

## **MycoCosm: Manual Gene Curation**

### **JGI MycoCosm: Gene models open to editing by collaborating scientists**


The JGI Fungal annotation pipeline uses several gene prediction algorithms, including ab-initio, homology, and EST-based gene modelers to produce multiple overlapping gene models for a given locus. A heuristic filtering process chooses the “best” model at each locus according to specific weights given to each model based on evidence, completeness, homology, presence of known domains, and structures. These filtered models are stored in the “FilteredModels” track on the JGI browser. A copy of the FilteredModels is stored as the GeneCatalog. Users with specific privileges (e.g. collaborating scientists) can modify, add, or remove models from the GeneCatalog using available manual curation tools. These corrected gene models eventually become the reference list of gene models for this organism.

---

# How to use the JGI MycoCosm platform

## The Transcript Annotation tool

If you are a registered user, you can annotate a genome with information about the gene you are viewing. This is accomplished via the **Transcript Annotation tool**, which displays annotation information for the gene, and allows a user to modify several fields, including a model's Disposition by promotion (or demotion) to (or from) GeneCatalog.



**MycoCosm**  
THE FUNGAL GENOMICS RESOURCE

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#) [PHYCOCOSM](#) [LOGOUT](#) [SARA CALHOUN \(SECALHOUN\)](#) [SUPERUSER](#)

Absidia padenii NRRL 2977 v1.0

SEARCH BLAST BROWSE ANNOTATIONS MCL CLUSTERS SYNTENY MITOCHONDRION DOWNLOAD INFO HOME QC ADMIN STATUS HELP

**TRANSCRIPT ANNOTATION**

Chlpad1/scaffold\_27:259337-263722 fgenes1 kg.27 #\_328 #\_TRINITY\_DN7619\_c0\_g1\_i1 Hide

Attribute	Value	Creator	Action
Name			<a href="#">add</a>
Description			<a href="#">add</a>
Model Notes			<a href="#">add</a>
Define	Cse1-domain containing protein	AUTOMATIC	<a href="#">add</a>
Disposition	Catalog	AUTOMATIC	<a href="#">edit</a>
Literature			<a href="#">add</a>
Evidence	Type	Creator	Action
			<a href="#">add</a>

**FUNCTIONAL (PROTEIN) ANNOTATION**

**User-Assigned Ontology** [add](#)

**ASPECT** **DETAILS**


function	
process	
component	
enzyme	
kog	

**Automatic Ontology and Best Protein Alignments for transcript 436917**

ASPECT	GO/EC	SUPPORT	Action
function	5488 The selective, often stoichiometric, interaction of a molecule with one or more specific sites on another molecule.	IPR016024 Armadillo-type fold	<a href="#">add</a>
	5515 Interacting selectively with any protein or protein complex (a complex of two or more proteins that may include other nonprotein molecules).	IPR005043 CAS/CSE, C-terminal	<a href="#">add</a>
	8536 Interacting selectively with Ran, a conserved Ras-like GTP-binding protein, implicated in nucleocytoplasmic transport, cell cycle progression, spindle assembly, nuclear organization and nuclear envelope (NE) assembly.	IPR001494 Importin-beta, N-terminal domain	<a href="#">add</a>
process	6886 The directed movement of proteins in a cell, including the movement of proteins between specific compartments or structures within a cell, such as organelles of a eukaryotic cell.	IPR001494 Importin-beta, N-terminal domain	<a href="#">add</a>
		IPR013713 Exportin/Importin, Cse1-like	<a href="#">add</a>
kog	KOG1992 Nuclear export receptor CSE1/CAS (importin beta superfamily)		<a href="#">add</a>
	Intracellular trafficking, secretion, and vesicular transport		

**High Scoring Alignments** [Change Hit Filter](#)


gi|384490090|gb|EIE81312.1|  
hypothetical protein RO3G\_06017 [Rhizopus oryzae RA 99-880] [no tax name]

1  718

[view alignment](#) [view info](#)

GO/EC Classification	Score	E value	% id	% target	% model
	2276	0.0	55%	100%	80%

gi|392579467|gb|EIW72594.1|  
hypothetical protein TREMEDRAFT\_41848 [Tremella mesenterica DSM 1558] [no tax name]

23  963

[view alignment](#) [view info](#)

**Name** (GenBank “gene”) provides a unique, organism-specific identifier which should be consistent with community standards.

**Description** (GenBank “note”) provides a place to record information. Can be as detailed as needed, provided that the information is accurate and useful to researchers not familiar with the type of protein.

**Define** (GenBank “product”) provides a precise description of the gene and gene product, and if possible, it should include the gene's main function(s). Very often, the define of a related entry in Swissprot can be used.

**Disposition** provides two options regarding a models inclusion in GeneCatalog:

- “Catalog” for addition
- “Demote” for removal

There are multiple ways of accessing the transcript annotation page for a given gene model:

1. Via the View/Modify manual annotation link on the gene model's Protein page:

**Browse • Absidia padonii NRRL 2977 v1.0**

SEARCH BLAST **BROWSE** ANNOTATIONS ▾ MCL CLUSTERS SYNTENY MITOCHONDRION DOWNLOAD INFO HOME QC ADMIN STATUS **HELP!**

Name: CE85965\_60  
Protein ID: 85966  
Location: scaffold\_2:303440-305355  
Strand: +  
Number of exons: 6  
Description: Longest ORF from: 166 to 1410 breakup#1  
Best Hit: gj|384487169|gb|EIE79349.1| hypothetical protein RO3G\_04054 [Rhizopus oryzae RA 99-880] (model%: 78, hit%: 96, score: 1030, %id: 56) [no tax name]  
total hits(shown) 162 (10)

**KOG GROUP** Metabolism  
**KOG Id** KOG1055  
**KOG Class** Amino acid transport and metabolism  
**KOG Desc** GABA-B ion channel receptor subunit GABABR1 and related subunits, G-protein coupled receptor superfamily

[View/modify manual annotation](#)  
[View nucleotide and 3-frame translation](#) [To Genome Browser](#)  
[NCBI blastp](#) Predicted number of transmembrane domains: 0

CE85965\_60 To Genome Browser

☐ PF01094: Receptor family ligand binding region [HMM/ Pfam]  
☐ SSF53822: [SUPERFAMILY]

Flip Start	End	Len	%C	%I	Score	Description [taxName]
10	355	358	97%	56%	1030	nr_b_b_384487169 hypothetical protein RO3G_04054 [Rhizopus oryzae RA 99-880] [no tax name]
9	237	300	76%	23%	414	nr_b_b_384500255 hypothetical protein RO3G_15457 [Rhizopus oryzae RA 99-880] [no tax name]
8	202	880	22%	14%	232	nr_b_b_384485528 hypothetical protein RO3G_02412 [Rhizopus oryzae RA 99-880] [no tax name]
73	221	858	17%	22%	183	nr_b_b_13994201 Taste receptor, type 1, member 3 [synthetic construct] [no tax name]
73	221	858	17%	22%	183	nr_b_b_14190002 AF368024.1 putative sweet taste receptor family 1 member 3 [Mus musculus] [no tax name]

[View/modify manual annotation](#)  
[View nucleotide and 3-frame translation](#) [To Genome Browser](#)  
[NCBI blastp](#) Predicted number of transmembrane domains: 0


CE85965\_60 To Genome Browser

☐

## 2. Via Advanced Searching directly against annotations

- Gene models which match the specified search criteria are returned as a table, sorted by relevance score. The Gene column provides the following links:

- Protein Id: Link to the Protein page
- Transcript Id: Link to the Transcript Annotation page
- Location: Link to the genome browser, zoomed on the gene model



**MycoCosm**  
THE FUNGAL GENOMICS RESOURCE

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#) [PHYCOCOSM](#) [LOGOUT](#) [SARA CALHOUN \(SECALHOUN\)](#) [SUPERUSER](#)

Search • **Absidia padenii** NRRL 2977 v1.0

SEARCH

BLAST

BROWSE

ANNOTATIONS

MCL CLUSTERS

SYNTENY

MITOCHONDRION

DOWNLOAD

INFO

HOME

QC

ADMIN

STATUS

HELP!

pyruvate kinase

Search

Search By:

Keywords

Across:

Default

Terms:

exact - fast

Download

as CSV

compressed by Gzip

Total genes found: 19

25 rows per page


Gene	Gene Ontology	Annotations	User Annotations
<div>Portal: <a href="#">Chlpad1</a> Portal Name: <b>Absidia padenii</b> NRRL 2977 v1.0 Protein Id: <a href="#">491985</a> Transcript Id: <a href="#">492191</a> Location: <a href="#">scaffold_1:1548637-1550252 (+)</a> Model Name: <b>estExt_fgenes1_pm.C_10277</b> Track: <b>estExt_fgenes1_pm</b></div>		<b>KOG0787</b> • Dehydrogenase kinase <b>EC:2.7.11.2</b> • [pyruvate dehydrogenase (acetyl-transferring)] kinase <b>PF10436</b> • Mitochondrial branched-chain alpha-ketoacid dehydrogenase kinase <a href="#">IPR018955</a> •	
<div>Portal: <a href="#">Chlpad1</a> Portal Name: <b>Absidia padenii</b> NRRL 2977 v1.0 Protein Id: <a href="#">500858</a> Transcript Id: <a href="#">501064</a> Location: <a href="#">scaffold_7:1186163-1187538 (+)</a> Model Name: <b>estExt_fgenes1_pg.C_70398</b> Track: <b>estExt_fgenes1_pg</b></div>		<b>KOG0787</b> • Dehydrogenase kinase <b>EC:2.7.11.2</b> • [pyruvate dehydrogenase (acetyl-transferring)] kinase <b>PF10436</b> • Mitochondrial branched-chain alpha-ketoacid dehydrogenase kinase <a href="#">IPR018955</a> •	
<div>Portal: <a href="#">Chlpad1</a> Portal Name: <b>Absidia padenii</b> NRRL 2977 v1.0 Protein Id: <a href="#">379092</a> Transcript Id: <a href="#">379298</a> Location: <a href="#">scaffold_19:621926-623925 (+)</a> Model Name: <b>estExt_Genewise1.C_190458</b> Track: <b>estExt_Genewise1</b></div>		<b>KOG0787</b> • Dehydrogenase kinase <b>EC:2.7.11.2</b> • [pyruvate dehydrogenase (acetyl-transferring)] kinase <b>PF02518</b> • Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase <b>PF10436</b> • Mitochondrial branched-chain alpha-ketoacid dehydrogenase kinase <a href="#">IPR005467</a> • <a href="#">IPR018955</a> • <a href="#">IPR003594</a> •	

Gene

Portal: [Chlpad1](#)  
Portal Name: **Absidia padenii** NRRL 2977 v1.0  
Protein Id: [491985](#)  
Transcript Id: [492191](#)  
Location: [scaffold\\_1:1548637-1550252 \(+\)](#)  
Model Name: **estExt\_fgenes1\_pm.C\_10277**  
Track: **estExt\_fgenes1\_pm**

### 3. Via the GO/KEGG/KOG functional tools

- These utilities provide dynamic lists of gene models which match functional search criteria specific to the particular functional category
- (GO) For gene models belonging to a particular GO category, the Links column contains the following:
  - P: Link to the Protein page
  - A: Link to the Transcript Annotation page



[JGI HOME](#)
[GENOME PORTAL](#)
[MYCOCOSM](#)
[PHYCOCOSM](#)
[LOGIN](#)

**Gene Ontology • Absidia padenii NRRL 2977 v1.0**

[SEARCH](#)
[BLAST](#)
[BROWSE](#)
[ANNOTATIONS](#)
[MCL CLUSTERS](#)
[SYNTENY](#)
[DOWNLOAD](#)
[INFO](#)
[HOME](#)
[STATUS](#)
[HELP!](#)

**Text Search:**

Select Model Set(s) to View:
 

Chlpad1:FilteredModels1 (run 1)  
 Gonbut1:FilteredModels1 (run 1)  
 Absrep1:FilteredModels1 (run 1)  
 Parpar1:FilteredModels1 (run 1)

Using GO dataset **go\_200804**


GO Term	Gene Models In Chlpad1	Total Gene Models
<input type="checkbox"/> all all	7234	7234
<input type="checkbox"/> GO:0008150 biological_process	4785	4785
<input type="checkbox"/> GO:0032502 developmental process	9	9

**Download:**

Name	ProteinId	Links	JGI DB/Batch	Quality	All Xref
<b>GO:0006915:apoptosis</b>					
<input type="checkbox"/> fgenes1_kg.27_#_328_#_TRINITY_DN7619_c0_g1_i1	436711	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR001494 IPR005043 IPR013713 IPR016024
<input type="checkbox"/> estExt_Genemark1.C_220019	516041	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR000626 IPR003103
<input type="checkbox"/> estExt_Genewise1Plus.C_190208	399972	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR000626 IPR003103
<input type="checkbox"/> fgenes1_kg.11_#_587_#_TRINITY_DN8456_c0_g2_i1	424453	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR003103
<input type="checkbox"/> fgenes1_kg.9_#_248_#_TRINITY_DN11231_c0_g1_i1	421633	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR003103
<input type="checkbox"/> e_gw1.8.763.1	349521	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR003103
<b>GO:0006916:anti-apoptosis</b>					
<input type="checkbox"/> fgenes1_kg.13_#_1156_#_TRINITY_DN5620_c0_g1_i1	427291	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR001370
<input type="checkbox"/> fgenes1_pg.2_#_591	448345	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR001370
<b>GO:0007275:multicellular organismal development</b>					
<input type="checkbox"/> fgenes1_pg.3_#_156	448656	<a href="#">P</a> <a href="#">A</a>	Chlpad1:1581	IEA	IPR003663 IPR005828 IPR005829 IPR016201 3.5.4.3

**Download:**

- c. (KEGG/KOG) For gene models belonging to a particular KEGG metabolic pathway (EC designation) or KOG functional group (KOG id), the Curated? column contains a YES/NO link to the Transcript Annotation page



**MycoCosm**  
THE FUNGAL GENOMICS RESOURCE

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#) [PHYCOCOSM](#) [LOGIN](#)

KEGG • Absidia padenii NRRL 2977 v1.0

SEARCH BLAST BROWSE ANNOTATIONS MCL CLUSTERS SYNTENY DOWNLOAD INFO HOME STATUS **HELP!**

Select Model Set(s) to Search:

Absidia padenii NRRL 2977 v1.0/FilteredModels1 (ver 1)

Gongronella butleri v1.0/FilteredModels1 (ver 1)

Absidia repens NRRL 1336 v1.0/FilteredModels1 (ver 1)

Parasitella parasitica v1.0/FilteredModels1 (ver 1)

Other Functions

[View KEGG Metabolic Pathways](#)

[View KEGG Regulatory Pathways](#)

[Search KEGG Enzyme Commission Numbers](#)

Search Options: 

EC Number

EC: 

1

4

1

2


search

EC Number	Definition	Alternative Name	Catalytic Activity	Cofactors	Associated Diseases
1.4.1.2	glutamate dehydrogenase	glutamic dehydrogenase	L-glutamate + H <sub>2</sub> O + NAD <sup>+</sup> = 2-oxoglutarate + NH <sub>3</sub> + NADH + H <sup>+</sup> [RN:R00243]		

Searching for gene models with association to EC 1.4.1.2 ... Done!

Found 2 model(s), displayed below:

Species	Model Set	Protein ID	Protein Name	Source	E-Value	Top KEGG Hit	Curated?
Absidia padenii NRRL 2977 v1.0	FilteredModels1 (ver 1)	411121	fgenes1_kg.2_#_1271_#_TRINITY_DN8017_c1_g1_i3	SW/KEGG	0	afm:AFUA_2G06000 K15371 (EC:1.4.1.2)	<a href="#">NO</a>
Absidia padenii NRRL 2977 v1.0	FilteredModels1 (ver 1)	500005	estExt_fgenes1_pg.C_50379	SW/KEGG	0	nfi:NFA_082760 K15371 (EC:1.4.1.2)	<a href="#">NO</a>



**MycoCosm**  
THE FUNGAL GENOMICS RESOURCE

[JGI HOME](#) [GENOME PORTAL](#) [MYCOCOSM](#) [PHYCOCOSM](#) [LOGIN](#)

KOG • Absidia padenii NRRL 2977 v1.0

SEARCH BLAST BROWSE ANNOTATIONS MCL CLUSTERS SYNTENY DOWNLOAD INFO HOME STATUS **HELP!**

CELLULAR PROCESSES AND SIGNALING

(N) Cell motility

Prot name	Prot id	KOG id	KOG Description	Curated?
<input type="checkbox"/> CE108962_1362	108963	KOG2116	Protein involved in plasmid maintenance/nuclear protein involved in lipid metabolism	<a href="#">NO</a>
<input type="checkbox"/> CE165210_9933	165211	KOG2116	Protein involved in plasmid maintenance/nuclear protein involved in lipid metabolism	<a href="#">NO</a>
<input type="checkbox"/> CE222428_6742	222429	KOG2116	Protein involved in plasmid maintenance/nuclear protein involved in lipid metabolism	<a href="#">NO</a>
<input type="checkbox"/> gw1.33.116.1	329174	KOG3896	Dynactin, subunit p62	<a href="#">NO</a>
<input type="checkbox"/> estExt_Genewise1.C_1_130133	365899	KOG4115	Dynein-associated protein Roadblock	<a href="#">NO</a>
<input type="checkbox"/> estExt_Genewise1Plus.C_1_110466	386557	KOG3905	Dynein light intermediate chain	<a href="#">NO</a>
<input type="checkbox"/> estExt_Genewise1Plus.C_5_120187	391566	KOG4229	Myosin VII, myosin IXB and related myosins	<a href="#">NO</a>
<input type="checkbox"/> fgenes1_kg.7_#_493_#_TRINITY_DN4302_c0_g2_i1	419153	KOG4229	Myosin VII, myosin IXB and related myosins	<a href="#">NO</a>
<input type="checkbox"/> fgenes1_kg.18_#_518_#_TRINITY_DN2973_c0_g2_i1	431242	KOG4081	Dynein light chain	<a href="#">NO</a>
<input type="checkbox"/> fgenes1_kg.36_#_150_#_TRINITY_DN6527_c0_g3_i1	439839	KOG2116	Protein involved in plasmid maintenance/nuclear protein involved in lipid metabolism	<a href="#">NO</a>
<input type="checkbox"/> fgenes1_pg.19_#_117	454278	KOG4242	Predicted myosin-I-binding protein	<a href="#">NO</a>
<input type="checkbox"/> fgenes1_pg.31_#_9	456013	KOG4229	Myosin VII, myosin IXB and related myosins	<a href="#">NO</a>
<input type="checkbox"/> fgenes1_pm.5_#_239	459836	KOG3896	Dynactin, subunit p62	<a href="#">NO</a>
<input type="checkbox"/> gm1.6859_g	470952	KOG3905	Dynein light intermediate chain	<a href="#">NO</a>
<input type="checkbox"/> estExt_Genemark1.C_4_t10092	509824	KOG2116	Protein involved in plasmid maintenance/nuclear protein involved in lipid metabolism	<a href="#">NO</a>

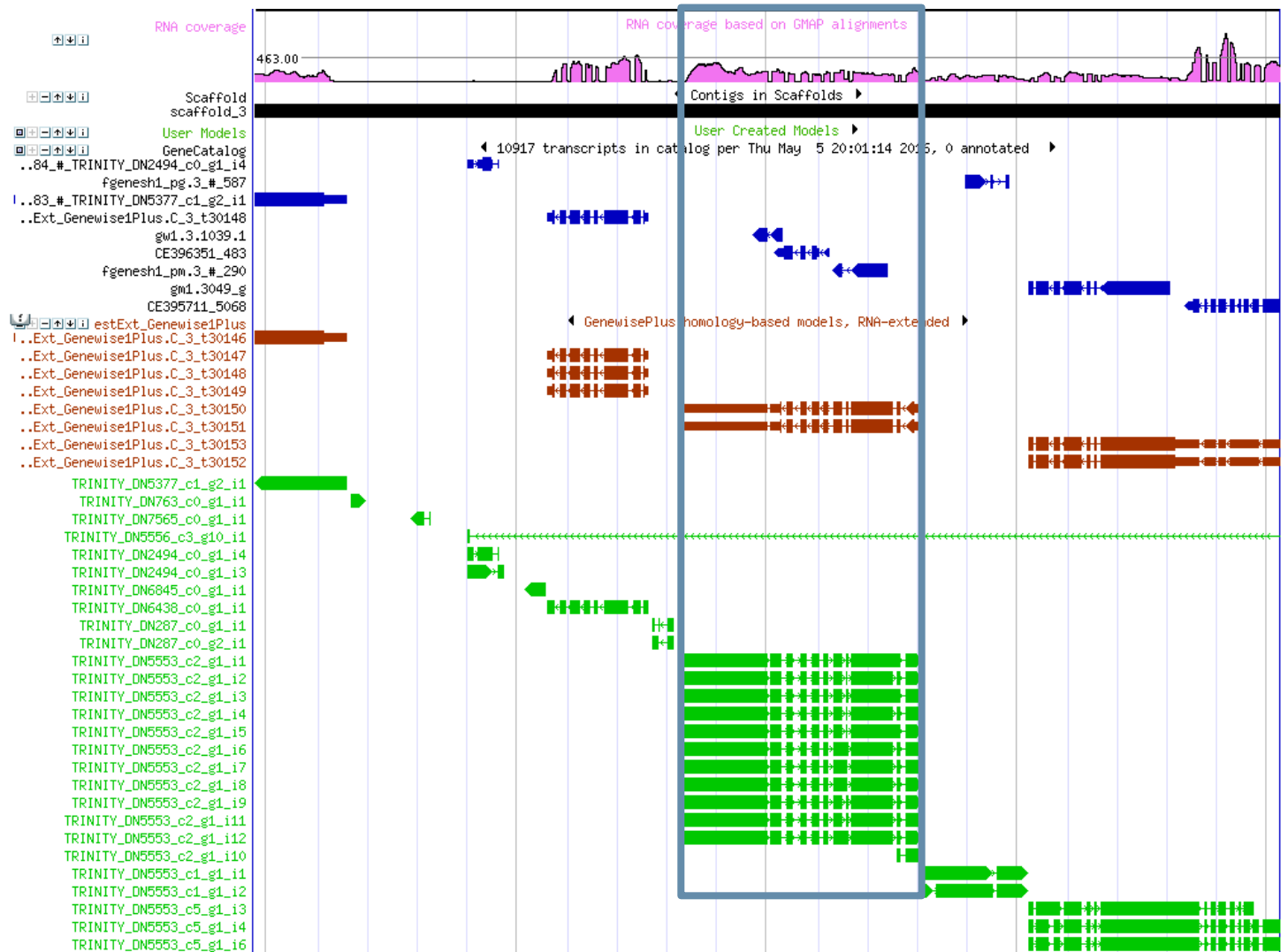
Retrieve FASTA

Clustalw

Uncheck all

## Model Promotion

To search for and evaluate alternative models at a given locus, expand all model tracks (red) and EST tracks (green). In many cases, a better model has already been generated by one of the gene predictors but was not promoted to GeneCatalog. For example, below is a view of select tracks displaying a long model covering three short fragment models, with EST and RNA coverage:



If an alternative model exists and is determined to be more accurate than the current model, it should be promoted to GeneCatalog. Use the Disposition field on the Transcript Annotation page to promote a model to GeneCatalog by setting the value to “Catalog”.

## Model Creation

If none of the alternative models are of acceptable quality, it will be necessary to create a model using the Track Editor tool: [http://genome.jgi.doe.gov/help/track\\_editor.html](http://genome.jgi.doe.gov/help/track_editor.html)

Using the Track Editor, it is possible to:

- Create a new model by copying an existing model
- Edit a new model
- Add existing exons to a new model
- Create an ab initio model

Once editing is finished, the model should be released in order to initiate protein analysis. However, since releasing a model does not automatically add it to the GeneCatalog, the model's Disposition must also be set to "Catalog" via the model's Transcript Annotation page (similar to Model Promotion).

## Model Demotion

Regardless of whether an existing model was promoted or a new model was created, the old/incorrect model should be demoted; otherwise, it will appear concurrently with the new/correct model. Similar to Model Promotion, use the Disposition field on the Transcript Annotation page to demote a model from GeneCatalog by setting the value to "Demote". This option does not delete the model or its annotation from the database. It simply removes it from the Catalog track.

---