# **GEP** Annotation Report

Student name: N'Djamena D. Marmon Student email: ndjamenamarmon@gmail.com Faculty advisor: Paul Szauter College/University: University of New Mexico

# **Project Details**

Project name: contig50 Project species: *Drosophila biarmipes* Date of submission: July 25, 2013 Size of project in base pairs: 40,000 Number of genes in project: 2

This report covers all genes and all isoforms in the project.

# **GENE REPORT FORM**

Gene name: eukaryotic translation initiation factor 4G Gene symbol: eIF4G Approximate location in project (from 5' end to 3' end): 39443-16562 Number of isoforms in *D. melanogaster*: 3 Number of isoforms in project: 3

Name of unique isoform based on coding sequence	List of isoforms with identical coding sequence			
eIF4G-PA	eIF4G-PC			
eIF4G-PB				

#### **ISOFORM REPORT FORM**

Gene-isoform name: eIF4G-PA Names of the isoforms with identical coding sequences as this isoform: eIF4G-PC Is the 5' end of this isoform missing from the end of the project: No Is the 3' end of this isoform missing from the end of the project: No

## **1. Gene Model Checker Checklist**

C	ene Model Checker											
C	onfigure Gene Model			~	C	hecklis	t Dot Plot	Transcript Sequence	Peptide	Sequence	Extracted Coding Exor	ns
	Model Details			-	ŧ	Expand	All   E Collap	ose All				
1	Fosmid Sequence File:	C:\fakepath\contig	50.fasta Browse.	1		Vi	Criteria		Status	Message		
	Ortholog in D. melanogaster:	eIF4G-PA			E	Q	Check for Star	t Codon	Pass			-
Ш	Cading Even Coordinates	20142 20266 2522	1 24024 24620		Ð		Acceptor for C	CDS 1	Skip	Already	checked for Start Codon	
Ш	county Exon coordinates.	34440,33735-3362	4,29181-29126,24626-		Ð	Q	Donor for CDS	1	Pass			
Ш		22327,21699-2145	6,21399-21236,21078-		Ð	Q	Acceptor for 0	CDS 2	Pass			
L		16917,16782-1668	1,16620-16562	E	Ð	Q	Donor for CDS	2	Pass			ш
L					Ð	Q	Acceptor for 0	CDS 3	Pass			
L	Annotated Untranslated	Yes	No		Ð	Q	Donor for CDS	3	🕑 Pass			
L	Regions?	0.00	0		Ð	Q	Acceptor for 0	CDS 4	Pass			
	Orientation of Gene Relative to	O Plus	Minus		Ð	Q	Donor for CDS	4	🕑 Pass			
Ш	Query sequence.		na terreta de la		Ð	Q	Acceptor for 0	CDS 5	Pass			
L	Completeness of Gene Model Translation:	<ul> <li>Complete</li> </ul>	O Partial		٠	Q	Donor for CDS	5	Pass			
Л	Stop Codon Coordinates:	16561-16559			Ð	Q	Acceptor for 0	CDS 6	Pass			
				+	Ð	Q	Donor for CDS	6	Pass			
	•	Ш		F	Ð	Q	Acceptor for 0	CDS 7	Pass			
			Verify Gene Model Reset Fo	orm	Ð	Q	Donor for CDS	7	🕑 Pass			
					H	0	Accentor for (	05.8	Pace			*

#### Gene Model Checker

Configure Gene Model			«	C	necklis	t Dot Plot	Transcript Sequence	e Peptide	Sequence Extra	acted Coding Exons
	22327,21699-214	56,21399-21236,21078-		ŧ	Expan	i All 📔 🔚 Colla	pse All			
	20604,20396-201 16917,16782-166	73,17346-17200,17139- 81,16620-16562			Vi	Criteria	50	Status	Message	112
				Ð	Q	Acceptor for	CDS 9	@ Pass		^
Annotated Untranslated	Yes	No		Ð	Q	Donor for CDS	S 9	© Pass		
Regions?				Ŧ	Q	Acceptor for	CDS 10	O Pass		
Orientation of Gene Relative to Query Sequence:	🔘 Plus	Minus		Ð	Q	Donor for CDS	S 10	© Pass		
Completeness of Gene Model	Complete	🙆 Dartial		Ð	Q	Acceptor for	CDS 11	Pass		
Translation:	Complete	U Forciar		Ð	Q	Donor for CDS	S 11	O Pass		
Stop Codon Coordinates:	16561-16559		Е	Ð	Q	Acceptor for	CDS 12	© Pass		
			_	Ð	Q	Donor for CDS	S 12	Pass		
Project Details			-	Ŧ	Q	Acceptor for	CDS 13	@ Pass		
Project Group:	D. biarmipes Dot	~		Ð	Q	Donor for CDS	S 13	O Pass		
Project Name:	contia50			Ð	Q	Acceptor for	CDS 14	O Pass		É
				Ð		Donor for CDS	S 14	Skip	Already checked	d for Stop Codon
•	m		•	Ð	Q	Check for Sto	p Codon	© Pass		
		7 10		Ð	Q	Additional Che	ecks	Pass		
		Verify Gene Model Reset Form	1	±	Q	Number of co	ding exons matche	© Pass		-



#### 2. View the gene model on the Genome Browser

# 3. Alignment between submitted model and D. melanogaster ortholog

# Alignment of eIF4G-PA vs. Submitted\_Seq

### View plain text version

Identity: 1307/1681 (77.8%), Similarity: 1430/1681 (85.1%), Gaps: 75/1681 (4.5%)

eIF4G-PA	1	MQQAIPTLFTQSDIDKAMQPHSAQNMYVSTGNNNSGNTRSNPQSGGIFRGPPSTPNAPRG	60
Submitted_Seq	1	MQQAIPTISTQSDIAKIMQPHSAQNM <mark>YVSTGNNSSGNSRSNPQSGAIFRGPPPTANTPRG</mark>	60
eIF4G-PA	61	ASGGATRHVHVQPMYSQPLHQNMVLQQYTQYNPRQQTFPASHLQYAPAPMPYYQYQYVPT	120
Submitted_Seq	61	ATSGATRHVHVQPMYSQPLHQNMVIQQYTQYNPRQQTFPTSHLQYAPAPMPYYQYQYVPT	120
eIF4G-PA	121	LQQQPP-HTRSAVTVNTNVNVGNNLQPVQSGPNGPLPVPGASSSQIQLLTSTVQPGASTV	179
Submitted_Seq	121	IQQQPPPHTRSAVSVSTNVNVGNTLQPVQSGPNGPLTAPGSTSSQLQLITSTVQPGTNNV	180
eIF4G-PA	180	MGVGGPGSTMGQVGVPPMVGVGVMTTSVQPQPVQVQPASRRHQHRLQIIDPTTKKNILD	239
Submitted_Seq	181	MGVGGPGSGMGQ	192
eIF4G-PA	240	DFDKTKSNTDNEFSDQVTSTNTPATVLSEGPIRIPQQESVGLNNLTSTSSQGSESRTNAP	299

eIF4G-PA	240 DFDKTKSNTDNEFSDQVTSTNTPATVLSEGPIRIPQQESVGLNNLTSTSSQGSESRTNAP	299
Submitted_Seq	193TNSNTDNDFSEQVTLPNTPTVVLSEGQIRIPQQDTVGINNLSNTTSQGSETRTNAS	248
eIF4G-PA	300 YIPIETTPISRTDVGPTPIVSAMTDAPSVEILPTPQRGRSKKIPIVSPKNVSESIVAPIT	359
Submitted_Seq	249 YTPVE <mark>PIPISRQDVGQTPIVSAMSDAPSVEILPTPQRGRSKK</mark> IPIVSPKNASEASAAPTT	308
eIF4G-PA	360 DETDDAGSTPISRATEESMPPNQTHPNLLISDSSKHKQAVSNSEISKDAPTGLKEMHVAE	419
Submitted_Seq	309 DETDDALSKPIVTTAKAPTEQSLAHQKLLTSESPQQKQSVSNTEITKDEPTKLEDIKIDE	368
eIF4G-PA	420 LESSVASENHPAGILDVNVNKSSQSLDFSADPEIDSIGAISPTSPNAVSPILHEVLDNTE	479
Submitted_Seq	369 LDSVVSSGNLQTELLSFNVKDSQPPSNFSEEPETASTVEIPPLDFIEDSSKMHTALDNSE	428
eIF4G-PA	480 LSKKLENSTTERFKDSQSVEKPTHQELSLRNATDETEISAMALQDVNSLDNNQIEKT	536
Submitted_Seq	429 STLSIEILEKSTVESFKDNQSAEQQTQQDINLRSVPDETEISSMALKEVTTLDNRQTENK	488
eIF4G-PA	537 YPSKPKLNVDVSEDISSRESAIKSTSTKNTGVDVGLQSDSKPETTLNDKQDSTDLK	592
Submitted_Seq	489 DTIKSKNNADISKELT-RETTMDSLLKNNTDEVVEHQSGTSTDSKPEEDLEDRLQSTDQK	547
eIF4G-PA	593 VK-VSAKISSIINYNEGQWSPNNPSGKKQYDREQLLQLREVKASRIQPEVKNVSILPQPN	651
Submitted_Seq	548 LEGTGITVSSFINYNEGQWSPSNPTGKKQYNREQLLQLREVKASRKQPEVKNISILPQPN	607

Submitted_Seq	548	LEGTGITVSSFINYNEGQWSPSNPTGKKQYNREQLLQLREVKASRKQPEVKNISILPQPN	607
eIF4G-PA	652	LMPSFIRNNNNKRVQSMVGIIGNRSNESAGNYIGKQMSMSGVQSGGGRSSMKGMIHVNL	711
Submitted_Seq	608	LMPSFIRNNNN-KRVQSMVGIIGNRSSESGGNYIGKQISMSGVQGGGGRSSMKGMIHVNL	666
eIF4G-PA	712	SLNQDVKLSENENAWRPRVLNKSDGDSDAKSALEKDELVRRVRGILNKLTPERFDTLVEE	771
Submitted_Seq	667	SLNQDVKLSENENAWRPRGLNKSDGDSEAKSTHEKDELIRRVRGILNKLTPERFDTLVEE	726
eIF4G-PA	772	IIKLKIDTPDKVDEVIVLVFEKAIDEPNFSVSYARLCQRLAAEVKVIDERMESETKSNSA	831
Submitted_Seq	727	IIKLKIDTPEKMDEVIVLVFEKAIDEPNFSVSYARLCHRLISEVKGRDERMESGTKSNLA	786
eIF4G-PA	832	HFRNALLDKTEQEFTQNVSQSTAKEKKLQPIVDKIKKCTDANEKAELEAFLEEEERKIRR	891
Submitted_Seq	787	HFRNALLDKTEREFTQNVSQSTAKEKKLQPIVDKIKKCKDANEKAELEAFLEEEERKIRR	846
eIF4G-PA	892	RSGGTVRFIGELFKISMLTGKIIYSCIDTLLNPHSEDMLECLCKLLTTVGAKFEKTPVNS	951
Submitted_Seq	847	RSGGTVRFIGELFKISMLTGKIIYSCIDTLLNPHSEDMLECLCKLLTTVGAKFEQTPVNS	906
eIF4G-PA	952	KDPSRCYSLEKSITRMQAIASKTDKDGARVSSRVRFMLQDVIDLRKNKWQTSRNEAPKTM	1011
Submitted_Seq	907	KDPSRCYSLEKSITKMQAIASKTDKDGAKVSSRVRFMLQDVIDLRKNKWQTSRNEAPKTM	966
eIF4G-PA	1012	GQIEKEAKNEQLSAQYFGTLSSTTPGGSQGGSGKRDDRGNSRYGESRSSSAYGGSHSQRG	1071

	RoF(6)		2022
eIF4G-PA	1012	GQIEKEAKNEQLSAQYFGTLSSTTPGGSQGGSGKRDDRGNSRYGESRSSSAYGGSHSQRG	1071
Submitted_Seq	967	GQIEKEAKNEQLSAQYFGTLSSTTPVGSQGGSGKRDDRGNTRYGDSRSGSGYGGSHSQRS	1026
eIF4G-PA	1072	DNGNLRHQQQNNVGGNVSGGAGHSNGNNDENTWHVQTSKGSRSLAVDSNKLEGL <mark>SKLSD</mark> Q	1131
Submitted_Seq	1027	DNGNLRHQQQNNTGGAGHSNGNNDDNTWHVQTSKGSRSQAVDSNKLEGL <mark>SKLSDQ</mark>	1081
eIF4G-PA	1132	NLETKKMGGLTQFIWISSDTTRLSSAPTPTPSNPFAVLSSLIDKNSNERDRDRSGPRN	1189
Submitted_Seq	1082	NLETKKMGGLGQFLW-PSNTTRQSSASTSTPSNPFAVLSSLIDKNGSDRDRDRDRSGPRN	1140
eIF4G-PA	1190	KGSYNKGSMERDRYDRGMHSRTGSSQGSRENSSSRGGQQGRTLLSSSVQKSTSHSKYT	1247
Submitted_Seq	1141	KGSYNKGSIERDRFDRGIHSRTGSSQGSRENSSSRAGQHGQGRSLLSSTVQKSTSHSKYT	1200
eIF4G-PA	1248	QQAPPTRHTVKAQSSVGSSNVNTGPLYRGSEQQTSATFSQTTRSVAPVAVFIEASETD	1305
Submitted_Seq	1201	QQAPPTRHAGKTPTSLVSSNVNTGGLYRGSEQQSPTSATFSQGSRSVAPVAVFKEAGETE	1260
eIF4G-PA	1306	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR	1365
Submitted_Seq	1261	LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR	1320
eIF4G-PA	1366	YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAGPLIVKKILTI	1425
Submitted Sea	1321	YLATAVFOLTOONYTSVDHERLAYNEESEYANDLTVDTPELWLYTLOFAGPLTVKKTLTT	1380
			1000
			1000
eIF4G-PA	1306	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR	1365
eIF4G-PA Submitted_Seq	1306 1261	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR ***********************************	1365
eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR	1365 1320 1425
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq	1306 1261 1366 1321	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark>	1365 1320 1425 1380
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366 1321 1426	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAGPLIVKKILTI YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAGPLIVKKILTI SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKENLKWSDFMPESEVADF	1365 1320 1425 1380 1485
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq	1306 1261 1366 1321 1426 1381	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAGPLIVKKILTI YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAGPLIVKKILTI SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKFNLKWSDFMPESEVADF	1365 1320 1425 1380 1485 1440
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366 1321 1426 1381 1486	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> SDLWNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKENLKWSDFMPESEVADF SDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWSKFNLKWSDFMPESEVSDF	1365 1320 1425 1380 1485 1440 1545
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq	1306 1261 1366 1321 1426 1381 1486 1441	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKENLKWSDFMPESEVADF SDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWSKENLKWSDFMPESEVADF IKENRLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEGTTADCIIDYSNGNIMVVD	1365 1320 1425 1380 1485 1440 1545
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366 1321 1426 1381 1486 1441 1546	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAGPLIVKKILTT YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAGPLIVKKILTT SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKFNLKWSDFMPESEVADF SJVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWSKFNLKWSDFMPESEVSDF IKENRLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEGTTADCIIDYSNGNIMVVD KLFIRGLTETLSNFSIHYKDNSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTLQLIV	1365 1320 1425 1380 1485 1440 1545 1500
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq	1306 1261 1366 1321 1426 1381 1486 1441 1546 1501	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAGPLIVKKILTI YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAGPLIVKKILTI SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKFNLKWSDFMPESEVADF SDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWSKFNLKWSDFMPESEVSDF IKFNRLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEGTTADCIIDYSNGNIMVVD KLFIRGLTETLSNFSIHYKDNSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTMQLLY	1365 1320 1425 1380 1485 1440 1545 1500 1605
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366 1321 1426 1381 1486 1441 1546 1501 1606	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFIYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI</mark> SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKFNLKWSDFMPESEVADF SDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWSKFNLKWSDFMPESEVADF KLFIRRLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEGTTADCIIDYSNGNIMVVD KLFIRGLTETLSNFSIHYKDNSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTLQLLV KLFIRGLTETLSNFSIQYKENSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTMQLLV	1365 1320 1425 1380 1485 1440 1545 1500 1605 1560
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366 1321 1426 1381 1486 1441 1546 1501 1606 1561	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAG <mark>PLIVKKILTI YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAGPLIVKKILTI SDLWNNNLKENSPSNVAKKELKTYLIYCTQEVGPNFARNMWIKFNLKWSDFMPESEVADF SDVWNKNLKDNSPSSVAKKELKTYLIYCTQDVGPKFARSMWSKFNLKWSDFMPESEVADF KLFNRLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEGTTADCIIDYSNGNIMVVD KLFIRGLTETLSNFSIHYKDNSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTLQLLV KLFIRGLTETLSNFSIQYKENSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTLQLLV HGLEHPRGLLSELIGELYDAFVIQKESLCKWRDSKDQSAGKGVAVKSLNPFFNSLLNDDA</mark>	1365 1320 1425 1380 1485 1440 1545 1500 1605 1560 1665
eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA Submitted_Seq eIF4G-PA	1306 1261 1366 1321 1426 1381 1486 1441 1546 1501 1606 1561 1666	LKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIYYILTDYLHLANVGKQYRR LKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLYYLLTDYLHLANVGKQYRR YLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELWLYILQFAGPLIVKKILTI YLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELWLYILQFAGPLIVKKILTI SDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWIKFNLKWSDFMPESEVADF SDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWSKENLKWSDFMPESEVSDF IKFNRLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEGTTADCIIDYSNGNIMVVD KLFIRGLTETLSNFSIHYKDNSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTLQLLV KLFIRGLTETLSNFSIQYKENSYKLESETFQKFCIPVLQRYIDSNEDHQLECLYTLQLLV HGLEHPRGLLSELIGELYDAFVIQKESLCKWRDSKDQSAGKGVAVKSLNPFFNSLLNDDA HGLEHPRGLLSELIGELYDAFVIQKESLCKWRDSKDQSAGKGVAVKSLNPFFNSLLNDDA	1365 1320 1425 1380 1485 1440 1545 1500 1605 1560 1665

# Interfactor of the second seco

eIF4G-PA

Dot plot of eIF4G-PA vs. Submitted\_Sequence

#### 4. Dot plot between the submitted model and the D. melanogaster ortholog

The gap seen in the dot plot is due to a deletion of 90 amino acids. A splice site was found at the location predicted by sequence similarity to the *D. melanogaster* ortholog, but this could not be accepted due to the presence of an in-frame stop codon. A secondary splice site was found that has RNA-seq data, TopHat, and Gene Predictors to support it. As seen in the following screenshot, this deletion is in a region of low sequence conservation among *Drosophila* species closely related to *D. biarmipes* and *D. melanogaster* (the highlighted region being the region of deletion in *D.* biarmipes). Good sequence conservation does not begin until further along the sequence.

Dbia Dmel	LQYAPAPMPYYQYQYVPTIQQQPPPHTRSAVSVSTNVNVGNTLQPVQSGPNGPLTAPGST LQYAPAPMPYYQYQYVPTLQQQ-PPHTRSAVTVNTNVNVGNNLQPVQSGPNGPLPVPGAS
Dsec	
Dyak	
Dere	
Dbia	SSQLQLITSTVQPGTNNVMGVGGPGSGMGQT
Dmel	SSQIQLLTSTVQPGASTVMGVGGPGSTMGQVGVPPMVGVGVMTTSVQPQPVQVQPASRRR
Dsec	
Dyak	
Dere	
Dbia	NSNTDNDFSEOVTLENTPTVVLSEGOIRIPOODTVGI
Dme 1	HOHRLOTTDETTKKNTLDDEDKTKSNTDNEESDOVTSTNTEATVLSEGETRIEOOESVGL
Dsec	
Dvak	
Dere	
Dbia Dmel	NNLSNTTSQGSETRTNASYTPVEPIPISRQDVGQTPIVSAMSDAPSVEILPTPQRGRSKK NNLTSTSSQGSESRTNAPYIPIETTPISRTDVGPTPIVSAMTDAPSVEILPTPQRGRSKK
Dsec	MTDAPSVEILPTPQRGRSKK
Dyak	MSDAPSVEILPTPQRGRSKK
Dere	MSDAPSVEILPTPQRGRSKK *:**********************************
Dbia	IPIVSPKNASEASAAPTTDETDDALSKPIVTTAKAPTEQSLAHQKLLTSESPQQKQSVSN
Dmel	IPIVSPKNVSESIVAPITDETDDAGSTPISRATEESMPPNQTHPNLLISDSSKHKQAVSN
Dsec	IPIVSPKNVSESIAAPITDERDDAGSTPISATTKEPI-PNQPHQNLLISDSSKHNQSVSN
Dyak	IPIMSPKNVSESIAAPITDETDDAGSTPISGATKETVQPNFPHQNLHISESPKQKQTVSN
Dere	IPIVPPKNVSEPIDAPITDETDDAGAIPISGATKEPVQPNLPHQNLHISESPKQNQTVSN ***: ***.** ** *** *** : ** ::: . * :* *:* :::***
Dbia	TEITKDEPTKLEDIKIDELDSVVSSGNLQTELLSFNVKDSQ-PPSNFSEEPETASTVEIP
Dmel	SEISKDAPTGLKEMHVAELESSVASENHPAGILDVNVN-KSSQSLDFSADPEIDSIGAIS
Dsec	TEILTDVPTVLKEMHIAELESSVASENHSAGILDVNVN-KSSQSSDFPADPEFGGIGTTS
Dyak	IEITTDVPTVLKETNIAELESSVSSENLLTGFQAVNANVKSPQSSNFSADPENASIDKSS
Dere	TEITTEVPTVLKETNIAELKSSVSSENVAVNVNKSPQPSDFSADPEIASIDKSS
	** .: ** *:: .: **.* *:* ** :** .

#### **ISOFORM REPORT FORM**

Gene-isoform name: eIF4G-PB Names of the isoforms with identical coding sequences as this isoform: *None* Is the 5' end of this isoform missing from the end of the project: No Is the 3' end of this isoform missing from the end of the project: No

## **1. Gene Model Checker Checklist**

Configure Const Madel				-	1.1		-	ll a ul a			
Configure Gene Piodel			<<	G	hecklis	t Dot Plot	Transcript Sequence	Peptide S	equence	Extracted Coding Exon	s
Model Details			-	t:	Expand	All 📘 Collaps	e All				
Fosmid Sequence File:	C:\fakepath\cont	ig50.fasta Browse			Vi	Criteria		Status	Message		
Ortholog in D. melanogaster:	eIF4G-PB			Ð	Q	Check for Start	Codon	Pass 🖉			-
Coding Dyon Coordinatory	20442 20266 26	077 26212 25221		Ð		Acceptor for CD	)S 1	Skip	Already of	checked for Start Codon	
Coung Exon Coordinates:	34834,34620-34	440,33735-33624,29181-		Ŧ	9	Donor for CDS 1	15 - 14	© Pass			
	29126,24626-22	327,21699-21456,21399-	Ξ	Ð	Q	Acceptor for CD	0S 2	O Pass			11
	17200,17139-16	917,16782-16681,16620-16562		Ð	Q	Donor for CDS 2	2	O Pass			
	Sector Se			Ð	Q	Acceptor for CD	IS 3	Pass			
Annotated Untranslated	Yes	No		Ð	Q	Donor for CDS 3	3	Pass			
Regions?	-			Ð	9	Acceptor for CD	)S 4	© Pass			
Orientation of Gene Relative to	🔘 Plus	Minus		Ð	Q	Donor for CDS 4	ŧ j	© Pass			
Query Sequence.				Ð	Q	Acceptor for CD	)S 5	Pass			
Translation:	<ul> <li>Complete</li> </ul>	O Partial		Ħ	9	Donor for CDS 5	i / )	© Pass			
Stop Codon Coordinates:	16561-16559		-	Ð	Q	Acceptor for CD	DS 6	Pass			
•		,		Ħ	Q	Donor for CDS 6	3	O Pass			
		Varify Case Medal Baset Form		€	Q	Acceptor for CD	)S 7	O Pass			
		verity Gene model Reset Form		Ð	Q	Donor for CDS 7	19 ( )	O Pass			-

#### Gene Model Checker

Configure Gene Model			~	C	heckli	t Dot Plot	Transcript Sequence	e Peptide	Sequence	Extracted Coding Exon	ıs
	21236,21078-206	04,20396-20173,17346-		1	Expan	d All 📔 E Collap	se All				
	17200,17139-169	17,16782-16681,16620-16562			Vi	Criteria		Status	Message	e	
				H	R	Acceptor for C	DS 10	© Pass			*
Annotated Untranslated	🔘 Yes	No		Ð	Q	Donor for CDS	10	O Pass			
		~		Ŧ	9	Acceptor for C	DS 11	Pass			
Query Sequence:	O Plus	Minus		Ŧ	Q	Donor for CDS	11	Pass			
Completeness of Gene Model	Complete	Partial		Ð	9	Acceptor for C	DS 12	O Pass			
Translation:	Complete	<b>O</b> Furdar		Ħ	Q	Donor for CDS	12	O Pass			
Stop Codon Coordinates:	16561-16559			Ð	Q	Acceptor for C	DS 13	@ Pass			
			-	Ð	Q	Donor for CDS	13	O Pass			
Project Details			-	Ð	Q	Acceptor for C	DS 14	O Pass			-
Project Group:	D. biarmipes Dot	v		Ð	Q	Donor for CDS	14	O Pass			
Project Name:	contia50			Ð	Q	Acceptor for C	DS 15	O Pass			
				Ð		Donor for CDS	15	Skip	Already	checked for Stop Codon	=
•	m		н т. Н	Ð	Q	Check for Stop	Codon	O Pass			
		1		Ð	Q	Additional Chee	cks	O Pass			
		Verify Gene Model Reset Forr	n	Ŧ	Q	Number of cod	ing exons matched	O Pass			+





# 3. Alignment between submitted model and D. melanogaster ortholog

# Alignment of eIF4G-PB vs. Submitted\_Seq

#### View plain text version

Identity: 1485/1939 (76.6%), Similarity: 1637/1939 (84.4%), Gaps: 83/1939 (4.3%)

eIF4G-PB	1	MQQAIPTLPTQSDIDKAMQPHSAQNM <mark>ILPANKKTKKYDQQVPTSKPQSLHQPLQPQHSH</mark>	P 60
Submitted_Seq	1	MQQAIPTISTQSDIAKIMQPHSAQNM <mark>ILPANKKTKKYAQQVLPSKPQSL-QTMQLQHN</mark> H	Q 59
eIF4G-PB	61	TAQPQFQINKAYNVVSILKASAQIAQQSPHLTNQQHPPIHHPQQTQQHQQSYTNVVNRS	120
Submitted_Seq	60	TPQPQFQINKAYNVVSILKATAQNAQQSSHLTHQQQPSLLL-QQTQQHQQSYANVVNRS	I 118
eIF4G-PB	121	SASEPVRA-QSSVICNGSSILTVNSRQLNSGDMNSTAIYNISSYRKLTGSLDGNVCFLN	179
Submitted_Seq	119	PGSGPVGAHQSTVICNGSNIMTVNSCQLNSGDVNSTAIYNLSNQRGLPGSQDGNVRFLN	V 178
eIF4G-PB	180	QDIKQNGN-ISGSVVSNKSIVGVGSEKSSCTGVSINNQIVLPNAQIGTSMGIA-GTT	235
Submitted_Seq	179	PDTTKKGNNLGASVVSSNSTTGVGNGTTSCTGVGITNNSQITLTSSHIGTTMGVALGTT	: A 238
eIF4G-PB	236	AGTSYMHEKNIVGVSVNCVNTSKKYDFNNSSLLSNNSYPASTAE <mark>YVSTGNNNSGNTRSN</mark>	P 295
Submitted Sea	239	AGTTYMHERNIVGVSVSCVDISRKYDERNSSLIKNNSFOAA-AFWUSTGNNSSGNSRSN	P 297
eIF4G-PB	236	AGTSYMHEKNIVGVSVNCVNTSKKYDFNNSSLLSNNSYPASTAE	295
Submitted_Seq	239	AGTTYMHEKNIVGVSVSCVDTSRKYDFKNSSLIKNNSFQAA-AE <mark>YVSTGNNSSGNSRSNF</mark>	297
eIF4G-PB	296	QSGGIFRGPPSTPNAPRGASGGATRHVHVQPMYSQPLHQNMVLQQYTQYNPRQQTFPASH	355
Submitted_Seq	298	QSGAIFRGPPPTANTPRGATSGATRHVHVQPMYSQPLHQNMVIQQYTQYNPRQQTFPTSH	357
eIF4G-PB	356	LQYAPAPMPYYQYQYVPTLQQQPP-HTRSAVTVNTNVNVGNNLQPVQSGPNGPLPVPGAS	414
Submitted_Seq	358	LQYAPAPMPYYQYQYVPTIQQQPPPHTRSAVSVSTNVNVGNTLQPVQSGPNGPLTAPGST	417
eIF4G-PB	415	SSQIQLLTSTVQPGASTVMGVGGPGSTMGQVGVPPMVGVGVMTTSVQPQPVQVQPASRRR	474
Submitted_Seq	418	SSQLQLITSTVQPGTNNVMGVGGPGSGMGQ	447
eIF4G-PB	475	HQHRLQIIDPTTKKNILDDFDKTKSNTDNEFSDQVTSTNTPATVLSEGPIRIPQQESVGL	534
Submitted_Seq	448	TNSNTDNDFSEQVTLPNTPTVVLSEGQIRIPQQDTVGI	485
eIF4G-PB	535	NNLTSTSSQGSESRTNAPYIPIETTPISRTDVGPTPIVSAMTDAPSVEILPTPQRGRSKK	594
Submitted_Seq	486	NNLSNTTSQGSETRTNASYTPVEPIPISRQDVGQTPIVSAMSDAPSVEILPTPQRGRSKK	545
eIF4G-PB	595	IPIVSPKNVSESIVAPITD ETDDAGSTPISRATEESMPPNQTHPNLLISDSSKHKQAVSN	654
Submitted_Seq	546	IPIVSPKNASEASAAPTTDETDDALSKPIVTTAKAPTEQSLAHQKLLTSESPQQKQSVSN	605

	********.**: .** ******* *.** ::: :* :** *:*.::***
Submitted_Seq	546 IPIVSPKNASEASAAPTTDETDDALSKPIVTTAKAPTEQSLAHQKLLTSESPQQKQSVSN 605
eIF4G-PB	655 SEISKDAPTGLKEMHVAELESSVASENHPAGILDVNVNKSSQSLDFSADPEIDSIGAISP 714
Submitted_Seq	606 TEITKDEPTKLEDIKIDELDSVVSSGNLQTELLSFNVKDSQPPSNFSEEPETASTVEIPP 665
eIF4G-PB	715 TSPNAVSPILHEVLDNTELSKKLENSTTERFKDSQSVEKPTHQELSLRNATDETEIS 771
Submitted_Seq	666 LDFIEDSSKMHTALDNSESTLSIEILEKSTVESFKDNQSAEQQTQQDINLRSVPDETEIS 725
eIF4G-PB	772 AMALQDVNSLDNNQIEKTYPSKPKLNVDVSEDISSRESAIKSTSTKNTGVDVGLQS 827
Submitted_Seq	726 SMALKEVTTLDNRQTENKDTIKSKNNADISKELT-RETTMDSLLKNNTDEVVEHQSGTST 784
eIF4G-PB	828 DSKPETTLNDKQDSTDLKVK-VSAKISSIINYNEGQWSPNNPSGKKQYDREQLLQLREVK 886
Submitted_Seq	785 DSKPEEDLEDRLQSTDQKLEGTGITVSSFINYNEGQWSPSNPTGKKQYNREQLLQLREVK 844
eIF4G-PB	887 ASRIQPEVKNVSILPQPNLMPSFIRNNNNKRVQSMVGIIGNRSNESAGNYIGKQMSMSG 946
Submitted_Seq	845 ASRKQPEVKNISILPQPNLMPSFIRNNNN-KRVQSMVGIIGNRSSESGGNYIGKQISMSG 903
eIF4G-PB	947 VQSGGGRSSMKGMIHVNLSLNQDVKLSENENAWRPRVLNKSDGDSDAKSALEKDELVRRV 1006
Submitted_Seq	904 VQGGGGRSSMKGMIHVNLSLNQDVKLSENENAWRPRGLNKSDGDSEAKSTHEKDELIRRV 963
eIF4G-PB	1007 RGILNKLTPERFDTLVEETIKLKIDTPDKVDEVIVLVFEKAIDEPNFSVSYARLCORLAA 1066
papartosca_ped	
eIF4G-PB	1007 RGILNKLTPERFDTLVEEIIKLKIDTPDKVDEVIVLVFEKAIDEPNFSVSYARLCQRLAA 1066
Submitted_Seq	964 RGILNKLTPERFDTLVEEIIKLKIDTPEKMDEVIVLVFEKAIDEPNFSVSYARLCHRLIS 1023
eIF4G-PB	1067 EVKVIDERMESETKSNSAHFRNALLDKTEQEFTQNVSQSTAKEKKLQPIVDKIKKCTDAN 1126
Submitted_Seq	1024 EVKGRDERMESGTKSNLAHFRNALLDKTEREFTQNVSQSTAKEKKLQPIVDKIKKCKDAN 1083
eIF4G-PB	1127 EKAELEAFLEEEERKIRRRSGGTVRFIGELFKISMLTGKIIYSCIDTLLNPHSEDMLECL 1186
Submitted_Seq	1084 EKAELEAFLEEEERKIRRRSGGTVRFIGELFKISMLTGKIIYSCIDTLLNPHSEDMLECL 1143
eIF4G-PB	1187 CKLLTTVGAKFEKTPVNSKDPSRCYSLEKSITRMQAIASKTDKDGARVSSRVRFMLQDVI 1246
Submitted_Seq	1144 CKLLTTVGAKFEQTPVNSKDPSRCYSLEKSITKMQAIASKTDKDGAKVSSRVRFMLQDVI 1203
eIF4G-PB	1247 DLRKNKWQTSRNEAPKTMGQIEKEAKNEQLSAQYFGTLSSTTPGGSQGGSGKRDDRGNSR 1306
Submitted_Seq	1204 DLRKNKWQTSRNEAPKTMGQIEKEAKNEQLSAQYFGTLSSTTPVGSQGGSGKRDDRGNTR 1263
eIF4G-PB	1307 YGESRSSSAYGGSHSQRGDNGNLRHQQQNNVGGNVSGGAGHSNGNNDENTWHVQTSKGSR 1366
Submitted_Seq	1264 YGDSRSGSGYGGSHSQRSDNGNLRHQQQNNTGGAGHSNGNNDDNTWHVQTSKGSR 1318
eIF4G-PB	1367 SLAVDSNKLEGISKLSDQNLETKKMGGLTQFIWISSDTTRLSSAPTPTPSNPFAVLSSLI 1426
Submitted Seq	1319 SQAVDSNKLEGISKLSDQNLETKKMGGIGQFIW-PSNTTRQSSASTSTPSNPFAVLSSLI 1377

eIF4G-PB	1367	SLAVDSNKLEGISKLSDQNLETKKMGGLTQFIWISSDTTRLSSAPTPTPSNPFAVLSSLI	1426
Submitted_Seq	1319	SQAVDSNKLEGI <mark>SKLSDQNLETKKMGGLGQFLW-PSNTTRQSSASTSTPSNPFAVLSSLI</mark>	1377
eIF4G-PB	1427	DKNSNERDRDRSGPRNKGSYNKGSMERDRYDRG <mark>MHSRTGSSQGSRENSSSRGGQQ</mark> G	1482
Submitted_Seq	1378	DKNGSDRDRDRDRSGPRNKGSYNKGSIERDRFDRG <mark>IHSRTGSSQGSRENSSSRAGQHGQ</mark> G	1437
eIF4G-PB	1483	RTLLSSSVQKSTSHSKYTQQAPPTRHTVKAQSSVGSSNVNTGPLYRGSEQQTSATFSQ	1540
Submitted_Seq	1438	RSLLSSTVQKSTSHSKYTQQAPPTRHAGKTPTSLVSSNVNTGGLYRGSEQQSPTSATFSQ	1497
eIF4G-PB	1541	TTRSVAPVAVFIEASETDLKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIY	1600
Submitted_Seq	1498	GSRSVAPVAVFKEAGETELKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLY	1557
eIF4G-PB	1 <mark>6</mark> 01	YILTDYLHLANVGKQYRRYLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELW	1660
Submitted_Seq	1558	YLLTDYLHLANVGKQYRRYLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELW	1617
eIF4G-PB	1661	LYILQFAGPLIVKKILTISDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWI	1720
Submitted_Seq	1 <mark>6</mark> 18	LYILQFAGPLIVKKILTLSDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWS	1677
eIF4G-PB	1721	KFNLKWSDFMPESEVADFIKFN <mark>RIE YVENESKS PVIDHRET PEKHVKNVIDHIEHLLKEG</mark>	1780
Submitted_Seq	1678	KFNLKWSDFMPESEVSDFIKSN <mark>RLEYIENESKSPVIEQRESPEKHVKNVIDHIEHLLKE</mark> G	1737
eIF4G-PB	1541	TTRSVAPVAVFIEASETDLKLIKSVVSEIVDLSAASKEVTPGAVSCIKRVPEKLRCSFIY	1600
Submitted_Seq	1498	GSRSVAPVAVFKEAGETELKLIKSVVSEMIELAAASKAVTPGVVSCMNRVPEDLRCSFLY	1557
eIF4G-PB	1601	YILTDYLHLANVGKQYRRYLSIAVSQLIQQNYISADHLRLAYNEFTVYANDLIVDIPELW	1660
Submitted_Seq	1558	YLLTDYLHLANVGKQYRRYLAIAVFQLIQQNYISVDHFRLAYNEFSEYANDLIVDIPELW	1617
eIF4G-PB	1661	LYILQFAGPLIVKKILTISDLWNNNLKENSPSNVAKKFLKTYLIYCTQEVGPNFARNMWI	1720
Submitted_Seq	1618	LYILQFAGPLIVKKILTLSDVWNKNLKDNSPSSVAKKFLKTYLIYCTQDVGPKFARSMWS	1677
eIF4G-PB	1721	KFNLKWSDFMPESEVADFIKFN <mark>RLEYVENESKSPVIDHRETPEKHVKNVIDHIEHLLKEG</mark>	1780
Submitted_Seq	1678	KFNLKWSDFMPESEVSDFIKSN <mark>RLEYIENESKSPVIEQRESPEKHVKNVIDHIEHLLKE</mark> G	1737
eIF4G-PB	1781	TTADCIIDYSNGNIMVVDKLFIRGLTETLSNFSIHYKDNSYKLESETFQKFCIPVLQRYI	1840
Submitted_Seq	1738	TTADCIIDYSN <mark>GNIILVDKLFIRGLTETLSNFSIQYKENSYKLESETFQKFCIPVLQRYI</mark>	1797
eIF4G-PB	1841	DSNEDHQLECLYTLQLLVHGLEHPRG <mark>LLSELIGELYDAFVIQKESLCKWRDSKDQSAGKG</mark>	1900
Submitted_Seq	1798	DSNEDHQLECLYTMQLLVHGLEHPRG <mark>LLSELIGELYDAYVIQKESLCKWRDSKDQSAGKG</mark>	1857
eIF4G-PB	1901	VAVKSLNPFFNSLLNDDAN 1919	
50.55 S. X S.	St. 5. 8 - 51		

Submitted\_Seq 1858 VAVKSLNPFFNSLLNDEAN 1876

#### Dot plot of eIF4G-PB vs. Submitted\_Sequence 1400-Submitted\_Sequence

# 4. Dot plot between the submitted model and the D. melanogaster ortholog

See section 4 for isoform A for an explanation of the gap seen on the dot plot.

# **GENE REPORT FORM**

Gene name: unc-13 Gene symbol: unc-13 Approximate location in project (from 5' end to 3' end): 10758-5636 Number of isoforms in *D. melanogaster*: 6 Number of isoforms in project: 4

Name of unique isoform based on coding sequence	List of isoforms with identical coding sequences			
unc-13-PA	unc-13-PC unc-13-PD			
unc-13-PF				

#### **ISOFORM REPORT FORM**

Gene-isoform name: unc-13-PA Names of the isoforms with identical coding sequences as this isoform: unc-13-PC, unc-13-PD Is the 5' end of this isoform missing from the end of the project: No Is the 3' end of this isoform missing from the end of the project: Yes

Number of exons missing from the 3' end: 22

## **1. Gene Model Checker Checklist**



#### 2. View the gene model on the Genome Browser



3. Alignment between submitted model and *D. melanogaster* ortholog

# Alignment of unc-13-PA vs. Submitted\_Seq

#### View plain text version

#### Identity: 956/2944 (32.5%), Similarity: 1156/2944 (39.3%), Gaps: 1327/2944 (45.1%)

unc-13-PA	1	MTHYVRHDYFHNTQNGALSSDTSRISYSQISYETQPSREYFSESYALSNQGPEECSRVSH	60
Submitted_Seq	1	MTHYARHDYFHNTQNGALSSDTGRSPYSHIPYRAQNSREYYAEPYDLGNHGLEEYSSECH	60
unc-13-PA	61	LNSDTVLTTVDNSNNSYGYDYLECYGANIQCDPEEDSVDNWNENTSVVADQYGLGHNNLN	120
Submitted_Seq	61	LTSDRVLTTIDKRNNSYEYDYIECYEAQEQRDVESESIDNWNENHSGVGAQYGFEYANQK	120
unc-13-PA	121	CTSSKLLPKLPNIENGRGSSNACAPQMDVKFNTKGMCIKIDHSYGVCMAKAHDFVGRLSP	180
Submitted_Seq	121	CTSAKVLPTLPVNKTGSGPSKPSATQMDIIFKTKGMCIEKDQRFGVCMAKADEYDLRILP	180
unc-13-PA	181	SDYQNILGNNLNGYAGCAYSSTFLDNAMSSAPLRVLPQSPRCSSYLGRNIIGFNADAAQR	240
Submitted_Seq	181	GDYQNIYADNLNGYAGFAYPSTFLNNAVPAAPSRALPQTNRSSFYLGQDIFGLNADEAQR	240
unc-13-PA	241	DGRGFDTDQTDAMGSESSTYEVYEKMQRPYTSMLPLDYSDYQEGCYNTDNLSTYSDT	297
Submitted Sec	241	FFOSKCDFCMDO&VTMDSGSTSYDVLFKMSRPYTSMLPLDYSDYHDSYYNTDNLSTYSDT	300
unc-13-PA	241	DGRGFDTDQTDAMGSESSTYEVYEKMQRPYTSMLPLDYSDYQEGCYNTDNLSTYSDT	297
Submitted_Seq	241	EEQSKCDFGMDQAVTMDSGSTSYDVLEKMSRPYTSMLPLDYSDYHDSYYNTDNLSTYSDT	300
unc-13-PA	298	PPSNNTQLKRQMQRKISLMMAMTTASVIASGEIRVPVHSKQSKKSTEIQTDSIIGNTIST	357
Submitted_Seq	301	PPTTNSQLKLQKQRKLSLMMAMTTASVIASGETRVAVHSKHSKKPTEFQTDSILGNNIST	360
unc-13-PA	358	NAAARDLDRCLATESCEVIVDTRDSGSVTSFPSSAVTAITKTRKLPKVLPTPLCKSSRHP	417
Submitted_Seq	361	NAATKASDRFLETECSGGIVTLSGPGAVTGTPPVALSTIIKTRKLPKVLPAPQFNSSLHL	420
unc-13-PA	418	ITIATDALSSS-YTSDPLPEKSHRPKQLPKLPISLPQSNDRASLNSNWATPPAPDALPFN	476
Submitted_Seq	421	NSSTANALSSPVYSSDTAAEKSHRPKQLPKLPTSLPQIKPYSIHNSNLTTLSAADELPSY	480
unc-13-PA	477	SFDHKSASSPTPTTTITKDTETTSYLVETDFIGARHNALYQYDSKEPNIVFSDKSVEAEH	536
Submitted_Seq	481	SLKSNTASSPLAITETVATTSYLSSTETKTCPSKKPNEVYLTKSIDDGL	529
unc-13-PA	537	SPTWTPLSPIQSKQSPCPPVALPSNIMQNVSLTCHLPEIEATRSDIEREPESSSIEPILE	596
Submitted_Seq	530	TPPWTPSPPTQLKQYFSPSAELPSQIVQKNSPTSHLPDIEKAKSDIKPDLHENLNESILE	589
unc-13-PA	597	IEKLADPYSGPGSALFNISEYLKPYTLNKPILSEEKKNHIANAASTSTTTPL	648
Submitted_Seq	590	CKKSPEPCSEPESALFNISEYPKPYTLDIIPFSGEKENHITNAASTSTTTHSYDDNGELC	649

		:* .:* * * ******** *****: :* **:***:***	
Submitted_Seq	590	CKKSPEPCSEPESALFNISEYPKPYTLDIIPFSGEKENHITNAASTSTTTHSYDDNGELC	649
unc-13-PA	649	NITSDDEFSSYSNKWTSTCVNFQPLDVESSLNISLKVNAGTNQAELLMTPL	699
Submitted_Seq	650	TELIKLQYSEPSGDSHLFPYFNTWTTSGGNYLPFEVGQGANISTQSYTVTTKAESVMSPL	709
unc-13-PA	700	KSSTP-LFISSNGTSDNFNLRKSSPPDSAFTTTVNVNSFETVLV	742
Submitted_Seq	710	TSFIPPLFISSNFNSKNLTEDIVFSTTSSPHDKNTTVSYSSSNNAFTKTVYVNNPEEVPV	769
unc-13-PA	743	SGSQTASPSPSNLKSPPSIAPLLSYSDYMKQFELPELPQPIMDLSENDTATQSDSFNV	800
Submitted_Seq	770	TTSENAIPPSSDLTLRSPSSLVPILSYTDYMKQFELPDLPQPIMELSENIPVTHSDSNSV	829
unc-13-PA	801	-INNTLTNADNLNSYNQMDVESKSSLQLPSYSSESFDPCSVPSFSIKNKEYKIVEKLDSL	859
Submitted_Seq	830	PINDIATNAEFSCPPNELDVTSKCSDPLPSYLGELFDAYNVPSLSIENQESPIKGQIGDL	889
unc-13-PA	860	SNVESVESPKTLVSPVNPLNCSKLLPGTESIVSNDVAFDDTFYDSFNVDIKELTAFVD	917
Submitted Seq	890	*: * . :*: . *: * * :*** :::: : ***** ********	949
unc-13-PA	918	HVAPEDGLYNFPNDKTSVEFSFDKTEDTIDMNQNLSSGECGYYKPSQAQQ	967
Submitted Sea	950	NIEPENGSRNEPSSSVENSTDIDDKPSDEFINEKTVGTRDLNONLSSGOGGYYKPSOA00	1009
unc-13-PA	968	KASVVASAASSVLDGISKGLKGGLDGVFSGVSSTVDVTOSNPSSKRGFSFNLASKTVPSV	1027
	200	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1027
papint occa ped	555	MILLINGRAFTSSSAFUSIDIDDVLSDFLINFVLAGIVDDVCKDDSQCQ11VLSCVCC	1005
unc-13-PA	968	KASVVASAASSVLDGISKGLKGGLDGVFSGVSSTVDVTQSNPSSKRGFSFNLASKIVPSV	1027
Submitted_Seq	1010	KASVVASAATSVLGGISKGLKGGLDGVFSGVSSTVEVSQTTNSTKKGFSFNLASKLVPSV	1069
unc-13-PA	1028	GGLLTSTSSTSIKQTGSETNPTLILISPENVSSRNSNY-IPTTSPSCTQKNGEENLY	1083
Submitted_Seq	1070	GGLLSSSSSNSTKQTQGQQMDPTPTFITTSAENFSSESSTANVTMTSPPRLYKNAEDTLY	1129
unc-13-PA	1084	SATVHNKSTKSNSYYNEVGEISSTLVRNVCDSYDNSYDEMILTNEMVNIGMLDSESEFGL	1143
Submitted_Seq	1130	AATVSNITDVNHFYSEGLIDSNSALEGNISDTYDESYNEMILTDKLVDEQIVGRDSGYGL	1189
unc-13-PA	1144	IENSYSYQVPDNEQIDSVNSYNNKTQNVTNNGIEKANTKNKPVPLHDPPTKKASTV	1199
Submitted_Seq	1190	TENPYSYHVSCVGQDDTTKACSQLHDIPIDTSFVIEQAEIKSGLVHFPESTTKKGGTTSG	1249
unc-13-PA	1200	GMFGSILGKAAAAVQSATQAVNQSASSVASVVAQKPTIVPRTNNVLLLSSVCSPNEIKRN	1259
Submitted_Seq	1250	GMFGSILGKAAAAAVQSATHAVNQGASSVASVVSQKQSVVPTAHNIRDLSPNAIKRD	1305
unc-13-PA	1260	SSSVEFDSEYGYQMPDVESLSSHYANTGGDYDNSNMKIHEFGTYADDRPYADYHT	1314
Submitted Seq	1306	*. ** *: :*.**: : ******* ***.**:*:*:*:*	1365
unc-13-PA	1315	NGNQSQFKEEAVIPGEPEVINTNILPIGPQATGKKLPTVNGKSALLIKQMPTEVYD	1370
		*******:::::.* *:.:**. *:*** * * ********	
Submitted Seg	1366	NGNQSQFQDDTAIFGQSKVIGNGTKILPTVPPAGSTGKKLPTVNGKSGLLIKQMPTEIYD	1425

unc-13-PA	1315 NGNQSQFKEEAVIPGEPEVINTNILPIGPQATGKKLPTVNGKSALLIKQMPTEVYD 1370
Submitted_Seq	1366 NGNQSQFQDDTAIFGQSKVIGNGTKILPTVPPAGSTGKKLPTVNGKSGLLIKQMPTEIYD 1425
unc-13-PA	1371 DESDTDELDVSPSTGKVPSYSIYSEQEDYYMDLQQTTPSIQPNGFYEQVNNGYDYREDYF 1430
Submitted_Seq	1426 DESELDELDGSPLIGKKPSYHIDSEQDDYYLGLQQTTPSNQANGYYEHVNNGYDYREDYF 1485
unc-13-PA	1431 NEEDEYKYLEQQREQEEHNQPKNKKYLKQAKISKIQPPS-LDFIDVGQDDDFIYDNYH 1487
Submitted_Seq	1486 NEEDEYKYLEQQREQEKLHQPKNKKYLKQTKGVLSTTQPQSSLDFIDEGQDDDFMYENYH 1545
unc-13-PA	1488 SEDDSGNYLEGSSSGSVGPIEGSIIKVDSNIEASFASLNKKSDSFTPTNDSLQKHDTVIG
Submitted_Seq	1546 SEEDSGNYLDESSSGSVGPSEGRNLKMDSNGDASLASTSNQMKRDSFTNNSLHKLDTVVG 1605
unc-13-PA	1548 ESTTKLTRLRTEKMCPDVDEEDENLSDHVSDLTDLSKLISQKKKTLLRGETEEVVGGHMQ 1607
Submitted_Seq	1606 ESTSNLTGIIKEKMCSDLDERSEDINDQLSDLTDINKLTLLKKKSLLRGETEEVVGGHMQ 1665
unc-13-PA	1608 VLRQTEITARQRWHWAYNKIIMQLNNGGGPGEVGLRTNGHPGDNPFYSNIDSMPDIRPRR 1667
Submitted_Seq	1666 MIRQPEITARQRWHWAYNKIIMQLN 1690
unc-13-PA	1668 KSIPLVSELTMAATKRNAGLTSAVPRATLNDEDLKMHVYKKALQALIYPISSTTPHNFLL 1727
Submitted_Seq	1691 1690
Submitted_Seq	1691 1690
unc-13-PA	1728 WTATSPTYCYECEGLLWGIARQGVRCTECGVKCHEKCKDLLNADCLQRAAEKSSKHGAED 1787
Submitted_Seq	1691 1690
unc-13-PA	1788 KANSIITAMKDRMKQREREKPEIFELIRMTFGVDPDTHIDSLEQAEHATVEGTSKWSCKL 1847
Submitted_Seq	1691 1690
unc-13-PA	1848 TITVICAQGLIAKDKSGTSDPYVTVQVSKVKKRTRTMPQELNPVWNEKFHFECHNSSDRI 1907
Submitted_Seq	1691 1690
unc-13-PA	1908 KVRVWDEDNDLKSKLRQKLTRESDDFLGQTIIEVRTLSGEMDVWYNLEKRTDKSAVSGAI 1967
Submitted_Seq	1691 1690
unc-13-PA	1968 RLHISVEIKGEEKVAPYHVQYTCLHENLFHYLCEENTGM <mark>VKLPTQKGDDAWKLYFDEIPE</mark> 2027
Submitted_Seq	1691 1690
unc-13-PA	2028 EIVDEFSMRYGIENIYQAMTHFHCLSAKYLCPGVPAVMSTLLANINAYYAHTTASSAVSA 2087
Submitted_Seq	1691 1690
unc-13-PA	2088 SDRFAASNFG <mark>KEKFVKLLDQLHNSLRIDLSMYRNNFPASSPEKLMDLKSTVDLLTSITFF</mark> 2147

unc-13-PA	2088 SDRFAASNFG <mark>KEKFVKLLDQLHNSLRIDLSMYRNNFPASSPEKLMDLKSTVDLLTSITFF</mark> 2147
Submitted_Seq	1691 1690
unc-13-PA	2148 RMK <mark>VQELSSPPRASTVVKDCVKACLRSTYQFLFENCYELYNREFQ</mark> VDPNEAKRAPDDHEP 2207
Submitted_Seq	1691 1690
unc-13-PA	2208 KLDSVDFWHKLIALIVSVIDEDKNSYGTVLNQFPQELNIGQLSASSMWHLFAVDMKYALE 2267
Submitted_Seq	1691 1690
unc-13-PA	2268 EHEQHRLCKSSAYMNLHFRVKWLYSNYVKEVPPYKGAVPDYPAWFEPFVMQWLNENDDVS 2327
Submitted_Seq	1691 1690
unc-13-PA	2328 LEYLHGAFKRDKKDGFQKSSEHALFSNSVVDVFTQLTQCFDVVSKLECPDPEIWKRYMRR 2387
Submitted_Seq	1691 1690
unc-13-PA	2388 FAKTIVKVLIAYADIVKLEFPEHMKDERI <mark>ACILMNNIQQLRVQLEKMFESMGGDKLEEDA</mark> 2447
Submitted_Seq	1691 1690
unc-13-PA	2448 ANILKELQQNLNSALDDLASQFAISLEPRITQSVRELGDMLLSIKGGSGTLAAGNLAAQR 2507
Submitted Seq	1691 1690
unc-13-PA	2508 NAVAVEADEVLRPLMDLLDGSLTLYAQSCEKTVLKRLLKELWKIVMRILEKTIVLPPMTD 2567
Submitted_Seq	1691 1690
unc-13-PA	2568 KTMMFKHLTDNAKNLASNAKIEDMGRLFKSHMAGKQDVKSALSGVMDISKEVEKNLSPKQ 2627
Submitted_Seq	1691 1690
unc-13-PA	2628 CAVLDVALDTIKQYFHAGGNGLKKTFLEKSPELQSLRYALSLYTQMTDTLIKTFISSQVH 2687
Submitted_Seq	1691 1690
unc-13-PA	2688 EVDLENSEESVGEISVQIDLFSHPGTGEHKVNVKVVAANDLKWQIPSGMFRPFVDINLIG 2747
Submitted_Seq	1691 1690
unc-13-PA	2748 PHLQEKKRKFATKSKSNNWSPKYNESFSFTIGNEEQLDFFELHICVKDYCFARDDRLVGV 2807
Submitted_Seq	1691 1690
unc-13-PA	2808 AVIPLKDISEKGSVACWLPLMRRIEMDETGWTILRILSQRNNDEVAKEFVKLKSEIRQEP 2867
Submitted_Seq	1691 1690
unc-13-PA	2868 TMGT 2871
Submitted_Seq	1691 1690

# 4. Dot plot between the submitted model and the *D. melanogaster* ortholog



Dot plot of unc-13-PA vs. Submitted\_Sequence

#### **ISOFORM REPORT FORM**

Gene-isoform name: unc-13-PF Names of the isoforms with identical coding sequences as this isoform: *None* Is the 5' end of this isoform missing from the end of the project: No Is the 3' end of this isoform missing from the end of the project: Yes Number of exons missing from the 3' end: 20

## **1. Gene Model Checker Checklist**

C	onfigure Gene Model				«	C	necklis	t Dot Plot	Transcript Se	quence P	eptide Sequence	Extracted Coding Ex
	Model Details						Expand	All 📘 Collar	ose All			
	Fosmid Sequence File:	C:\fakepath\cont	ig50.fasta	Browse			Vi	Criteria		Status	Message	
	Ortholog in D. melanogaster:	unc-13-PF			Ð	Check for S		itart Codon 🦿 🍘	O Pas	S		
	Coding Exon Coordinates:	10759 10752 10600 10242 0000 0715 0642				Ð		Acceptor for (	CDS 1	Skip	Already che	ecked for Start Codon
	coung Exon coordinates.	5636			Ŧ	Q	Donor for CDS	1	@ Pas	S		
						Ŧ	Q	Acceptor for (	CDS 2	@ Pas	S	
					-	Ð	Q	Donor for CDS	2	@ Pas	s	
					=	Ð	Q	Acceptor for (	CDS 3	@ Pas	s	
	Annotated Untranslated	Yes	No			۲	Q	Donor for CDS	3	O Pas	s	
	Regions?	0.11				€	Q	Acceptor for (	CDS 4	@ Pas	s	
	Orientation of Gene Relative to Ouery Sequence:	🔘 Plus	Minus			Ð	٩	Donor for CDS	4	© Pas	s	
	Completeness of Gene Model	Complete	Dartial			€		Check for Stop	o Codon	Skip	Partial gene	with 3' end missing
	Translation:	O complete	Partia			Ŧ	Q	Additional Che	cks	@ Pas	S	
	Region Missing:	Missing 3' end of	translated region	~								
	Project Details											
			Verify Gene	Model Depet Fr	arm							



#### 2. View the gene model on the Genome Browser

# 3. Alignment between submitted model and D. melanogaster ortholog

# Alignment of unc-13-PF vs. Submitted\_Seq

#### View plain text version

Identity: 893/2786 (32.1%), Similarity: 1078/2786 (38.7%), Gaps: 1276/2786 (45.8%)

unc-13-PF	1 MTHYVRHDYFHNTQNGALSSDTSRISYSQISYETQPSREYFSESYALSNQGPEECSRVSH 60
Submitted_Seq	1 MTHYARHDYFHNTQNGALSSDTGRSPYSHIPYRAQNSREYYAEPYDLGNHGLEEYSSECH 60
unc-13-PF	61 LNSDTVLTTVDNSNNSYGYDYLECYGANIQCDPEEDSVDNWNENTSVVADQYGLGHNNLN 120
Submitted_Seq	61 LTSDRVLTTIDKRNNSYEYDYIECYEAQEQRDVESESIDNWNENHSGVGAQYGFEYANQK 120
unc-13-PF	121 CTSSKLLPKLPNIENGRGSSNACAPQMDVKFNTKDAAQRDGRGFDTDQTDAMGSESS 177
Submitted_Seq	121 CTSAKVLPTLPVNKTGSGPSKPSATQMDIIFKTKD <mark>EAQREEQSKCDFGMDQAVTMDSGST</mark> 180
unc-13-PF	178 TYEVYEKMQRPYTSMLPLDYSDYQEGCYNTDNLSTYSDTPPSNNTQLKRQMQRKISLMMA 237
Submitted_Seq	181 SYDVLEKMSRPYTSMLPLDYSDYHDSYYNTDNLSTYSDTPPTTNSQLKLQKQRKLSLMMA 240
unc-13-PF	238 MTTASVIASGNTISTNAAARDLDRCLATESCEVIVDTRDSGSVTSFPSSAVTAITKTRKL 297
2012/00/2010/00/2012/10/10/2012	**********
Submitted Sea	241 MTTASVIASGINI ISTNAATKASDREIAETECSGELVII ISGRGAVIGIPPVAISTIIKURKI 300
unc-13-PF	238 MTTASVIASGNTISTNAAARDLDRCLATESCEVIVDTRDSGSVTSFPSSAVTAITKTRKL 297
Submitted_Seq	241 MITASVIASCNNISTNAATKASDRFLETECSGGIVILSGPGAVTGTPPVALSTIIKTRKL 300
unc-13-PF	298 PKVLPTPLCKSSRHPITIATDALSSS-YTSDPLPEKSHRPKQLPKLPISLPQSNDRASLN 356
Submitted_Seq	301 PKVLPAPQFNSSLHLNSSTANALSSPVYSSDTAAEKSHRPKQLPKLPTSLPQIKPYSIHN 360
unc-13-PF	357 SNWATPPAPDALPFNSFDHKSASSPTPTTTITKDTETTSYLVETDFIGARHNALYQYDSK 416
Submitted_Seq	361 SNLTTLSAADELPSYSLKSNTASSPLAITETVATTSYLSSTETKTCPSK 409
unc-13-PF	417 EPNIVFSDKSVEAEHSPTWTPLSPIQSKQSPCPPVALPSNIMQNVSLTCHLPEIEATRSD 476
Submitted Seq	:** *: **:: :*.*** .* * ** .*. ***:*:* * *.***:** ::** 410 KPNEVYLTKSIDDGLTPPWTPSPPTQLKQYFSPSAELPSQIVQKNSPTSHLPDIEKAKSD 469
unc-13-PF	477 IEREPESSSIEPILEIEKLADPYSGPGSALFNISEYLKPYTLNKPILSEEKKNHIANAAS 536
	*: : *.*** :* .:* * * ********* *****: :* **:***
Submitted_Seq	470 IKPDLHENLNESILECKKSPEPCSEPESALFNISEYPKPYTLDIIPFSGEKENHITNAAS 529
unc-13-PF	537 TSTTTPLNITSDDEFSSYSNKWTSTCVNFQPLDVESSLNISLK 579
Submitted_Seq	530 TSTTTHSYDDNGELCTELIKLQYSEPSGDSHLFPYFNTWTTSGGNYLPFEVGQGANISTQ 589
unc-13-PF	580 VNAGTNQAELLMTPLKSSTP-LFISSNGTSDNFNLRKSSPPDSA 622
Submitted Seg	590 SYTVTTKAESVMSPLTSFIPPLFISSNFNSKNLTEDIVFSTTSSPHDKNTTVSYSSSNNA 649

	· *.:** :*:**:* * ****** .*.*:*** * .*	
Submitted_Seq	590 SYTVTTKAESVMSPLTSFIPPLFISSNFNSKNLTEDIVFSTTSSPHDKNTTVSYSSSNNA 649	
unc-13-PF	623 FTTTVNVNSFETVLVSGSQTASPSPSNLKSPPSIAPLLSYSDYMKQFELPELPQPIMD 680	
Submitted_Seq	650 FTKTVYVNNPEEVPVTTSENAIPPSSDLTLRSPSSLVPILSYTDYMKQFELPDLPQPIME 709	
unc-13-PF	681 LSENDTATQSDSFNV-INNTLTNADNLNSYNQMDVESKSSLQLPSYSSESFDPCSVPSFS 739	
Submitted_Seq	710 LSENIPVTHSDSNSVPINDIATNAEFSCPPNELDVTSKCSDPLPSYLGELFDAYNVPSLS 769	
unc-13-PF	740 IKNKEYKIVEKLDSLSNVESVESPKTLVSPVNPLNCSKLLPGTESIVSNDVAFDDTFY 797	
Submitted_Seq	770 IENQESPIKGQIGDLFNIAIVPQVPINTIECSVDPPNRFNLLPEADAVENHAVAFDDNFY 829	
unc-13-PF	798 DSFNVDIKELTAFVDHVAPEDGLYNFPNDKTSVEFSFDKTEDTIDMNQNL 847	
Submitted_Seq	830 DSFNVDIKELTASVANIEPENGSRNFPSSSVENSTDIDDKPSDEFINEKTVGTRDLNQNL 889	
unc-13-PF	848 SSGECGYYKPSQAQQKASVVASAASSVLDGISKGLKGGLDGVFSGVSSTVDVTQSNPSSK 907	
Submitted_Seq	890 SSGQGGYYKPSQAQQKASVVASAATSVLGGISKGLKGGLDGVFSGVSSTVEVSQTTNSTK 949	
unc-13-PF	908 RGFSFNLASKIVPSVGGLLTSTSSTSIKQTGSETNPTLILISPENVSSRNSNY-IPT 963	
Submitted Seq	.*************************************	
unc-13-PF	964 TSPSCTQKNGEENLYSATVHNKSTKSNSYYNEVGEISSTLVRNVCDSYDNSYDEMILTNE 1023	
Second Chick state	···· ·· ·· ······· · · · · · · · · · ·	
unc-13-PF	964 TSPSCTQKNGEENLYSATVHNKSTKSNSYYNEVGEISSTLVRNVCDSYDNSYDEMILTNE 1023	
Submitted_Seq	1010 TSPPRLYKNAEDTLYAATVSNITDVNHFYSEGLIDSNSALEGNISDTYDESYNEMILTDK 1069	
unc-13-PF	1024 MVNIGMLDSESEFGLIENSYSYQVPDNEQIDSVNSYNNKTQNVTNNGIEKANTKNKPV 1081	
Submitted_Seq	1070 LVDEQIVGRDSGYGLTENPYSYHVSCVGQDDTTKACSQLHDIPIDTSFVIEQAEIKSGLV 1129	
unc-13-PF	1082 PLHDPPTKKASTVGMFGSILGKAAAAAVQSATQAVNQSASSVASVVAQKPTIVPRTNNV 1139	
Submitted_Seq	1130 HFPESTTKKGGTTSGGMFGSILGKAAAAVQSATHAVNQGASSVASVVSQKQSVVPTAHNI 1189	
unc-13-PF	1140 LLLSSVCSPNEIKRNSSSVEFDSEYGYQMPDVESLSSHYANTGGDYDNSNMKIHE 1194	
Submitted_Seq	1190 RDLSPNAIKRDSNRDSVGFNVMNVDYSYQLGNEESLSSHYENTGDDYENSNIKMHE 1245	
unc-13-PF	1195 FGTYADDRPYADYHTNGNQSQFKEEAVIPGEPEVINTNILPIGPQATGKKLPTVNG 1250	
Submitted_Seq	1246 YGTYTDNKAFVNYHSNGNQSQFQDDTAIFGQSKVIGNGTKILPTVPPAGSTGKKLPTVNG 1305	
unc-13-PF	1251 KSALLIKQMPTEVYDDESDTDELDVSPSTGKVPSYSIYSEQEDYYMDLQQTTPSIQPNGF 1310	
Submitted_Seq	1306 KSGLLIKQMPTEIYDDESELDELDGSPLIGKKPSYHIDSEQDDYYLGLQQTTPSNQANGY 1365	
unc-13-PF	1311 YEQVNNGYDYREDYFNEEDEYKYLEQQREQEEHNQPKNKKYLKQAKISKIQPPS-LDF 1367	
Submitted Seq	1366 YEHVNNGYDYREDYFNEEDEYKYLEQQREQEKLHQPKNKKYLKQTKGVLSTTQPQSSLDF 1425	

unc-13-PF	1311 YEQVNNGYDYREDYFNEEDEYKYLEQQREQEEHNQPKNKKYLKQAKISKIQPPS-LDF 1367
Submitted_Seq	1366 YEHVNNGYDYREDYFNEEDEYKYLEQQREQEKLHQPKNKKYLKQTKGVLSTTQPQSSLDF 1425
unc-13-PF	1368 IDVGQDDDFIYDNYHSEDDSGNYLEGSSSGSVGPIEGSIIKVDSNIEASFASLNKKSD 1425
Submitted_Seq	1426 IDEGQDDDFMYENYHSEEDSGNYLDESSSGSVGPSEGRNLKMDSNGDASLASTSNQMKRD 1485
unc-13-PF	1426 SFTPTNDSLQKHDTVIGESTTKLTRLRTEKMCPDVDEEDENLSDHVSDLTDLSKLISQKK 1485
Submitted_Seq	1486 SFTNNSLHKLDTVVGESTSNLTGIIKEKMCSDLDERSEDINDQLSDLTDINKLTLLKK 1543
unc-13-PF	1486 <mark>KTILLRGETEEVVGGHMQVLRQTEITARQRWHWAYNKIIMQLN</mark> NGGGPGEVGLRTNGHPGD 1545
Submitted_Seq	1544 KSLLRGETEEVVGGHMQMIRQPEITARQRWHWAYNKIIMQLN 1585
unc-13-PF	1546 NPFYSNIDSMPDIRPRRKSIPLVSEL <mark>VLK</mark> ANYPNAPSKMDQCYLLSLQENVETT <mark>IMAATK</mark> 1605
Submitted_Seq	1586 1585
unc-13-PF	1606 RNAGLTSAVPRATINDEDI <mark>KMHVYKKALQALIYPISSTTPHNFLLWTATSPTYCYECEGL</mark> 1665
Submitted_Seq	1586 1585
unc-13-PF	1666 LWGIARQGVRCTECGVKCHEKCKDLLNADCLQRAAEKSSKHGAEDKANSIITAMKDRMKQ 1725
Submitted_Seq	1586 1585

Submitted_Seq	1586 1585
unc-13-PF	1726 REREKPETFELISECHNSSDRIKVRVWDEDNDLKSKLRQKLTRESDDFLGQTIIEVRTLS 1785
Submitted_Seq	1586 1588
unc-13-PF	1786 GEMDVWYNLEKRTDKSAVSGAIRLHISVEIKGEEKVAPYHVQYTCLHENLFHYLCEENTG 1845
Submitted_Seq	1586 1588
unc-13-PF	1846 MVKLPTQKGDDAWKLYFDEIPEEIVDEFSMRYGIENIYQAMTHFHCLSAKYLCPGVPAVM 1905
Submitted_Seq	1586 1588
unc-13-PF	1906 STLLANINAYYAHTTASSAVSASDRFAASNFG <mark>KEKFVKLLDQLHNSLRIDLSMYRNNFPA</mark> 1965
Submitted_Seq	1586 1588
unc-13-PF	1966 SSPEKIMDLKSTVDLLTSITFFRMRVQELSSPPRASTVVKDCVKACLRSTYQFLFENCYE 2025
Submitted_Seq	1586 1588
unc-13-PF	2026 LYNREFQVDPNEAKRAPDDHEPKLDSVDFWHKLIALIVSVIDEDKNSYGTVLNQFPQELN 2085
Submitted_Seq	1586 1585
unc-13-PF	2086 IGQLSASSMWHLFAVDMKYALEEHEQHRLCKSSAYMNLHFRVKWLYSNYVKEVPPYKGAV 2145



# 4. Dot plot between the submitted model and the *D. melanogaster* ortholog



Dot plot of unc-13-PF vs. Submitted\_Sequence

unc-13-PF

# **Project Gene Models**

