

GEP Annotation Report

Student name: Kelly Hutchins
Student email: khutchin@unm.edu
Faculty Advisor: Dr. Paul Szauter
College/University: University of New Mexico

Project details

Project name: Fosmid 2594008
Project species: Drosophila ananassae
Date of submission: 7/9/2013
Size of project in base pairs: 36,277
Number Of genes in project: 1

Does this report cover all genes and all isoforms or is it a partial report? Full Report

Gene report form

Gene name (i.e. *D. mojavensis eyeless*): Enzyme olf413
Gene symbol (i.e. dmoj_ey): olf413-Pb
Approximate location in project (from 5' end to 3' end): 25,103-3370
*Missing 5' 2 exons

Number of isoforms in *D. melanogaster*: 2
Number of isoforms in this project: 2

Complete the following table for all the isoforms in this project:

Name of unique isoform based on coding sequence	List of isoforms with identical coding sequences
Olf413-Pb	Olf413-Pc

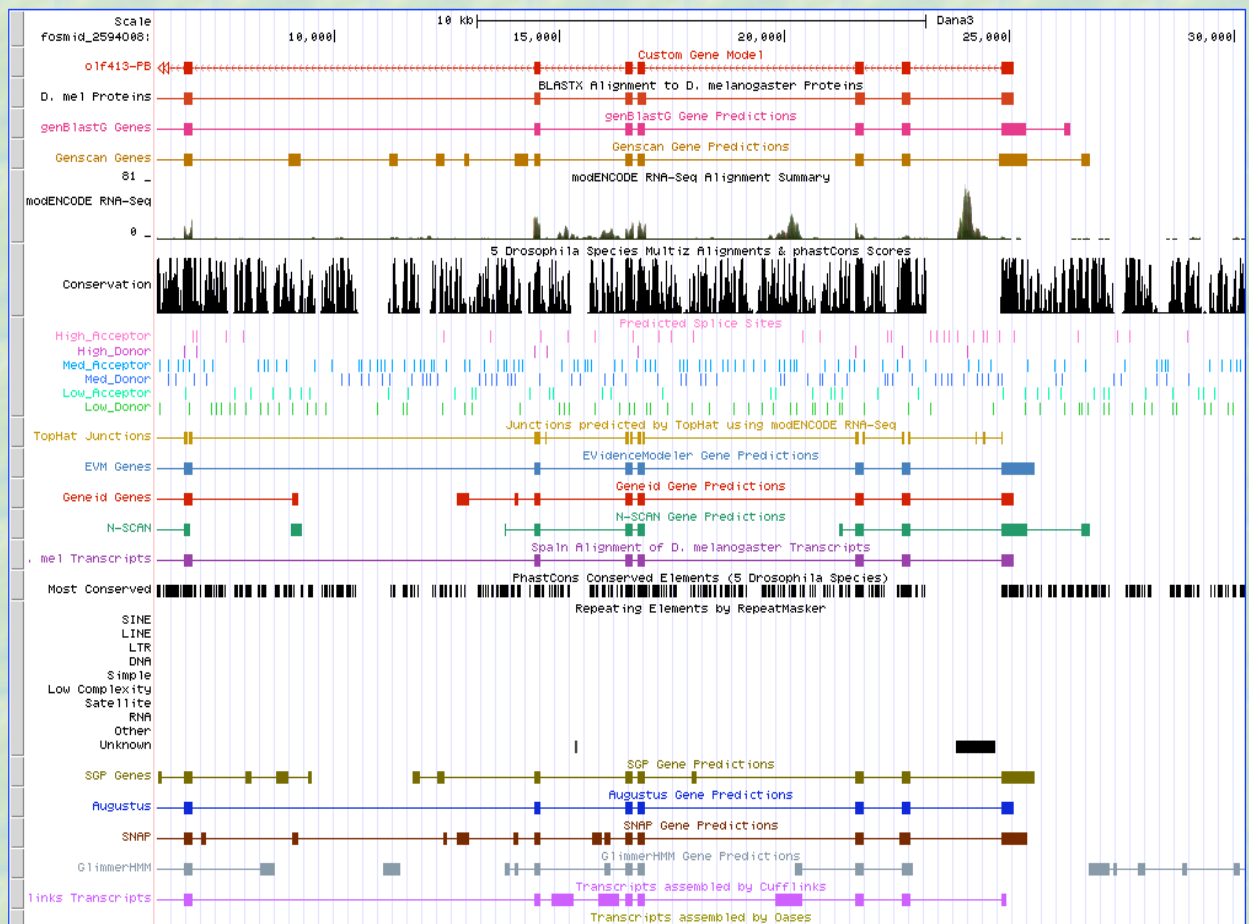
Isoform report form

Gene---isoform name (i.e. dmoj_ey---PA): olf413-Pb
Names of the isoforms with identical coding sequences as this isoform
olf413-Pc
Is the 5' end of this isoform missing from the end of project: Yes missing two exons
Is the 3' end of this isoform missing from the end of project: No

GEP UCSC Genome Browser on *D. ananassae* Jan. 2013 (GEP/3L Reference) Assembly (Dana3)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search fosmid_2594008:6,047-30,23 jump clear size 24,185 bp. configure



Gene Model Checker

Configure Gene Model

Model Details

Fosmid Sequence File: [Browse...](#)

Ortholog in *D. melanogaster*:

Coding Exon Coordinates:

Annotated Untranslated Regions? ☐ Yes ☒ No

Orientation of Gene Relative to Query Sequence: ☐ Plus ☒ Minus

Completeness of Gene Model Translation: ☐ Complete ☒ Partial

Region Missing: [v](#)

Phase of First Exon: [v](#)

Stop Codon Coordinates:

Project Details

Project Group: [v](#)

Project Name:

[Verify Gene Model](#) [Reset Form](#)

Checklist

[Dot Plot](#) [Transcript Sequence](#) [Peptide Sequence](#) [Extracted Coding Exons](#) [Downloads](#)

[Expand All](#) [Collapse All](#)

View	Criteria	Status	Message
	Check for Start Codon	Skip	Partial gene with 5' end missing
	Acceptor for CDS 1	Pass	
	Donor for CDS 1	Pass	
	Acceptor for CDS 2	Pass	
	Donor for CDS 2	Pass	
	Acceptor for CDS 3	Pass	
	Donor for CDS 3	Pass	
	Acceptor for CDS 4	Pass	
	Donor for CDS 4	Pass	
	Acceptor for CDS 5	Pass	
	Donor for CDS 5	Pass	
	Acceptor for CDS 6	Pass	
	Donor for CDS 6	Pass	
	Acceptor for CDS 7	Pass	
	Donor for CDS 7	Pass	
	Acceptor for CDS 8	Pass	
	Donor for CDS 8	Skip	Already checked for Stop Codon
	Check for Stop Codon	Pass	
	Additional Checks	Pass	

Alignment of olf413-PB vs. Submitted_Seq

[View plain text version](#)

Identity: 516/764 (67.5%), **Similarity:** 522/764 (68.3%), **Gaps:** 226/764 (29.6%)

olf413-PB	1	MAHPRKAVATPATLQLGPPAQTAQSPAATLRHSRTASSSRRLSLIRCFISCHTFNLFLLI	60
Submitted_Seq	1	-----	0
olf413-PB	61	LLLASGVRAASKLATRSNKTSGSSTASGVGACTAATSAAAAAASGTPIWHDHAIDLNDDFR	120
Submitted_Seq	1	-----	0
olf413-PB	121	ILWQIINQDITFEIQARTLG YVGFSPDGNLAGADMAIGWVDKGQTYFQDRHVTRNGDP	180
Submitted_Seq	1	-----	0
olf413-PB	181	EPVVDPSQDYMLMLGYENATHTVLRFRRKLDTCDP SHDIATTNDTMRLLYMYHAQDPPHG	240
Submitted_Seq	1	-----NDTMRLLYMYHAHDPPHG	18
olf413-PB	241	SVRPGTLPDPAARFRPYRPMVLMQRAQLPMPSPTHDERVRVLELRNEDVELPAGDLPLFW	300
Submitted_Seq	19	SVRPGTLPDPAGAFRPMVLMQRAQMPLPSPTHDERVRVLELRNEDVELPAGDLPLFW	78
olf413-PB	301	CKMFKLEDINRKHHLIRYEPIYDSSSSVHYLQHITLHECQGAHAELEEMAREQGRPCIGA	360
Submitted_Seq	79	CKMFKLEDINRKHHLIRYEPIYDSSSSVHYLQHITLHECQGAHAELEEMAREQGRQCLGA	138
olf413-PB	361	RSIPLACNAIVASWSRGSEGFYTPHEAGYPIESRQAKYYLMETHYNNLKPDPFAQLHARQM	420
Submitted_Seq	139	RSIPLACNAIVASWSRGSEGFYTPHEAGYPIESRQAKYYLMETHYNNLKPDPFAQLHARQM	198
olf413-PB	421	ADNSGLKIYFTHVLRPNDA GTLSIGMDPNWRHIIPPGQKRVVSEGQCIEDCTGYAFPPQG	480
Submitted_Seq	199	ADNSGLKIYFTHVLRPNDA GILSIGMDPNWRHIIPPGQKHVVSEGQCIEDCTGYAFPPQG	258
olf413-PB	481	INIFAVMMRTHQIGKEVKLRQIRQTEELPPIAHSNIDVAYQDFRRLPQSVHSMPCDRLI	540
Submitted_Seq	259	INIFAVMMRTHQIGKEVKLRQIRQTEELAPIAHSNIDVAYQDFRRLPQSVHSMPCDRLI	318
olf413-PB	541	AECIYDSSSRKAITLGGLTMKEESCTVLTLYYPRQKKLTTC HSLPSLPTVLHSLGIEQLA	600
Submitted_Seq	319	AECIYDSSSRKAITLGGLTMKEESCTVLTLYYPRQKKLTTC HSLPSLPTVLHSLGIEQLA	378
olf413-PB	601	TDSNPVLISPPPELAGMTLEARLISYDWENQFGEFQEATRKG SFKPICWGAKNHVVP GSE	660
Submitted_Seq	379	TDSNPVLISPPPELAGMTLEARLISYDWENQFGEFQEATRKG SFKPICWGAKNHVVP GSE	438
olf413-PB	661	FLEGYSINVTKTYKKHRRCKPKRPLAPPTERT-APPPA-SDLSELPVLHELDNNNIIEGA	718
Submitted_Seq	439	FLEGYSINVTKTYKKHRRCKPKRLFAPPTERTVAPQPTGSDLSELPVLHELDNNNIIEGA	498
olf413-PB	719	ARSSRSSATDVHSLSRGSGR--HFISCLLWLGASSWLLMLRT	760
Submitted_Seq	499	ARSSRSSATDVLGMSQGTGRAGEFSSRLLWLGASTWLLMLRT	542

Configure Gene Model

Model Details

Fosmid Sequence File:

fosmid_2594008.fasta

Browse...

Ortholog in *D. melanogaster*:

olf413-PB

Coding Exon Coordinates:

25103-24819, 22793-22608, 21773-21578,
16906-16737, 16631-16469, 14567-14432,
6826-6658, 3690-3370

Annotated Untranslated Regions?

No

Yes

Orientation of Gene Relative to Query Sequence:

Minus

Plus

Completeness of Gene Model Translation:

Partial

Complete

Region Missing:

Missing 5' end of translated region

Phase of First Exon:

Phase 0

Stop Codon Coordinates:

3369-3367

Project Details

Project Group:

D. ananassae 3L Control

Project Name:

fosmid_2594008

Verify Gene Model

Reset Form

Checklist

Dot Plot

Transcript Sequence

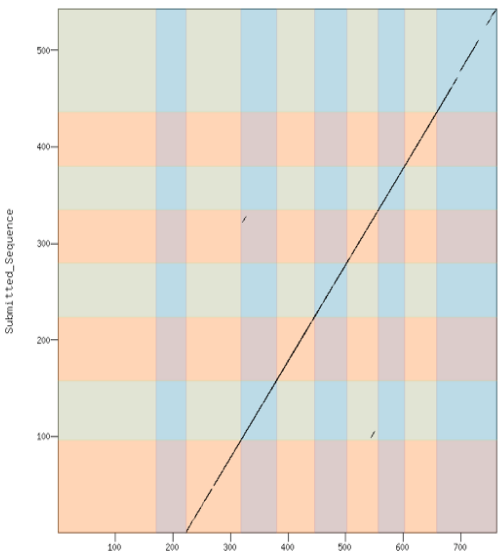
Peptide Sequence

Extracted Coding Exons

Downloads

View protein alignment

Dot plot of olf413-PB vs. Submitted_Sequence



Submitted_Sequence

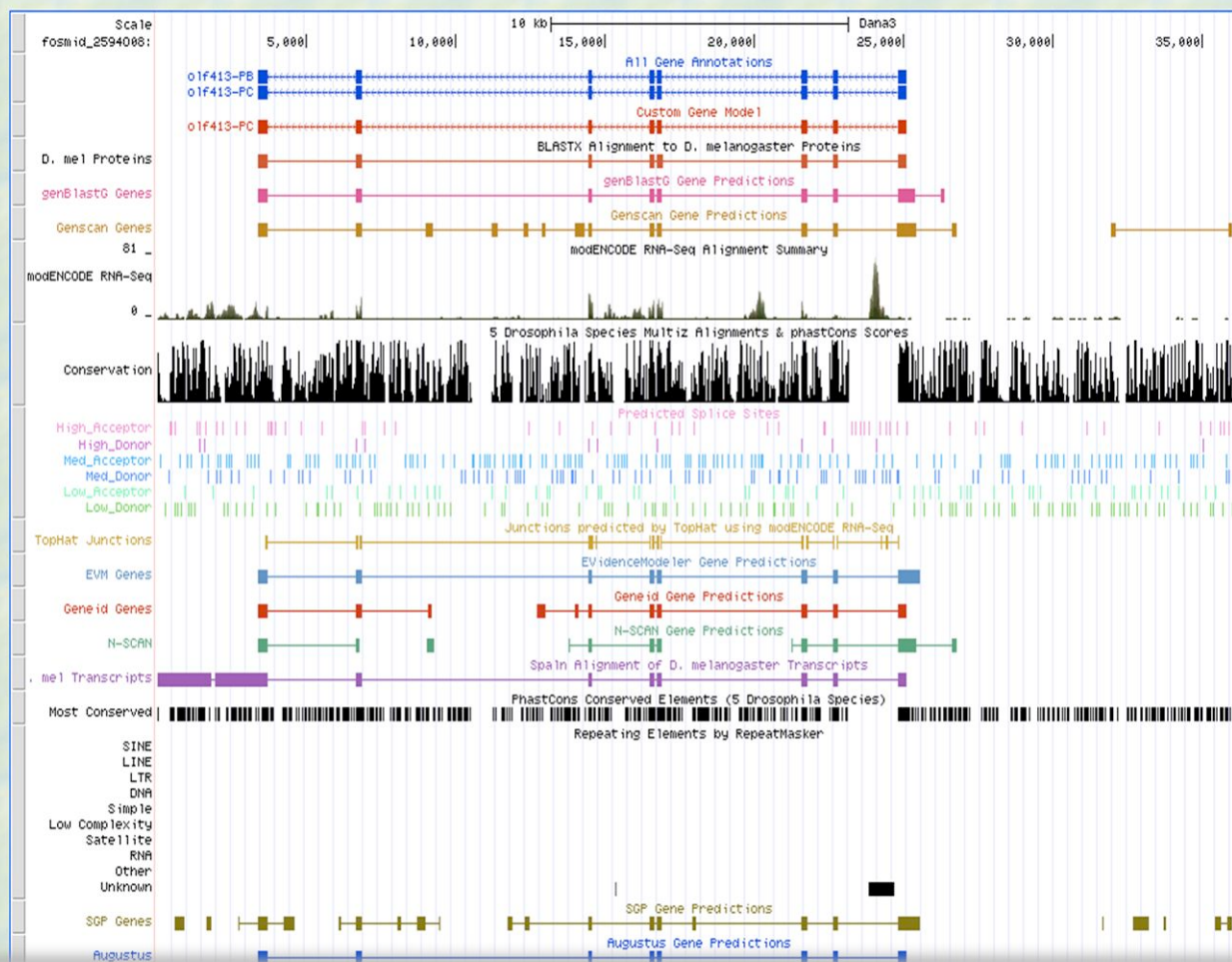
olf413-PB

External Links: [Old Gene Checker](#)

GEP UCSC Genome Browser on *D. ananassae* Jan. 2013 (GEP/3L Reference) Assembly (Dana3)

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x

position/search fosmid_2594008:1-36,277 jump clear size 36,277 bp. configure



Query ID Icd|81899
Description None
Molecule type amino acid
Query Length 109

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.2.28+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

Graphic Summary

[Show Conserved Domains](#)

No putative conserved domains have been detected

Distribution of 4 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments

Color key for alignment scores

<40	40-50	50-80	80-200	>=200
-----	-------	-------	--------	-------

Query 1 20 40 60 80 100

Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	extracellular matrix protein papilin 3 [Drosophila melanogaster]	26.6	26.6	40%	9.3	33%	AAO84908.1
☐	papilin, isoform E [Drosophila melanogaster] >sp Q868Z9.2 PPN_DROME RecName: Full=Papilin; Flags: Precursor >gb AAI	26.6	26.6	40%	9.3	33%	NP_788752.2
☐	papilin, isoform G [Drosophila melanogaster] >gb ACZ95055.1 papilin, isoform G [Drosophila melanogaster]	26.6	26.6	40%	9.6	33%	NP_001163761.1
☐	extracellular matrix protein papilin 2 [Drosophila melanogaster]	26.6	26.6	40%	9.9	33%	AAO84907.1

BLAST
[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

Basic Local Alignment Search Tool
My NCBI [Sign In](#) [Register](#)

NCBI/BLAST/blastp suite/Formatting Results - VF4H9Z96015

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Protein Sequence (109 letters)

Query ID Icd|32213
Description None
Molecule type amino acid
Query Length 109

Database Name nr
Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
Program BLASTP 2.2.28+ [Citation](#)

No significant similarity found. For reasons why, [click here](#)

Other reports: [Search Summary](#)