapopen
euraff	CLEC010822-RA:cDNA	Scaffold26	72.941	255	67	2
euraff	CLEC010822-RA:cDNA	Scaffold26	66.116	242	76	3
euraff	CLEC010822-RA:cDNA	Scaffold26	70.588	102	27	1
euraff	CLEC010822-RA:cDNA	Scaffold26	74.667	75	15	2
euraff	CLEC010822-RA:cDNA	Scaffold26	84.615	39	2	2
euraff	CLEC010822-RA:cDNA	Scaffold26	95.652	23	1	0
euraff	CLEC010822-RA:cDNA	EAFF012888-RA	68.861	395	111	5
euraff	CLEC010822-RA:cDNA	EAFF012888-RA	70.854	199	53	3
euraff	CLEC010822-RA:cDNA	EAFF027773-RA	68.243	444	128	7
euraff	CLEC010822-RA:cDNA	EAFF027773-RA	68.773	269	75	3
euraff	CLEC010822-RA:cDNA	EAFF003903-RA	71.138	246	63	3

2 of 2

qseqid filter

sseqid filter

Download

2022 - National Agricultural Library

BLAST Report

FASTA

Sbjct 258169 GTGAACGAGTTCAAGCGCAAGCACAAG---AAGGATAT 258135

Score = 142 bits (157), Expect = 3e-31

Identities = 186/255 (73%), Gaps = 2/255 (1%)

Strand=Plus/Minus

Query 1722 CAGGTTTATGAGGAGAAAGGCCAATGACCAAGATAATCATTGCTCGGAAAATTTCGAT 1781

Sbjct 257309 CAGGTTTACGAGGGTGAGAGGGCAATGACCAAGGACAATCACCTCCTTGTTAAATTTGAC 257250

Query 1782 TTGACAGGAATACCGCCTGCACCAAGGGAGTGCCACAGATTGAAGTCACCTTTTGATC 1841

Sbjct 257249 CTGAACGCCATCCCCCGCGCCAGGGGAGTCCCTCAGATTGAAGTCACCTTTGACATC 257190

Query 1842 GATGCCAACGGTATCCTTCAGGTGTCCGCGAGGACAAGGGAACGGGCAA-CAGAGAGAA 1900

Sbjct 257189 GACCGGAACGGCATCCTAAACGTTTCTGCATGCGACAAGTCGACTGGGAAGCAGAGTAAG 257130

Query 1901 AATAGTCATCACAACGACCAAGACAGGTTGACTCCAGACGACATCGATAGGATGATCAA 1960

Sbjct 257129 ATCACT-ATCAGCAACGACAAGGGTAGGCTGAGCAAGGAGGAGATTGAGAGGATGTCAC 257071

Query 1961 GGACGCCGAGAAGTT 1975

Sbjct 257070 CGACGCCGAGAAGTT 257056

Score = 59.9 bits (65), Expect = 3e-06

Identities = 160/242 (66%), Gaps = 6/242 (2%)

Strand=Plus/Minus

View in Jbrowse, turn on evidence tracks

The image displays two screenshots of the JBrowse genome browser interface, illustrating the process of turning on evidence tracks.

Left Screenshot: The interface shows the "Available Tracks" panel on the left. The "Evidence" track is currently turned off. The main view displays a genomic region on Scaffold188, with a red box highlighting the position 400,000. The "BLAST Results" track is visible below the main view.

Right Screenshot: The interface shows the same genomic region, but the "Evidence" track is now turned on. The "Evidence" track displays several tracks, including "EAFV_v0.5.3-Models" and "EaffTtmpM021200-RA". The "BLAST Results" track is also visible below the main view.

The "Available Tracks" panel on the left lists various tracks, including:

- 1. Gene Sets (3 tracks)
 - Primary Gene Sets: Protein Coding (1 track)
 - ☐ EAFV_v0.5.3-Models
 - Supplementary Gene Predictions (2 tracks)
 - ☐ augustus_masked
 - ☐ snap_masked
- 2. Evidence (2 tracks)
 - Repeats (2 tracks)
 - ☐ repeatmasker
 - ☐ repeatrunner
- 3. Mapped Proteins (41 tracks)
 - Other (1 track)
 - ☐ cegma
 - Protein2genome (20 tracks)
 - ☐ protein2genome_Annelida
 - ☐ protein2genome_Arthropoda
 - ☐ protein2genome_Atelerata
 - ☐ protein2genome_Cephalochordata
 - ☐ protein2genome_Chelicerata
 - ☐ protein2genome_Cnidaria
 - ☐ protein2genome_Craniala
 - ☐ protein2genome_Crustacea
 - ☐ protein2genome_Echinodermata
 - ☐ protein2genome_Mollusca
 - ☐ protein2genome_Nemata
 - ☐ protein2genome_Nematomorpha
 - ☐ protein2genome_Onychophora
 - ☐ protein2genome_Parazoa
 - ☐ protein2genome_Placozoa
 - ☐ protein2genome_Platyhelminthes
 - ☐ protein2genome_Priapulida
 - ☐ protein2genome_Tardigrada
 - ☐ protein2genome_Tunicata
 - ☐ protein2genome_UNCATEGORISED
- Supplementary: BlastX (20 tracks)
 - ☐ blastx_Annelida

Login to Apollo

Available Tracks

filter tracks

GenBank Annotations 2

1. Gene Sets 2

Protein Coding 1

GenBank Annotations, protein-coding

Pseudogenes 1

GenBank Annotations, pseudogenes

Reference Assembly 2

GC Content

Gaps in assembly

apollo File View Help

0 200,000 400,000 600,000 800,000 1,000,000 1,200,000 1,400,000 1,600,000

GenBank Annotations, protein-coding

GenBank Annotations, pseudogenes

GC Content

Gaps in assembly

apollo

https://apollo.nal.usda.gov/apollo/auth/login?targetUri=/annotator/loadLink?&loc=Scaffold188:397981..403760&addStores.url.type=JBrowse/Store/SeqFeature

Upon Login You Agree to the Following Information

- You are accessing a U.S. Government information system, which includes (1) this computer, (2) this computer network, (3) all computers connected to this network, and (4) all devices and storage media attached to this network or to a computer on this network. This information system is provided for U.S. Government authorized use only.
- Unauthorized or improper use of this system may result in disciplinary action, as well as civil and criminal penalties.
- By using this information system, you understand and consent to the following:
 - You have no reasonable expectation of privacy regarding any communications or data transmitted or stored on this information system. At any time, the government may for any lawful government purpose monitor, intercept, search and seize and communication or data transmission or storage on this information system.
 - Any communications or data transmitting or storing on this information system may be disclosed or used for any lawful government purpose.
 - Your consent is final and irrevocable. You may not rely on any statements or informal policies purporting to provide you with any expectation of privacy regarding communications on this system, where oral or written, by your supervisor or any other official, except USDA's Chief Information Officer.

Username (email)

Password

Login

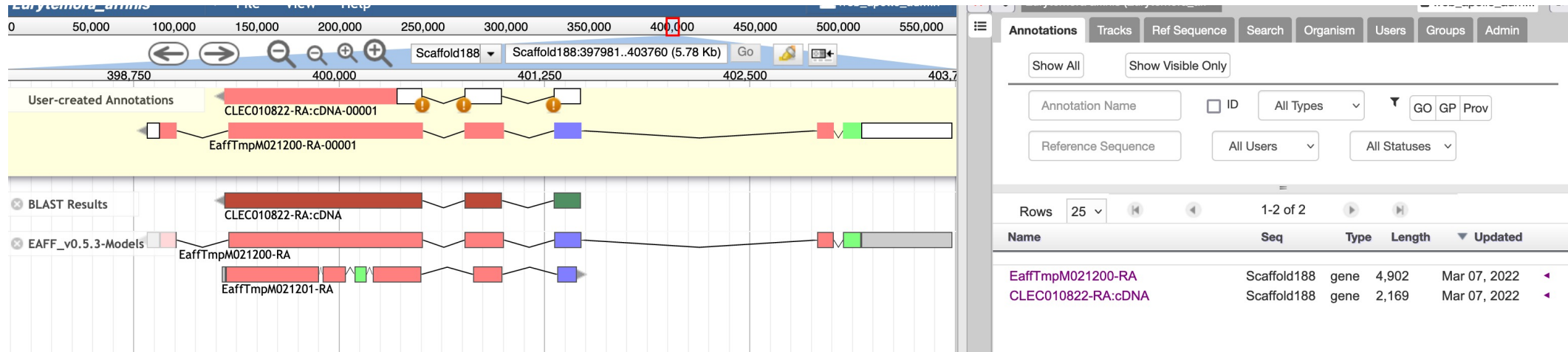
☐ Remember me

Start curating

The screenshot displays a genome browser interface. The top track shows a genomic region from 0 to 550,000. Below this, a track labeled 'User-created Annotations' is visible. A specific annotation, 'CLEC010822-RA: cDNA', is highlighted in red. A context menu is open over this annotation, offering options: 'View details', 'Highlight this match', and 'Create new annotation'. The 'Create new annotation' menu is expanded, showing a list of annotation types: 'gene', 'Pseudogenes', 'Non-coding RNA', 'repeat region', 'terminator', and 'transposable element'. On the right side of the interface, there is a sidebar with tabs for 'Annotations', 'Tracks', 'Ref Sequence', 'Search', 'Organism', 'Users', 'Groups', and 'Admin'. The 'Annotations' tab is active, showing a search interface with fields for 'Annotation Name', 'ID', 'All Types', 'GO', 'GP', and 'Prov'. Below these fields, there are filters for 'Reference Sequence', 'All Users', and 'All Statuses'. A table at the bottom of the sidebar shows 'Rows' (25), '1-1 of 0', and a table header with columns: 'Name', 'Seq', 'Type', 'Length', and 'Updated'. The table currently displays 'No results'.

- Annotations may be created by
- Click-dragging a model to the User created Annotations panel
- Right-clicking a model then clicking through the “Create new annotation” menu

A few Apollo tips



The screenshot displays the Apollo genome browser interface. The top track shows a genomic region on Scaffold188 from 397,981 to 403,760 (5.78 Kb). Below this, the 'User-created Annotations' track shows two gene models: CLEC010822-RA:cDNA-00001 and EaffTmpM021200-RA-00001. The 'BLAST Results' track shows a comparison of these models. The 'EAFf_v0.5.3-Models' track shows several other models, including CLEC010822-RA:cDNA, EaffTmpM021200-RA, and EaffTmpM021201-RA. The right sidebar contains the 'Annotations' panel with filters for 'Show All', 'Show Visible Only', 'Annotation Name', 'ID', 'All Types', 'GO', 'GP', 'Prov', 'Reference Sequence', 'All Users', and 'All Statuses'. Below the filters is a table with 25 rows, showing 1-2 of 2 results.

Name	Seq	Type	Length	Updated
EaffTmpM021200-RA	Scaffold188	gene	4,902	Mar 07, 2022
CLEC010822-RA:cDNA	Scaffold188	gene	2,169	Mar 07, 2022

- The Color by CDS feature is useful for fast comparisons
- The CDS is auto-calculated whenever a change is made
- Special symbols show if exons don't end with canonical splice sites
- Edge matching – select a model and other models that have the same splice sites become highlighted in red (next slide)

Right-click menu for many options

The screenshot displays a genomic browser interface. The top track shows a genomic map with coordinates from 0 to 550,000. A specific region, Scaffold188:400423..400805 (383 b), is highlighted. Below this, several tracks are visible, including 'User-created Annotations', 'BLAST Results', and 'EAFF_v0.5.3-Models'. A right-click context menu is open over the 'User-created Annotations' track, listing various actions such as 'Get Sequence (meta-click)', 'Zoom to Base Level', 'Change annotation type', 'Associate Transcript to Gene', 'Delete', 'Merge', 'Split', 'Duplicate', 'Make Intron', 'Move to Opposite Strand', 'Set Translation Start', 'Set Translation End', 'Set Longest ORF', 'Set Optimal ORF', 'Remove CDS', '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'.

Annotations

Tracks Ref Sequence Search Organism Users Groups Admin

Show All Show Visible Only

Annotation Name ID All Types GO GP Prov

Reference Sequence All Users All Statuses

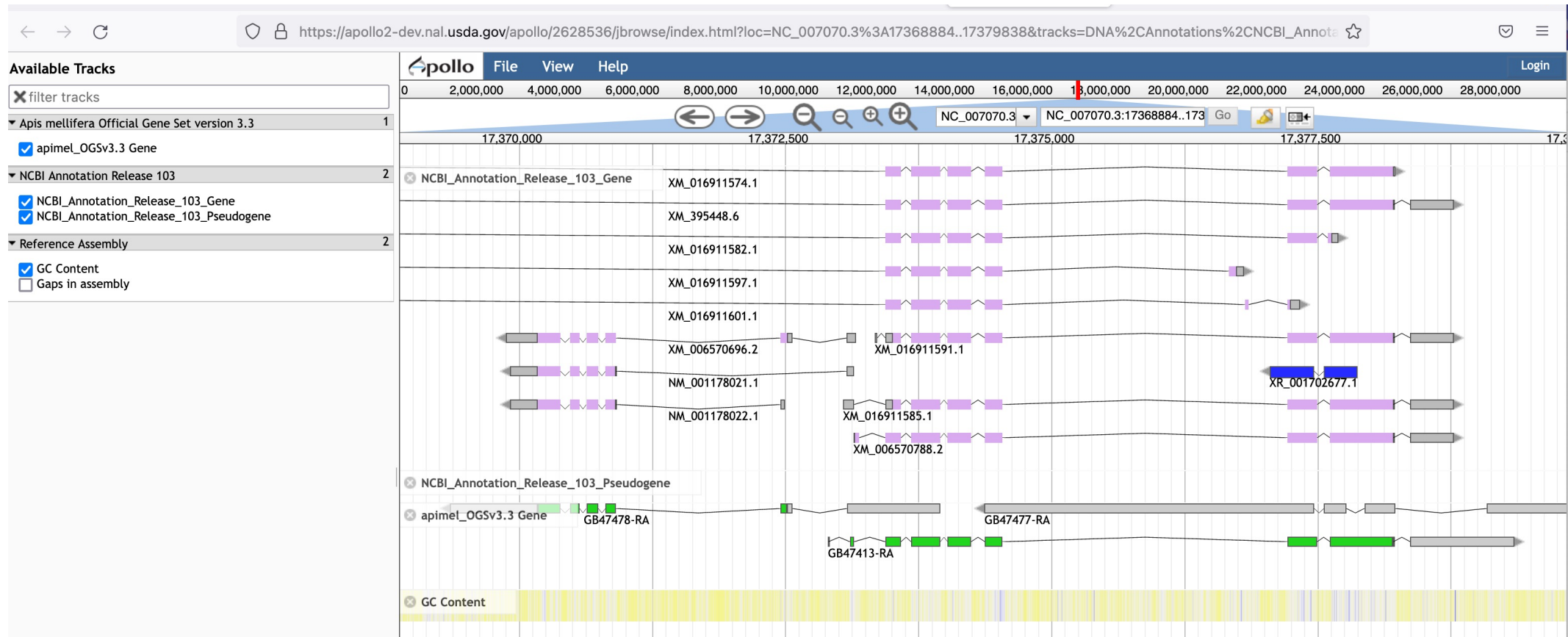
Rows 25 1-2 of 2

Name	Seq	Type	Length	Updated
EaffTmpM021200-RA	Scaffold188	gene	4,902	Mar 07, 2022
CLEC010822-RA:cDNA	Scaffold188	gene	2,169	Mar 07, 2022

Updates in Apollo 2.6.5

- Many under-the-hood updates
 - Bug fixes, improved responsiveness
 - Supports newer versions of Jbrowse
- User interface changes
 - Better
 - Faster
 - More stable
 - Updated styles

Updated styles



“NeatFeatures” – draw introns to show the direction a model goes
New built-in support for color by feature type (previously was a plugin)

Apollo updates in the latest release

- More functions are moved into the right side panel
- Updated Search features
 - New results
 - Hits can be used to create annotations

Annotation Editor updates

- More functions moved to the right side panel
- Gene is the default level
- GO Annotations, name, etc
- Many more validation steps

The screenshot displays the Eurytemora affinis Annotation Editor interface. The left panel shows a genomic track with various annotations, including 'User-created Annotations', 'manual-annotations', 'EAFF_v0.5.3-Models', 'augustus_masked', and 'snap_masked'. The right panel provides a detailed view of a specific annotation, 'Scaffold1:4210025-4211858-00001', which is a gene. The right panel includes a search bar, filters for 'All Types', 'GO', 'GP', and 'Prov', and a table of annotations. The table has columns for 'Name', 'Seq', 'Type', 'Length', and 'Updated'. The selected annotation is a gene with a length of 1,833 and was updated on Mar 29, 2022. Below the table, the 'Details' tab is active, showing fields for 'Name', 'Symbol', 'Aliases', 'Description', 'Location', 'Ref Sequence', 'Owner', 'Created', and 'Updated'.

Name	Seq	Type	Length	Updated
Scaffold1:4210025-4211858	Scaffold1	gene	1,833	Mar 29, 2022
Scaffold1:4210025-4211858-00001		mRNA	1,833	Mar 29, 2022

gene: Scaffold1:4210025-4211858 [Link to annotation](#) [Close\(x\)](#)

Details GO Gene Product Provenance DbXref Comment Attributes

☐ Sync name with transcript ☐ Obsolete

Type **gene** Status No status selected

Name Scaffold1:4210025-4211858

Symbol

Aliases (!" separated)

Description

Location 4210026 - 4211858 strand(+) Partial: ☐ 3' ☐ 5'

Ref Sequence Scaffold1

Owner cchilders

Created Mar 29, 2022 12:38 AM

Updated Mar 29, 2022 12:38 AM

Built in BLAT search

- Moved to the right-side panel
- Updated Search features
 - New results
 - Hits can be used to create annotations

The screenshot displays the Eurytemora affinis genome browser interface. The left panel shows a genomic track with a yellow background and a blue line representing a BLAT search hit. The right panel shows the BLAT search results and a table of hits.

BLAT Search Results:

Search: ☐ All genome sequences

MLKGRTRTTTPSKHLDPSTPIKQMDTPPTPSLKERSPDKRLKNSVLRSSMKLNHWSKD
MKEVCLVTTQDGVYNFRIAGGSDKGEFPIISDVLPDPTRGVEYKGDGELVTQDVLLEVQ
QAVSGYTQSDVVSWINOCTRGSNPNLK

1-2 of 2

ID	Start	End	Strand	Score	Significance	Identity	Action
Scaffold1	4,210,398	4,210,730	1	228	0	1	✓ --
Scaffold1	4,211,748	4,211,858	1	75	0	1	Save sequence Create annotation

Special instructions for *E. affinis*

- This is a community that transitioned from Apollo1 to the newest Apollo during their annotation window.
 - Special considerations
 - The previously created annotations were migrated as a new data track
 - Only visible to logged in users
 - Will have to be harmonized with newly created annotations for generation of the new OGS/submission to NCBI
 - Annotation editor window now defaults to Gene level.
 - Annotation information may still need to be entered into the transcript section

Training session for *E. affinis*

- <https://i5k.nal.usda.gov/training/webapp/blast>

Thanks!

If you have questions or comments reach out to us at

i5k@usda.gov