GEP Annotation Report for *D. biarmipes* contig34

Erick LeBrun

Student name:	Erick LeBrun
Student email:	lebrun@unm.edu
Faculty advisor:	Dr. Paul Szauter
College/University:	University of New Mexico

Project Details

Project name:	Dbiarmipes_dot_Aug2012_contig34
Project species:	*Drosophila biarmipes*
Date of submission:	07/23/13
Size of project in base pairs:	50,000
Number of genes in project:	5

Does this report cover all genes and all isoforms or is it a partial report?
All five genes are covered and all isoforms are covered by this report. Genes are Pur-alpha, CG1970, Ephrin, CG1909, and onecut.

The region of the project covered by this report:
from base _1_ to base _50,000_

Gene report form

Gene name (i.e. *D. mojavensis eyeless*): __*D. biarmipes Pur-alpha*___________
Gene symbol (i.e. dmoj_ey): ____dbia_Pur-α____
Approximate location in project (from 5' end to 3' end): __230-5377__________
Number of isoforms in *D. melanogaster:* _6_______________
Number of isoforms in this project: ____6____________

Complete the following table for all the isoforms in this project:
If you are annotating untranslated regions then all isoforms are unique (by definition)

Name of unique isoform based on coding sequence	**List of isoforms with identical coding sequences**
Pur-alpha-PA	*Pur-alpha-PC, Pur-alpha-PE*
Pur-alpha-PB	*Pur-alpha-PF*
Pur-alpha-PD	
Pur-alpha-PG	

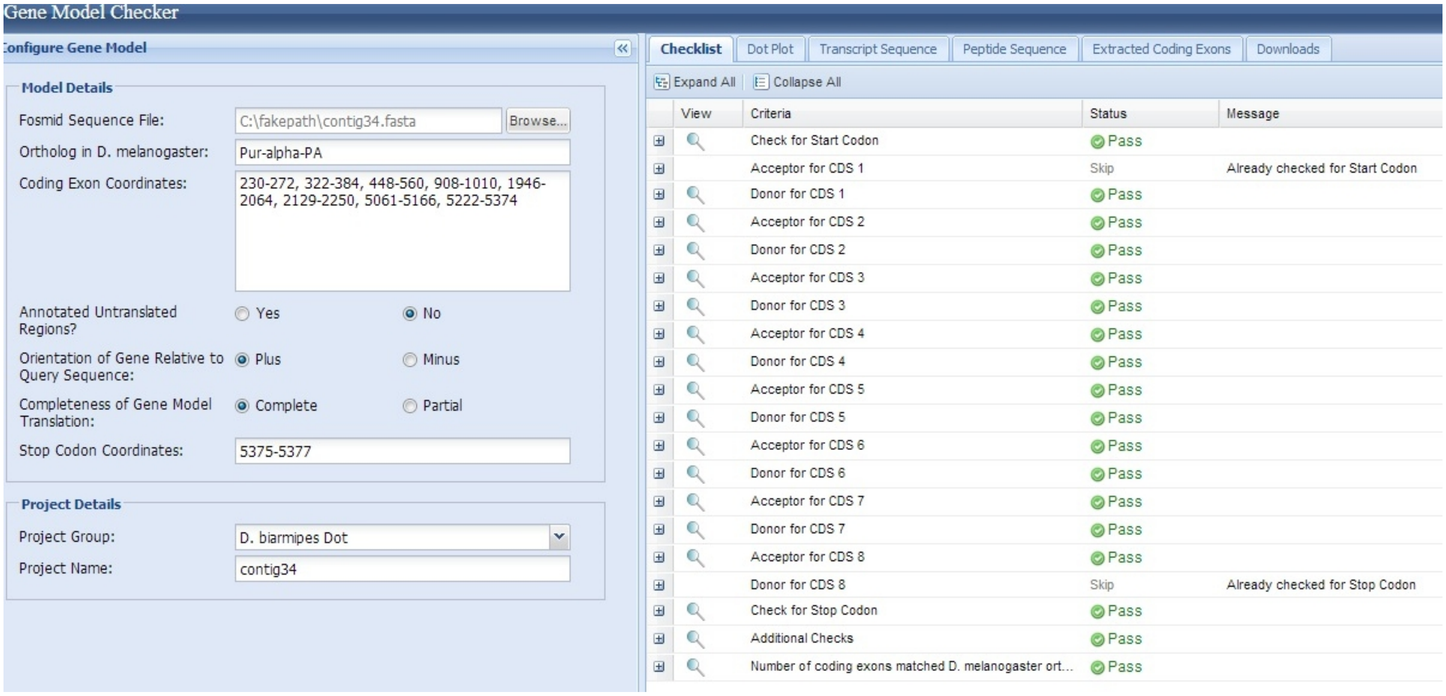
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): ___dbia_*Pur-α-PA*_______
Names of the isoforms with identical coding sequences as this isoform
___dbia *Pur-α-PC*, dbia *Pur-α-PE*_____________
Is the 5' end of this isoform missing from the end of project: ___No____________
If so, how many exons are missing from the 5' end: ____NA____________
Is the 3' end of this isoform missing from the end of the project: __No___________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



Model Details

Fosmid Sequence File: C:\fakepath\contig34.fasta
Ortholog in D. melanogaster: Pur-alpha-PA
Coding Exon Coordinates: 230-272, 322-384, 448-560, 908-1010, 1946-2064, 2129-2250, 5061-5166, 5222-5374
Annotated Untranslated Regions? ○ Yes ● No
Orientation of Gene Relative to Query Sequence: ● Plus ○ Minus
Completeness of Gene Model Translation: ● Complete ○ Partial
Stop Codon Coordinates: 5375-5377

Project Details

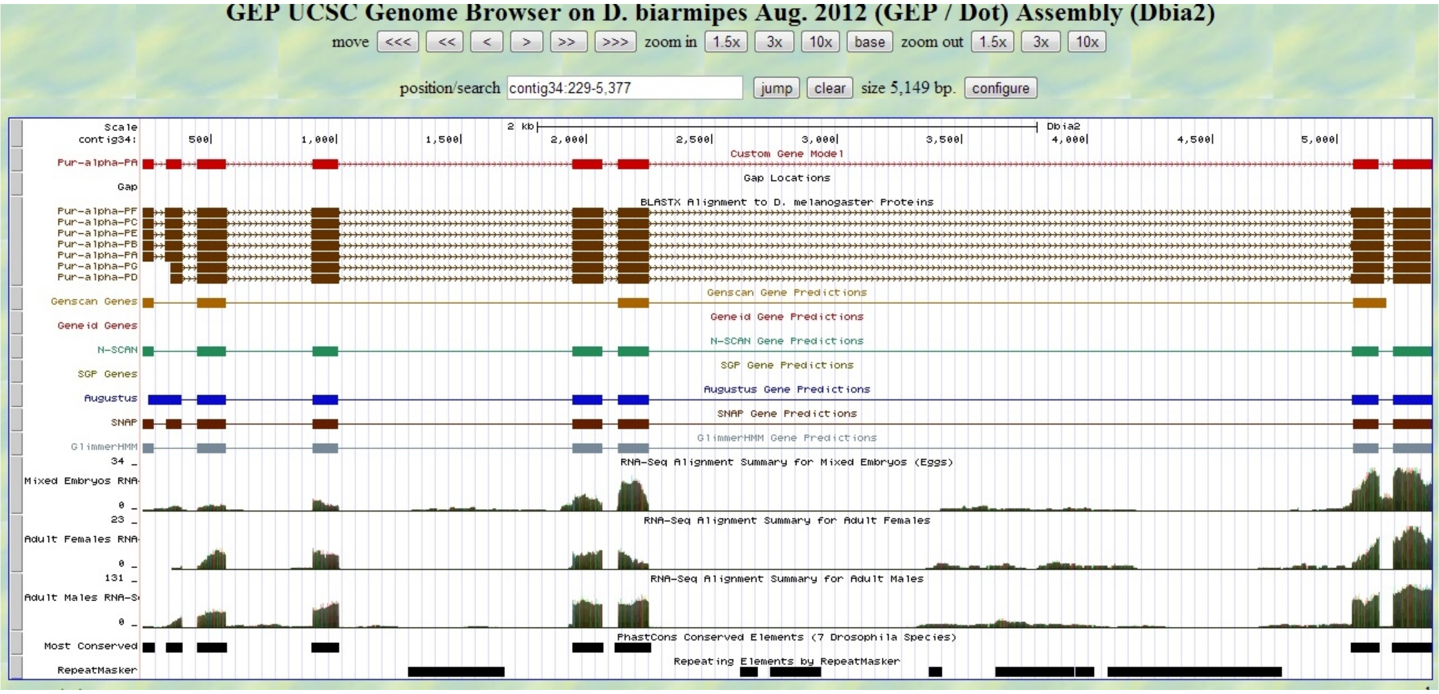
Project Group: D. biarmipes Dot
Project Name: contig34

View	Criteria	Status	Message
	Check for Start Codon	Pass	
	Acceptor for CDS 1	Skip	Already checked for Start Codon
	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	
	Number of coding exons matched D. melanogaster ort...	Pass	

2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker, capture a screenshot of your gene model shown on the Genome

Paste the screenshot of your gene model as shown on the Genome Browser below:



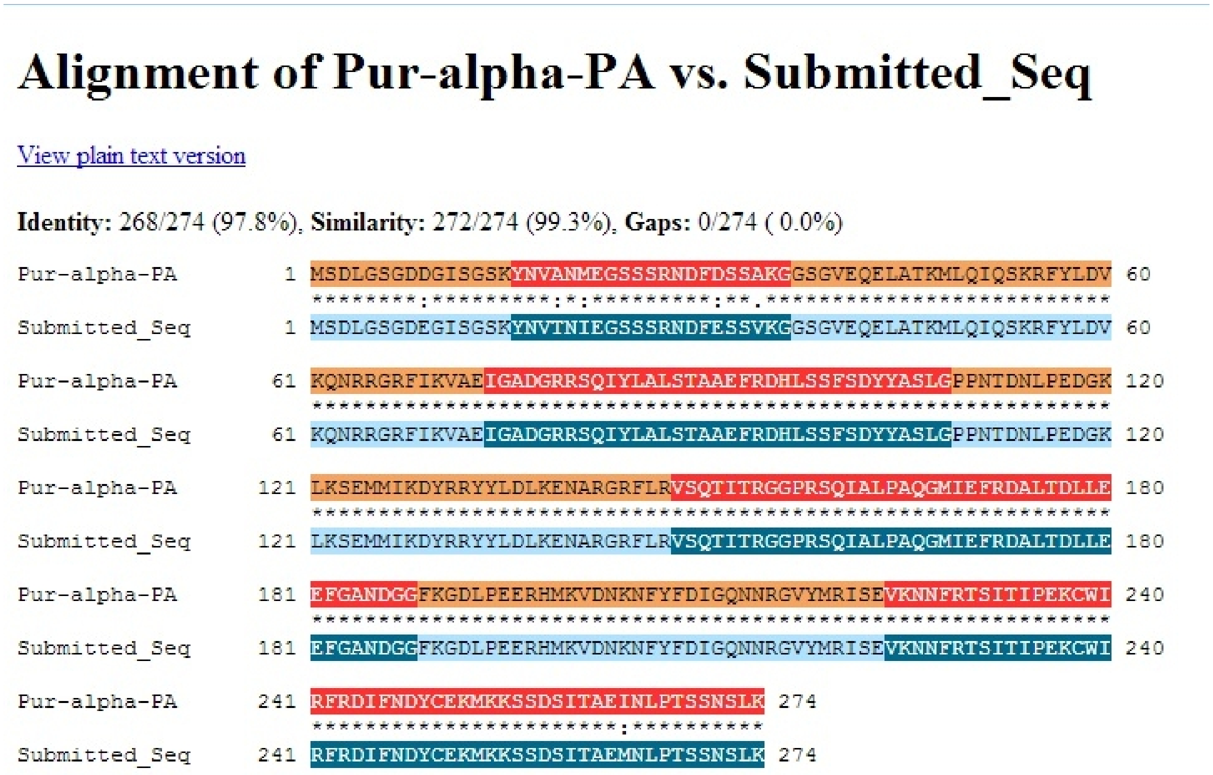
GEP UCSC Genome Browser on D. biarmipes Aug. 2012 (GEP / Dot) Assembly (Dbia2)

position/search contig34:229-5,377 size 5,149 bp.

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

Show an alignment between the protein sequence for your gene model and the protein sequence from the putative *D. melanogaster* ortholog. You can use the protein alignment generated by the Gene Model Checker or you can generate a new alignment using BLAST 2 Sequences (*bl2seq*).

Copy and paste the alignment below:



Alignment of Pur-alpha-PA vs. Submitted_Seq

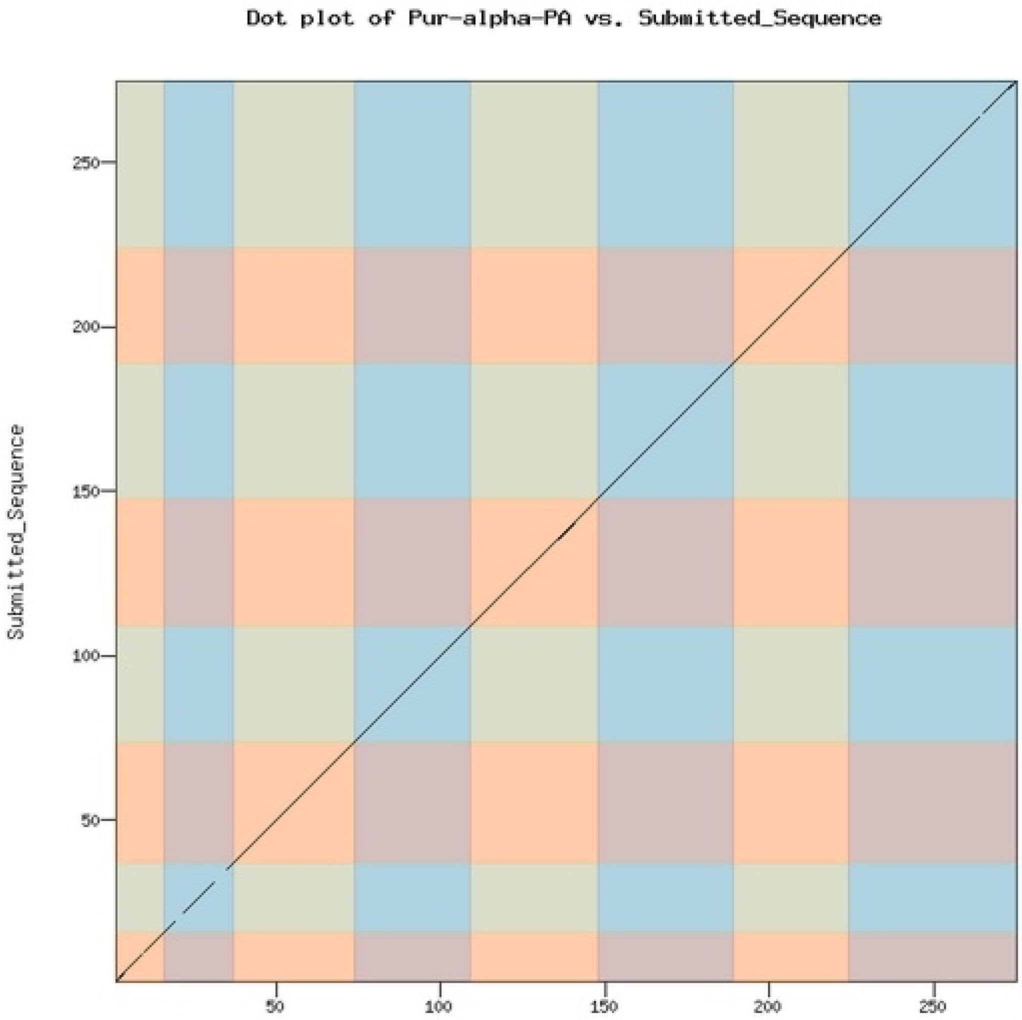
View plain text version

Identity: 268/274 (97.8%), **Similarity:** 272/274 (99.3%), **Gaps:** 0/274 (0.0%)

```
Pur-alpha-PA       1 MSDLGSGDDGISGSKYNVANMEGSSSRNDFDSSAKGGSGVEQELATKMLQIQSKRFYLDV  60
                     ********:*********:*:*********:**.**************************
Submitted_Seq      1 MSDLGSGDEGISGSKYNVTNIEGSSSRNDFESSVKGGSGVEQELATKMLQIQSKRFYLDV  60

Pur-alpha-PA      61 KQNRRGRFIKVAEIGADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGK 120
                     ************************************************************
Submitted_Seq     61 KQNRRGRFIKVAEIGADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGK 120

Pur-alpha-PA     121 LKSEMMIKDYRRYYLDLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLE 180
                     ************************************************************
Submitted_Seq    121 LKSEMMIKDYRRYYLDLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLE 180

Pur-alpha-PA     181 EFGANDGGFKGDLPEERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCWI 240
                     ************************************************************
Submitted_Seq    181 EFGANDGGFKGDLPEERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCWI 240

Pur-alpha-PA     241 RFRDIFNDYCEKMKKSSDSITAEINLPTSSNSLK 274
                     ***********************:**********
Submitted_Seq    241 RFRDIFNDYCEKMKKSSDSITAEMNLPTSSNSLK 274
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).



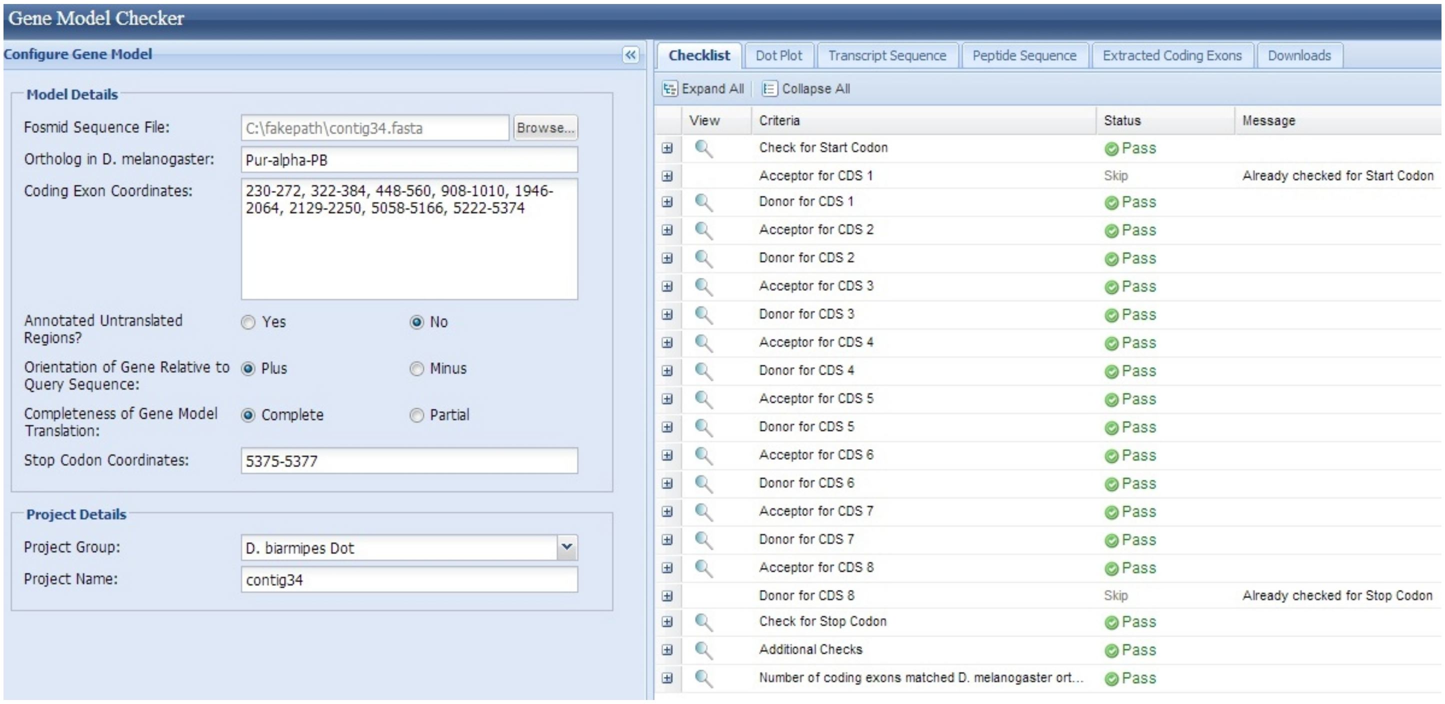
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): _____dbia_*Pur-α-PB*________
Names of the isoforms with identical coding sequences as this isoform
_____dbia_*Pur-α-PF*_____________
Is the 5' end of this isoform missing from the end of project: ___No_____________
If so, how many exons are missing from the 5' end: ____NA_____________
Is the 3' end of this isoform missing from the end of the project: __No____________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



Model Details

Fosmid Sequence File: C:\fakepath\contig34.fasta
Ortholog in D. melanogaster: Pur-alpha-PB
Coding Exon Coordinates: 230-272, 322-384, 448-560, 908-1010, 1946-2064, 2129-2250, 5058-5166, 5222-5374
Annotated Untranslated Regions? ○ Yes ● No
Orientation of Gene Relative to Query Sequence: ● Plus ○ Minus
Completeness of Gene Model Translation: ● Complete ○ Partial
Stop Codon Coordinates: 5375-5377

Project Details

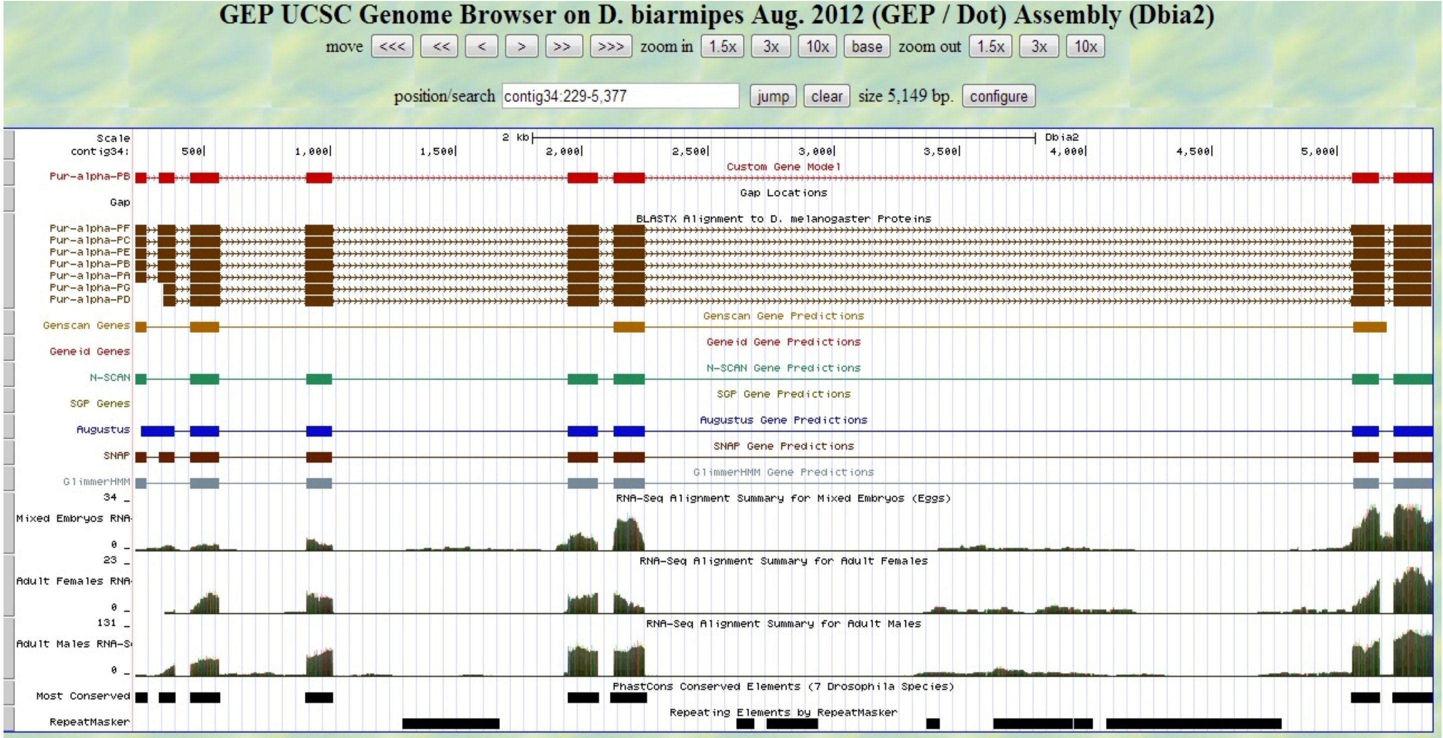
Project Group: D. biarmipes Dot
Project Name: contig34

View	Criteria	Status	Message
🔍	Check for Start Codon	✓ Pass	
	Acceptor for CDS 1	Skip	Already checked for Start Codon
🔍	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	
🔍	Number of coding exons matched D. melanogaster ort...	✓ Pass	

2. View the gene model on the Genome Browser

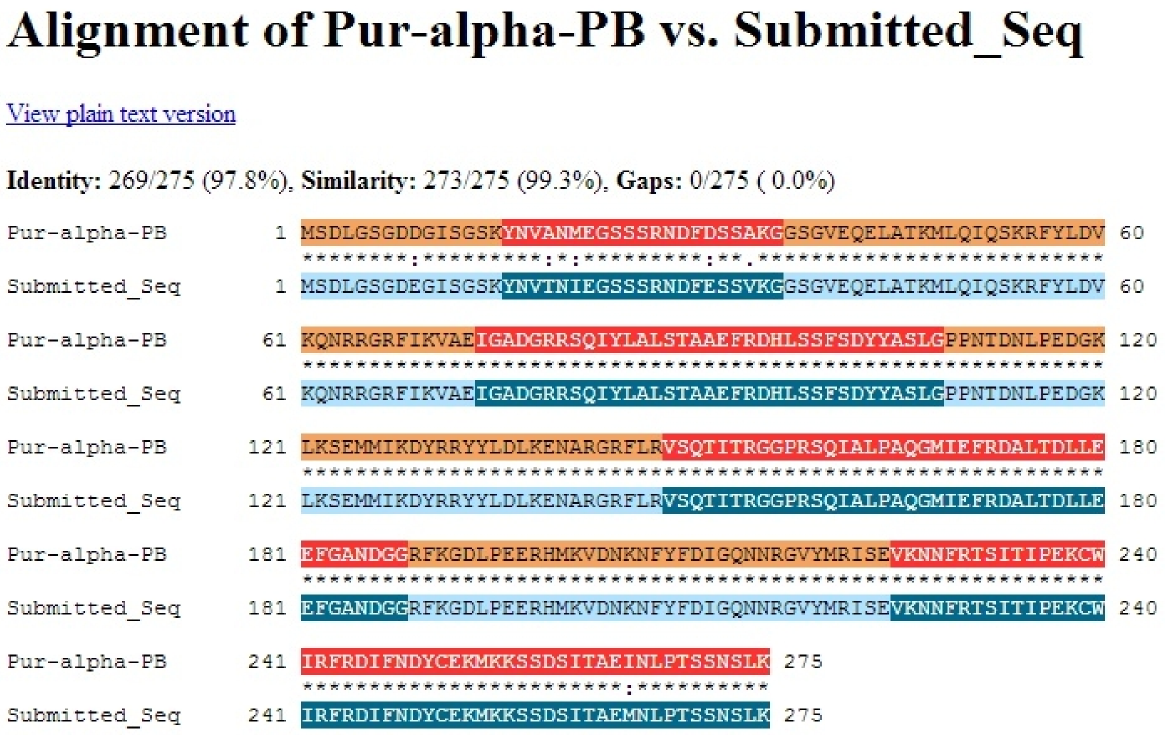
Using the custom track feature from the Gene Model Checker, capture a screenshot of your gene model shown on the Genome

Paste the screenshot of your gene model as shown on the Genome Browser below:



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

Copy and paste the alignment below:



Alignment of Pur-alpha-PB vs. Submitted_Seq

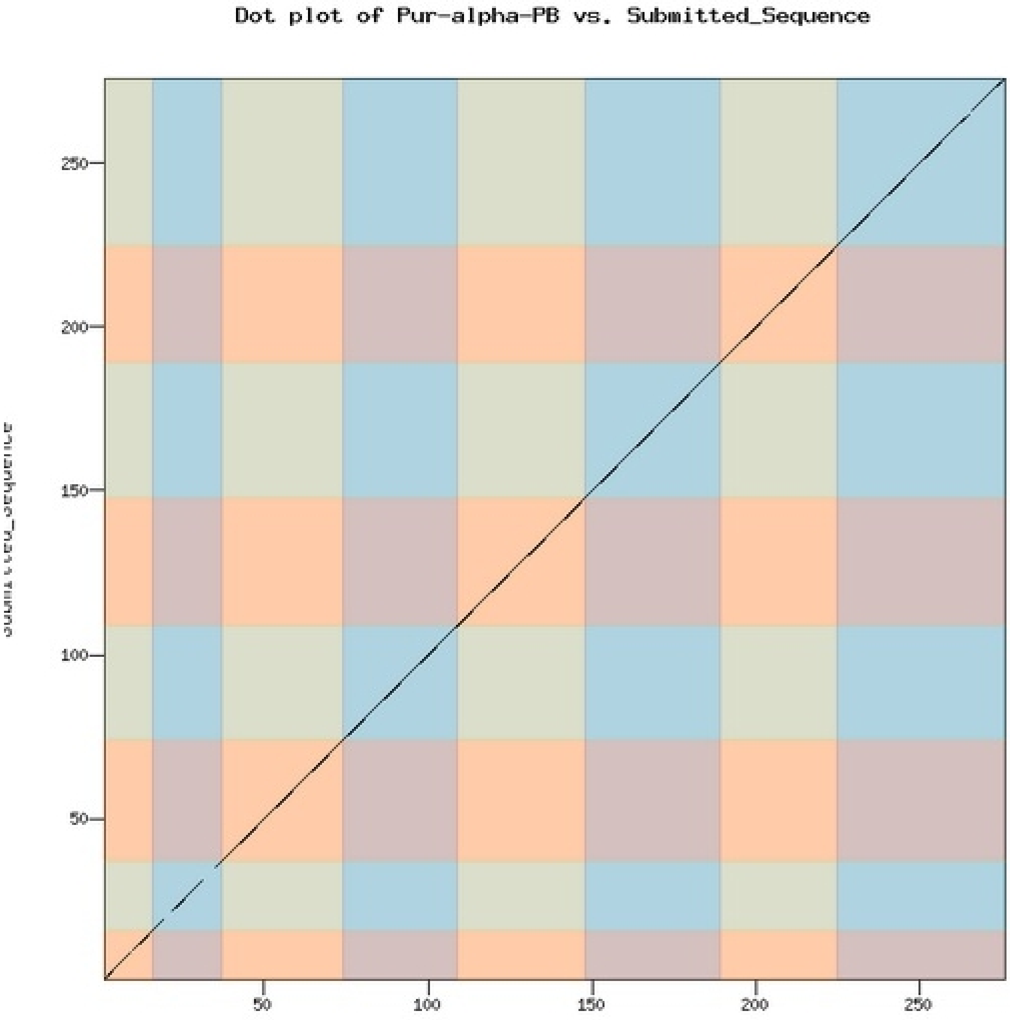
View plain text version

Identity: 269/275 (97.8%), **Similarity:** 273/275 (99.3%), **Gaps:** 0/275 (0.0%)

```
Pur-alpha-PB        1 MSDLGSGDDGISGSKYNVANMEGSSSRNDFDSSAKGGSGVEQELATKMLQIQSKRFYLDV  60
                      ********:*********:*:*********:**.**************************
Submitted_Seq       1 MSDLGSGDEGISGSKYNVTNIEGSSSRNDFESSVKGGSGVEQELATKMLQIQSKRFYLDV  60

Pur-alpha-PB       61 KQNRRGRFIKVAEIGADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGK 120
                      ************************************************************
Submitted_Seq      61 KQNRRGRFIKVAEIGADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGK 120

Pur-alpha-PB      121 LKSEMMIKDYRRYYLDLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLE 180
                      ************************************************************
Submitted_Seq     121 LKSEMMIKDYRRYYLDLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLE 180

Pur-alpha-PB      181 EFGANDGGRFKGDLPEERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCW 240
                      ************************************************************
Submitted_Seq     181 EFGANDGGRFKGDLPEERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCW 240

Pur-alpha-PB      241 IRFRDIFNDYCEKMKKSSDSITAEINLPTSSNSLK 275
                      ************************:**********
Submitted_Seq     241 IRFRDIFNDYCEKMKKSSDSITAEMNLPTSSNSLK 275
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).



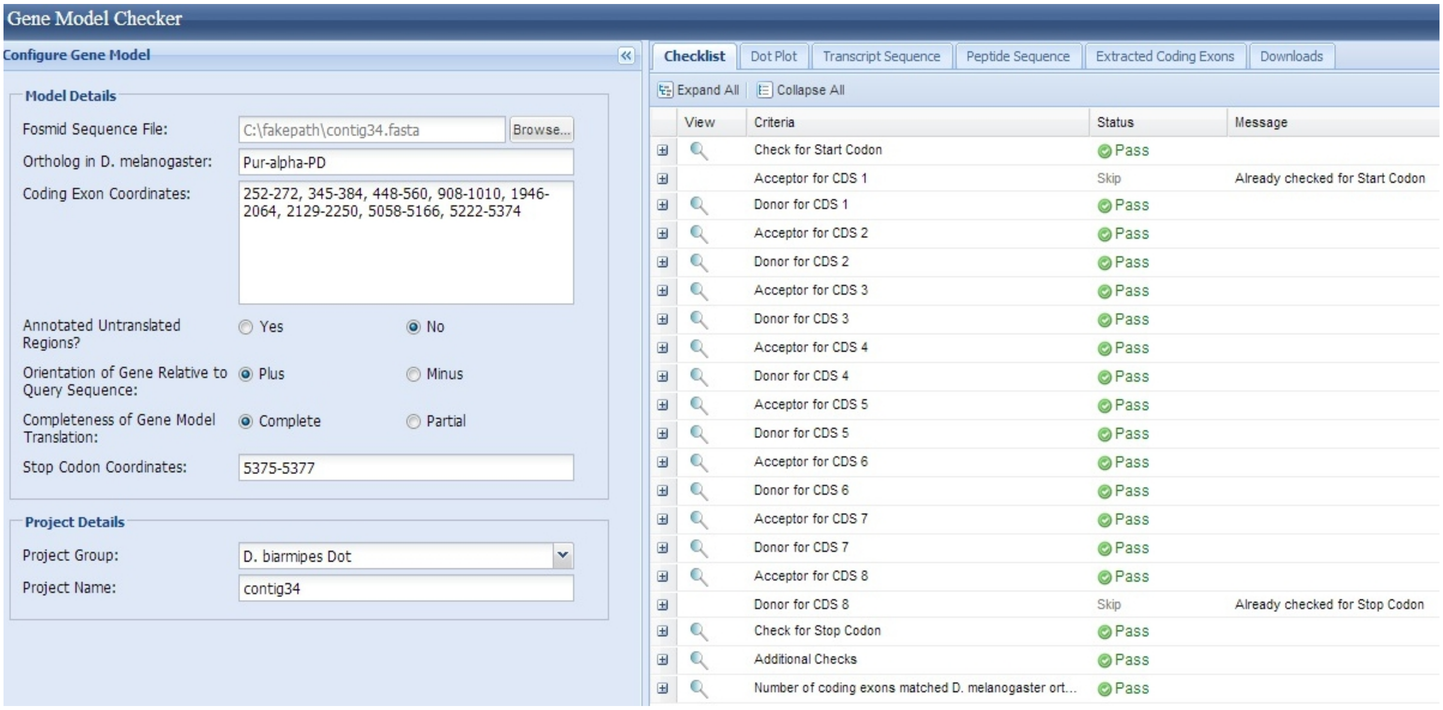
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): ____dbia_*Pur-α-PD*_______
Names of the isoforms with identical coding sequences as this isoform
____NA_____________
Is the 5' end of this isoform missing from the end of project: ___No____________
If so, how many exons are missing from the 5' end: ____NA____________
Is the 3' end of this isoform missing from the end of the project: __No___________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



Model Details

Fosmid Sequence File: C:\fakepath\contig34.fasta
Ortholog in D. melanogaster: Pur-alpha-PD
Coding Exon Coordinates: 252-272, 345-384, 448-560, 908-1010, 1946-2064, 2129-2250, 5058-5166, 5222-5374
Annotated Untranslated Regions? ○ Yes ● No
Orientation of Gene Relative to Query Sequence: ● Plus ○ Minus
Completeness of Gene Model Translation: ● Complete ○ Partial
Stop Codon Coordinates: 5375-5377

Project Details

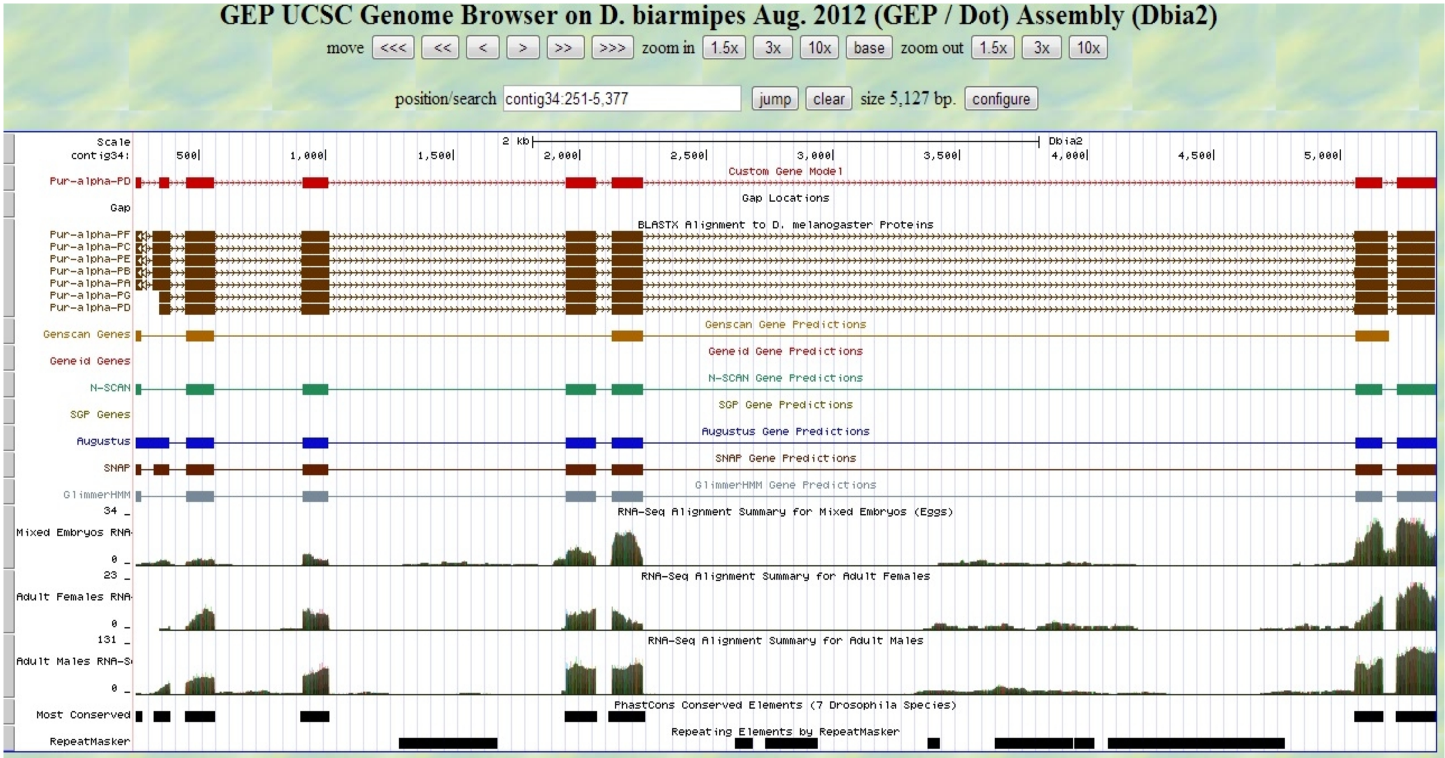
Project Group: D. biarmipes Dot
Project Name: contig34

View	Criteria	Status	Message
🔍	Check for Start Codon	✓ Pass	
	Acceptor for CDS 1	Skip	Already checked for Start Codon
🔍	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	
🔍	Number of coding exons matched D. melanogaster ort...	✓ Pass	

2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker, capture a screenshot of your gene model shown on the Genome

Paste the screenshot of your gene model as shown on the Genome Browser below:

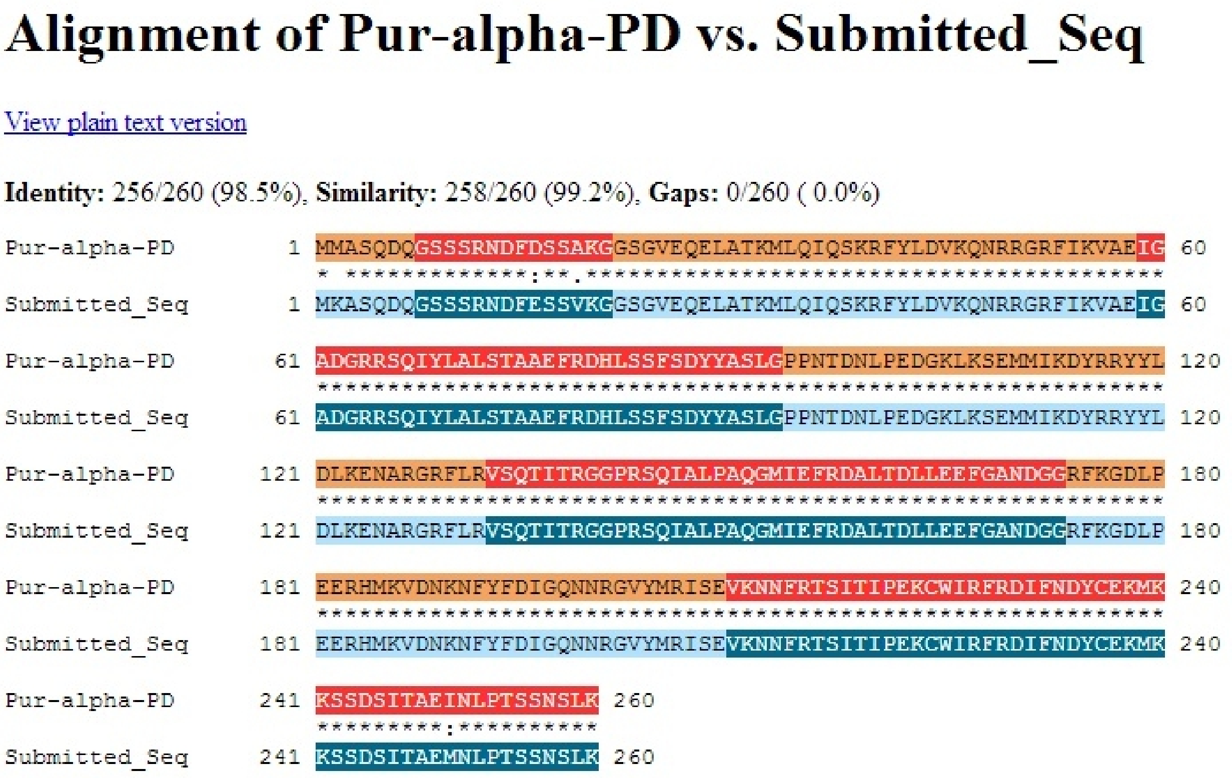


GEP UCSC Genome Browser on D. biarmipes Aug. 2012 (GEP / Dot) Assembly (Dbia2)

position/search contig34:251-5,377 size 5,127 bp.

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

Copy and paste the alignment below:



Alignment of Pur-alpha-PD vs. Submitted_Seq

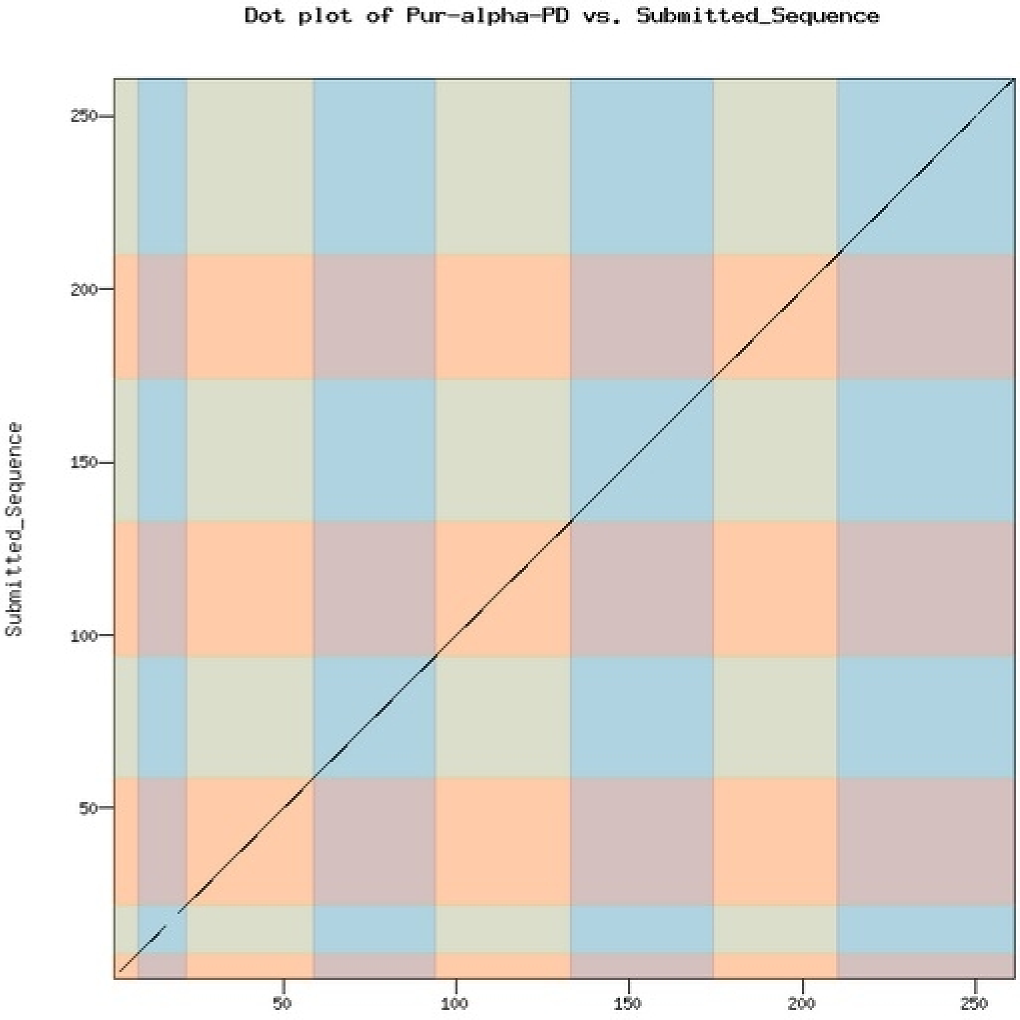
View plain text version

Identity: 256/260 (98.5%), **Similarity:** 258/260 (99.2%), **Gaps:** 0/260 (0.0%)

```
Pur-alpha-PD         1 MMASQDQGSSSRNDFDSSAKGGSGVEQELATKMLQIQSKRFYLDVKQNRRGRFIKVAEIG 60
                       * *************:**.*****************************************
Submitted_Seq        1 MKASQDQGSSSRNDFESSVKGGSGVEQELATKMLQIQSKRFYLDVKQNRRGRFIKVAEIG 60

Pur-alpha-PD        61 ADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGKLKSEMMIKDYRRYYL 120
                       ************************************************************
Submitted_Seq       61 ADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGKLKSEMMIKDYRRYYL 120

Pur-alpha-PD       121 DLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLEEFGANDGGRFKGDLP 180
                       ************************************************************
Submitted_Seq      121 DLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLEEFGANDGGRFKGDLP 180

Pur-alpha-PD       181 EERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCWIRFRDIFNDYCEKMK 240
                       ************************************************************
Submitted_Seq      181 EERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCWIRFRDIFNDYCEKMK 240

Pur-alpha-PD       241 KSSDSITAEINLPTSSNSLK 260
                       *********:**********
Submitted_Seq      241 KSSDSITAEMNLPTSSNSLK 260
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).



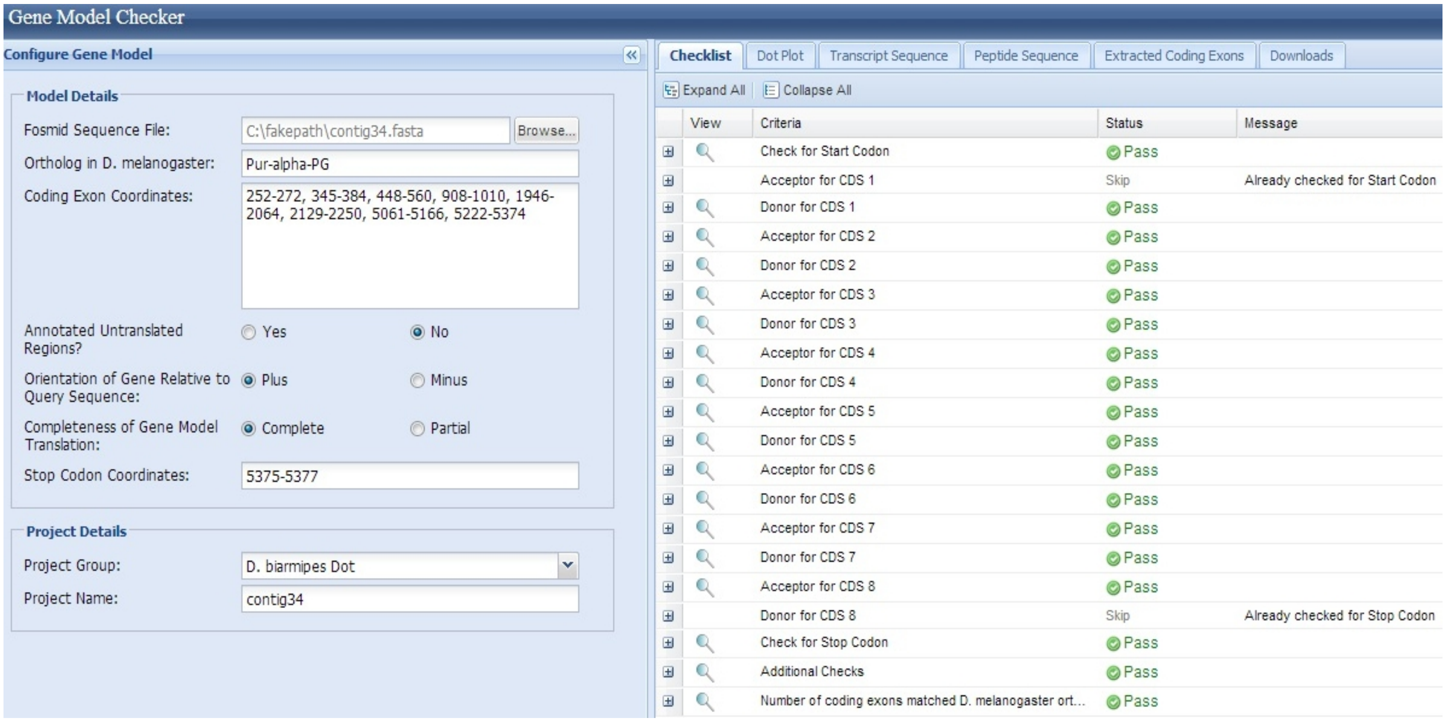
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): ____dbia_*Pur-α-PG*_______
Names of the isoforms with identical coding sequences as this isoform
____NA_____________
Is the 5' end of this isoform missing from the end of project: ___No____________
If so, how many exons are missing from the 5' end: ____NA____________
Is the 3' end of this isoform missing from the end of the project: __No___________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



Model Details

Fosmid Sequence File: C:\fakepath\contig34.fasta
Ortholog in D. melanogaster: Pur-alpha-PG
Coding Exon Coordinates: 252-272, 345-384, 448-560, 908-1010, 1946-2064, 2129-2250, 5061-5166, 5222-5374
Annotated Untranslated Regions? ○ Yes ● No
Orientation of Gene Relative to Query Sequence: ● Plus ○ Minus
Completeness of Gene Model Translation: ● Complete ○ Partial
Stop Codon Coordinates: 5375-5377

Project Details

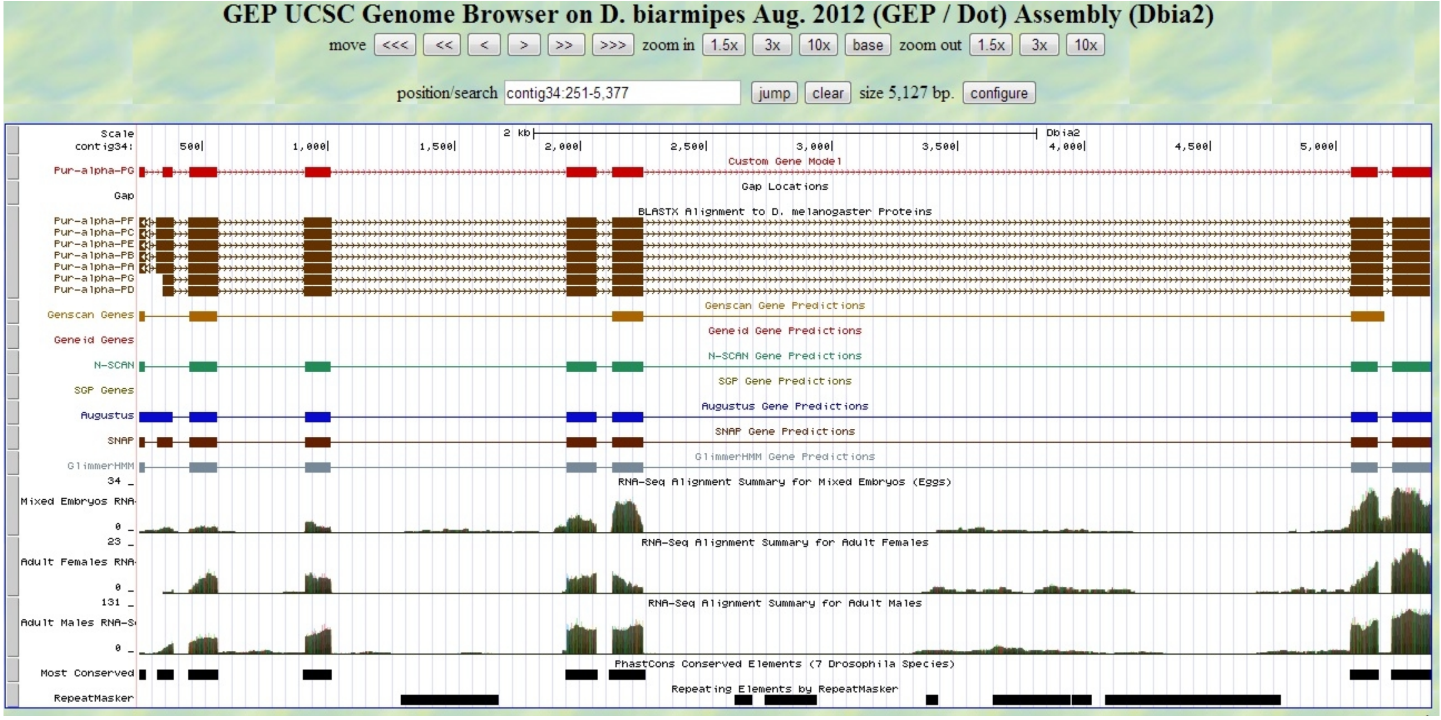
Project Group: D. biarmipes Dot
Project Name: contig34

Criteria	Status	Message
Check for Start Codon	Pass	
Acceptor for CDS 1	Skip	Already checked for Start Codon
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	
Number of coding exons matched D. melanogaster ort...	Pass	

2. View the gene model on the Genome Browser

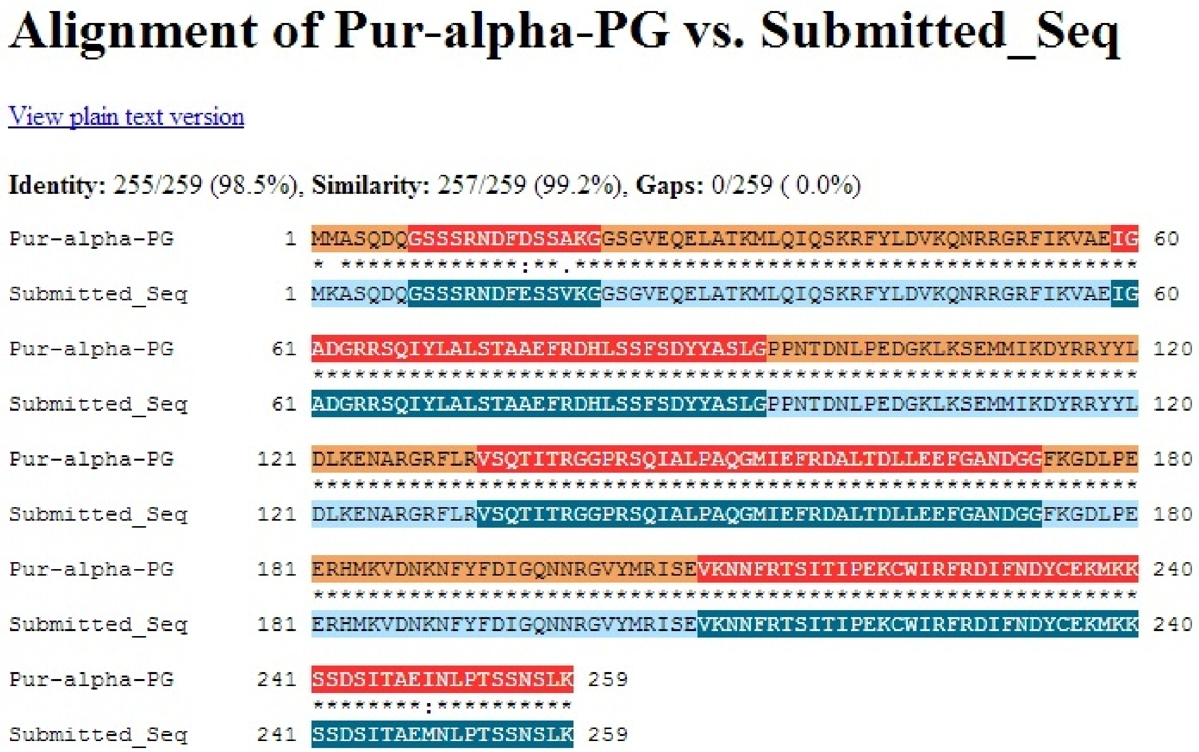
Using the custom track feature from the Gene Model Checker, capture a screenshot of your gene model shown on the Genome

Paste the screenshot of your gene model as shown on the Genome Browser below:



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

Copy and paste the alignment below:



Alignment of Pur-alpha-PG vs. Submitted_Seq

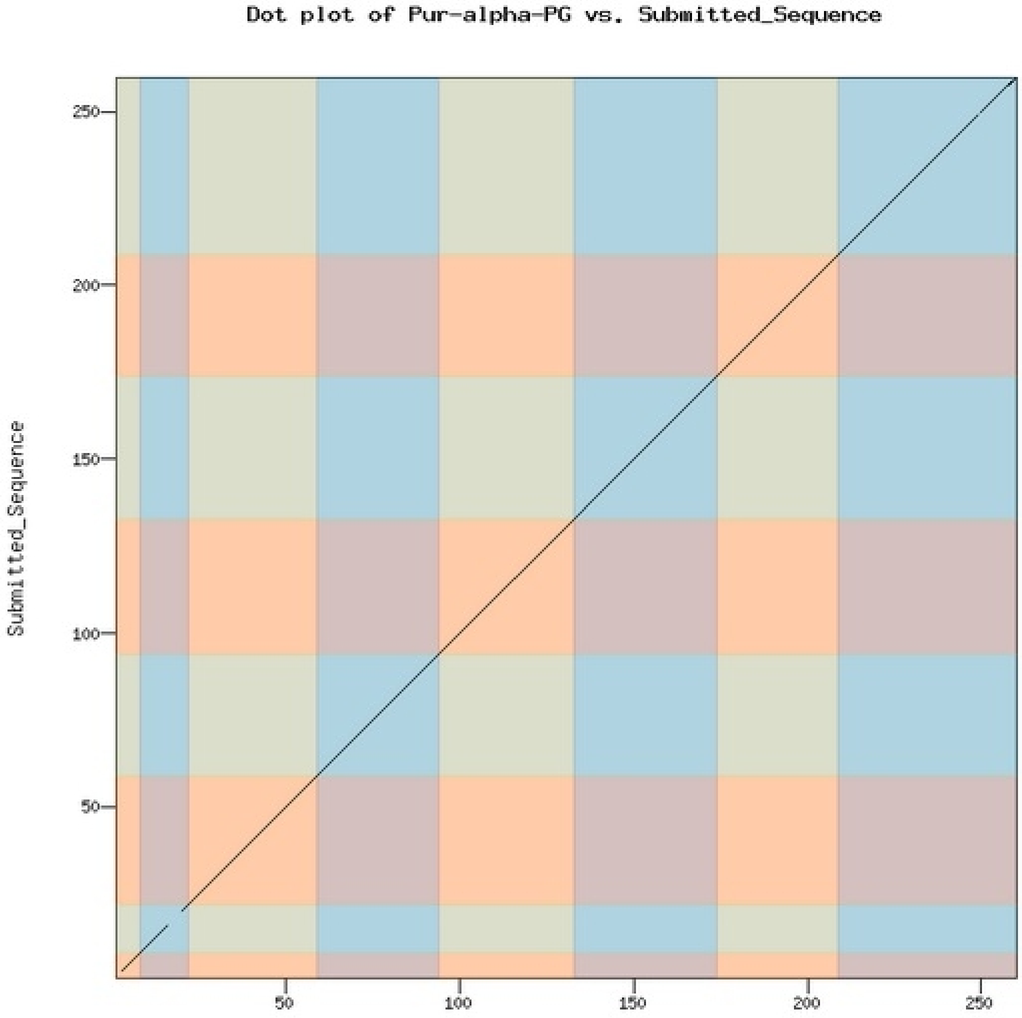
View plain text version

Identity: 255/259 (98.5%), **Similarity:** 257/259 (99.2%), **Gaps:** 0/259 (0.0%)

```
Pur-alpha-PG        1 MMASQDQGSSSRNDFDSSAKGGSGVEQELATKMLQIQSKRFYLDVKQNRRGRFIKVAEIG  60
                      * *************:**.*****************************************
Submitted_Seq       1 MKASQDQGSSSRNDFESSVKGGSGVEQELATKMLQIQSKRFYLDVKQNRRGRFIKVAEIG  60

Pur-alpha-PG       61 ADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGKLKSEMMIKDYRRYYL 120
                      ************************************************************
Submitted_Seq      61 ADGRRSQIYLALSTAAEFRDHLSSFSDYYASLGPPNTDNLPEDGKLKSEMMIKDYRRYYL 120

Pur-alpha-PG      121 DLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLEEFGANDGGFKGDLPE 180
                      ************************************************************
Submitted_Seq     121 DLKENARGRFLRVSQTITRGGPRSQIALPAQGMIEFRDALTDLLEEFGANDGGFKGDLPE 180

Pur-alpha-PG      181 ERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCWIRFRDIFNDYCEKMKK 240
                      ************************************************************
Submitted_Seq     181 ERHMKVDNKNFYFDIGQNNRGVYMRISEVKNNFRTSITIPEKCWIRFRDIFNDYCEKMKK 240

Pur-alpha-PG      241 SSDSITAEINLPTSSNSLK 259
                      ********:**********
Submitted_Seq     241 SSDSITAEMNLPTSSNSLK 259
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).



Gene report form

Gene name (i.e. *D. mojavensis eyeless*): __*D. biarmipes CG1970*___________
Gene symbol (i.e. dmoj_ey): ____dbia_CG1970____
Approximate location in project (from 5' end to 3' end): __12260-8439__________
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:
If you are annotating untranslated regions then all isoforms are unique (by definition)

Name of unique isoform based on coding sequence	List of isoforms with identical coding sequences
CG1970-PA	*CG1970-PB*

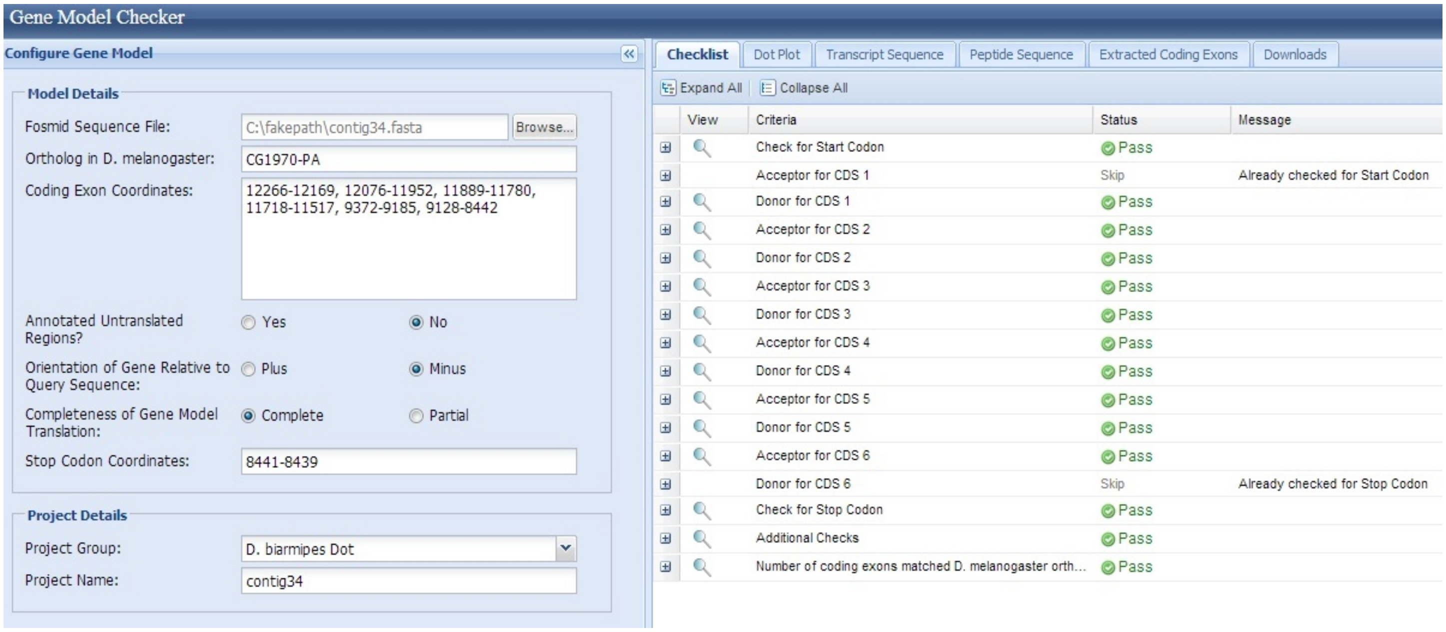
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): ___dbia_*CG1970-PA*_______
Names of the isoforms with identical coding sequences as this isoform
___dbia *CG1970-PB*_____________
Is the 5' end of this isoform missing from the end of project: ___No____________
If so, how many exons are missing from the 5' end: ____NA____________
Is the 3' end of this isoform missing from the end of the project: __No___________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



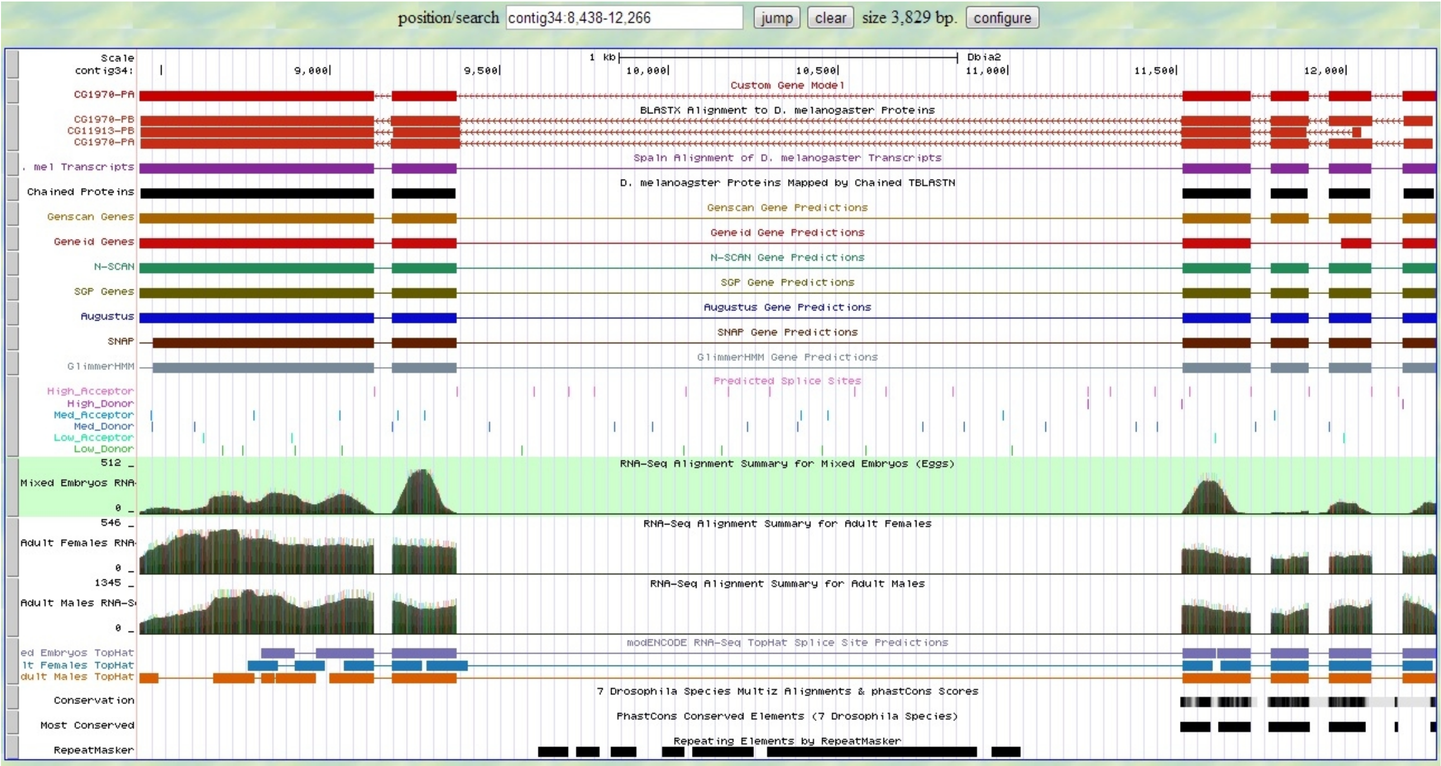
Fosmid Sequence File: C:\fakepath\contig34.fasta
Ortholog in D. melanogaster: CG1970-PA
Coding Exon Coordinates: 12266-12169, 12076-11952, 11889-11780, 11718-11517, 9372-9185, 9128-8442
Annotated Untranslated Regions? ○ Yes ● No
Orientation of Gene Relative to Query Sequence: ○ Plus ● Minus
Completeness of Gene Model Translation: ● Complete ○ Partial
Stop Codon Coordinates: 8441-8439
Project Group: D. biarmipes Dot
Project Name: contig34

Criteria	Status	Message
Check for Start Codon	Pass	
Acceptor for CDS 1	Skip	Already checked for Start Codon
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	Skip	Already checked for Stop Codon
Check for Stop Codon	Pass	
Additional Checks	Pass	
Number of coding exons matched D. melanogaster orth...	Pass	

2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker, capture a screenshot of your gene model shown on the Genome

Paste the screenshot of your gene model as shown on the Genome Browser below:

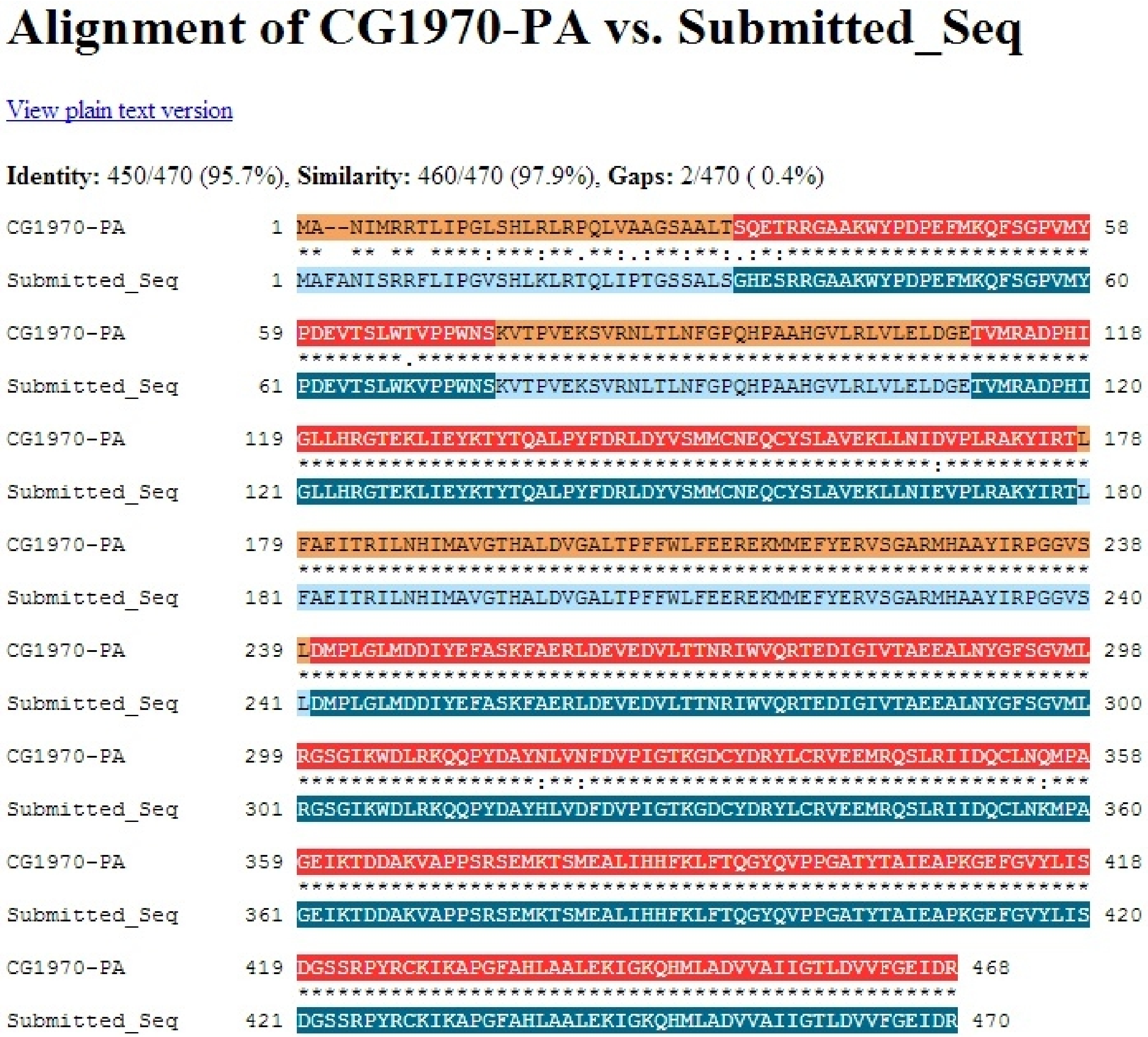


position/search contig34:8,438-12,266 size 3,829 bp.

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

Show an alignment between the protein sequence for your gene model and the protein sequence from the putative *D. melanogaster* ortholog. You can use the protein alignment generated by the Gene Model Checker or you can generate a new alignment using BLAST 2 Sequences (*bl2seq*).

Copy and paste the alignment below:



Alignment of CG1970-PA vs. Submitted_Seq

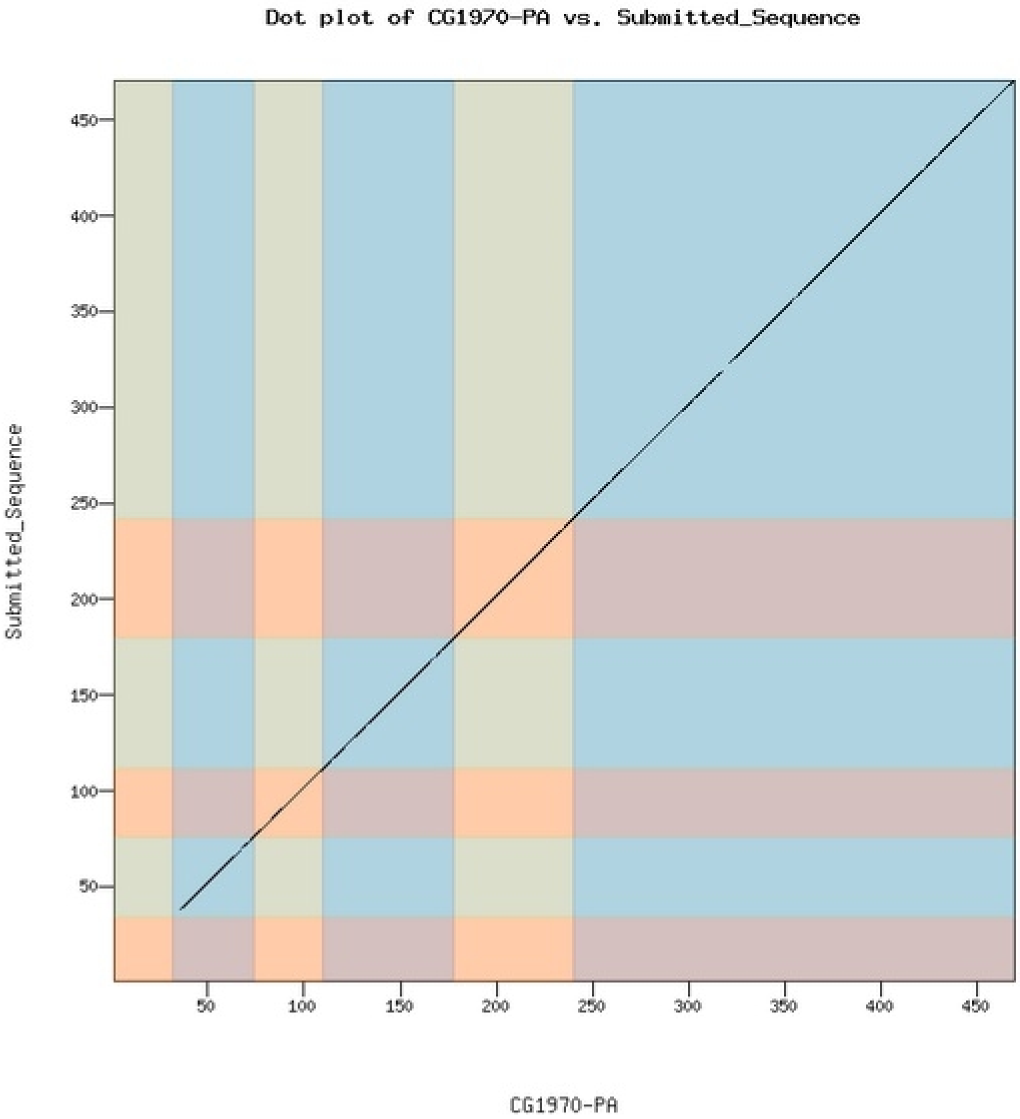
View plain text version

Identity: 450/470 (95.7%), **Similarity:** 460/470 (97.9%), **Gaps:** 2/470 (0.4%)

```
CG1970-PA           1 MA--NIMRRTLIPGLSHLRLRPQLVAAGSAALTSQETRRGAAKWYPDPEFMKQFSGPVMY 58
                      **  ** ** ****:***:**.**:.:**:**:.:*:***********************
Submitted_Seq       1 MAFANISRRFLIPGVSHLKLRTQLIPTGSSALSGHESRRGAAKWYPDPEFMKQFSGPVMY 60

CG1970-PA          59 PDEVTSLWTVPPWNSKVTPVEKSVRNLTLNFGPQHPAAHGVLRLVLELDGETVMRADPHI 118
                      ********.***************************************************
Submitted_Seq      61 PDEVTSLWKVPPWNSKVTPVEKSVRNLTLNFGPQHPAAHGVLRLVLELDGETVMRADPHI 120

CG1970-PA         119 GLLHRGTEKLIEYKTYTQALPYFDRLDYVSMMCNEQCYSLAVEKLLNIDVPLRAKYIRTL 178
                      ************************************************:***********
Submitted_Seq     121 GLLHRGTEKLIEYKTYTQALPYFDRLDYVSMMCNEQCYSLAVEKLLNIEVPLRAKYIRTL 180

CG1970-PA         179 FAEITRILNHIMAVGTHALDVGALTPFFWLFEEREKMMEFYERVSGARMHAAYIRPGGVS 238
                      ************************************************************
Submitted_Seq     181 FAEITRILNHIMAVGTHALDVGALTPFFWLFEEREKMMEFYERVSGARMHAAYIRPGGVS 240

CG1970-PA         239 LDMPLGLMDDIYEFASKFAERLDEVEDVLTTNRIWVQRTEDIGIVTAEEALNYGFSGVML 298
                      ************************************************************
Submitted_Seq     241 LDMPLGLMDDIYEFASKFAERLDEVEDVLTTNRIWVQRTEDIGIVTAEEALNYGFSGVML 300

CG1970-PA         299 RGSGIKWDLRKQQPYDAYNLVNFDVPIGTKGDCYDRYLCRVEEMRQSLRIIDQCLNQMPA 358
                      ******************:**:**********************************:***
Submitted_Seq     301 RGSGIKWDLRKQQPYDAYHLVDFDVPIGTKGDCYDRYLCRVEEMRQSLRIIDQCLNKMPA 360

CG1970-PA         359 GEIKTDDAKVAPPSRSEMKTSMEALIHHFKLFTQGYQVPPGATYTAIEAPKGEFGVYLIS 418
                      ************************************************************
Submitted_Seq     361 GEIKTDDAKVAPPSRSEMKTSMEALIHHFKLFTQGYQVPPGATYTAIEAPKGEFGVYLIS 420

CG1970-PA         419 DGSSRPYRCKIKAPGFAHLAALEKIGKQHMLADVVAIIGTLDVVFGEIDR 468
                      **************************************************
Submitted_Seq     421 DGSSRPYRCKIKAPGFAHLAALEKIGKQHMLADVVAIIGTLDVVFGEIDR 470
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).



Gene report form

Gene name (i.e. *D. mojavensis eyeless*): __*D. biarmipes Ephrin*___________
Gene symbol (i.e. dmoj_ey): ____dbia_Ephrin____
Approximate location in project (from 5' end to 3' end): __17540-21945__________
Number of isoforms in *D. melanogaster:* _3_______________
Number of isoforms in this project: ____3____________

Complete the following table for all the isoforms in this project:
If you are annotating untranslated regions then all isoforms are unique (by definition)

Name of unique isoform based on coding sequence	**List of isoforms with identical coding sequences**
CG1970-PA	*CG1970-PB*

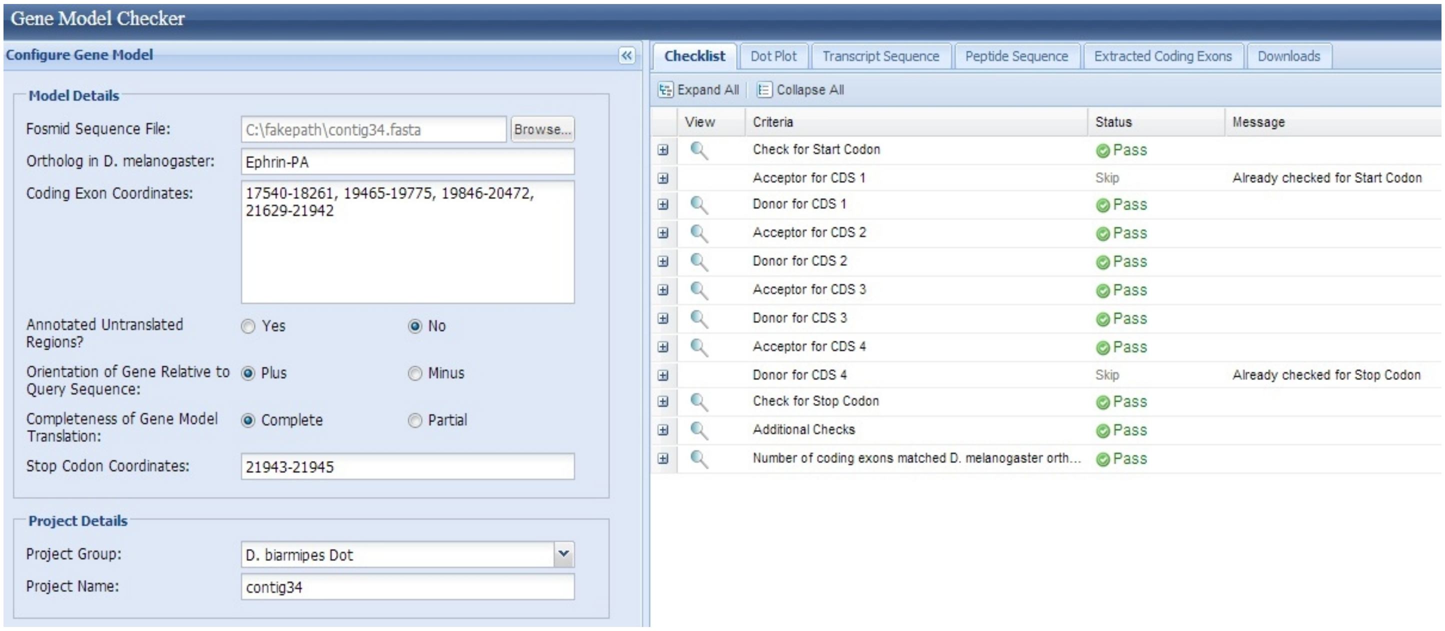
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): ___dbia_*Ephrin-PA*_______
Names of the isoforms with identical coding sequences as this isoform
___dbia *Ephrin-PB, dbia Ephrin-PC*_____________
Is the 5' end of this isoform missing from the end of project: ___No____________
If so, how many exons are missing from the 5' end: ____NA____________
Is the 3' end of this isoform missing from the end of the project: __No___________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



Model Details

Fosmid Sequence File: C:\fakepath\contig34.fasta
Ortholog in D. melanogaster: Ephrin-PA
Coding Exon Coordinates: 17540-18261, 19465-19775, 19846-20472, 21629-21942
Annotated Untranslated Regions? ○ Yes ● No
Orientation of Gene Relative to Query Sequence: ● Plus ○ Minus
Completeness of Gene Model Translation: ● Complete ○ Partial
Stop Codon Coordinates: 21943-21945

Project Details

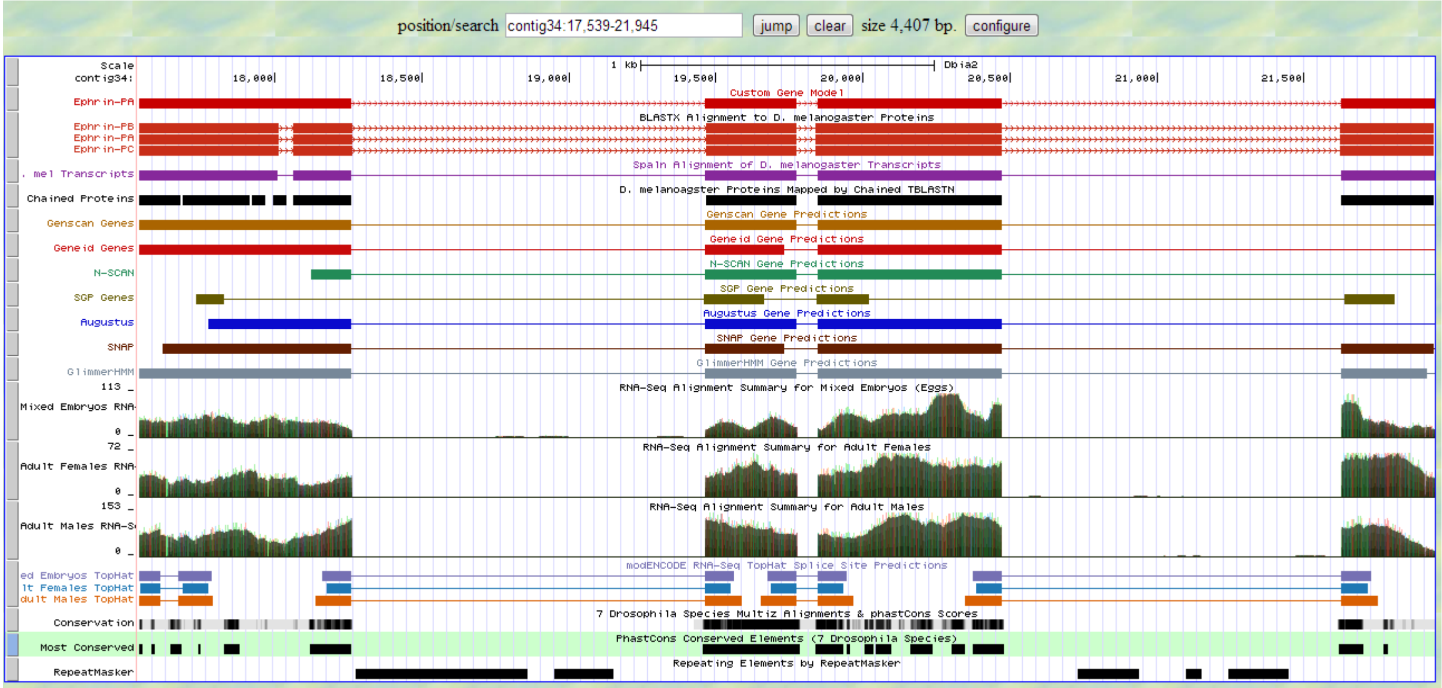
Project Group: D. biarmipes Dot
Project Name: contig34

Criteria	Status	Message
Check for Start Codon	Pass	
Acceptor for CDS 1	Skip	Already checked for Start Codon
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	Skip	Already checked for Stop Codon
Check for Stop Codon	Pass	
Additional Checks	Pass	
Number of coding exons matched D. melanogaster orth...	Pass	

2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker, capture a screenshot of your gene model shown on the Genome

Paste the screenshot of your gene model as shown on the Genome Browser below:

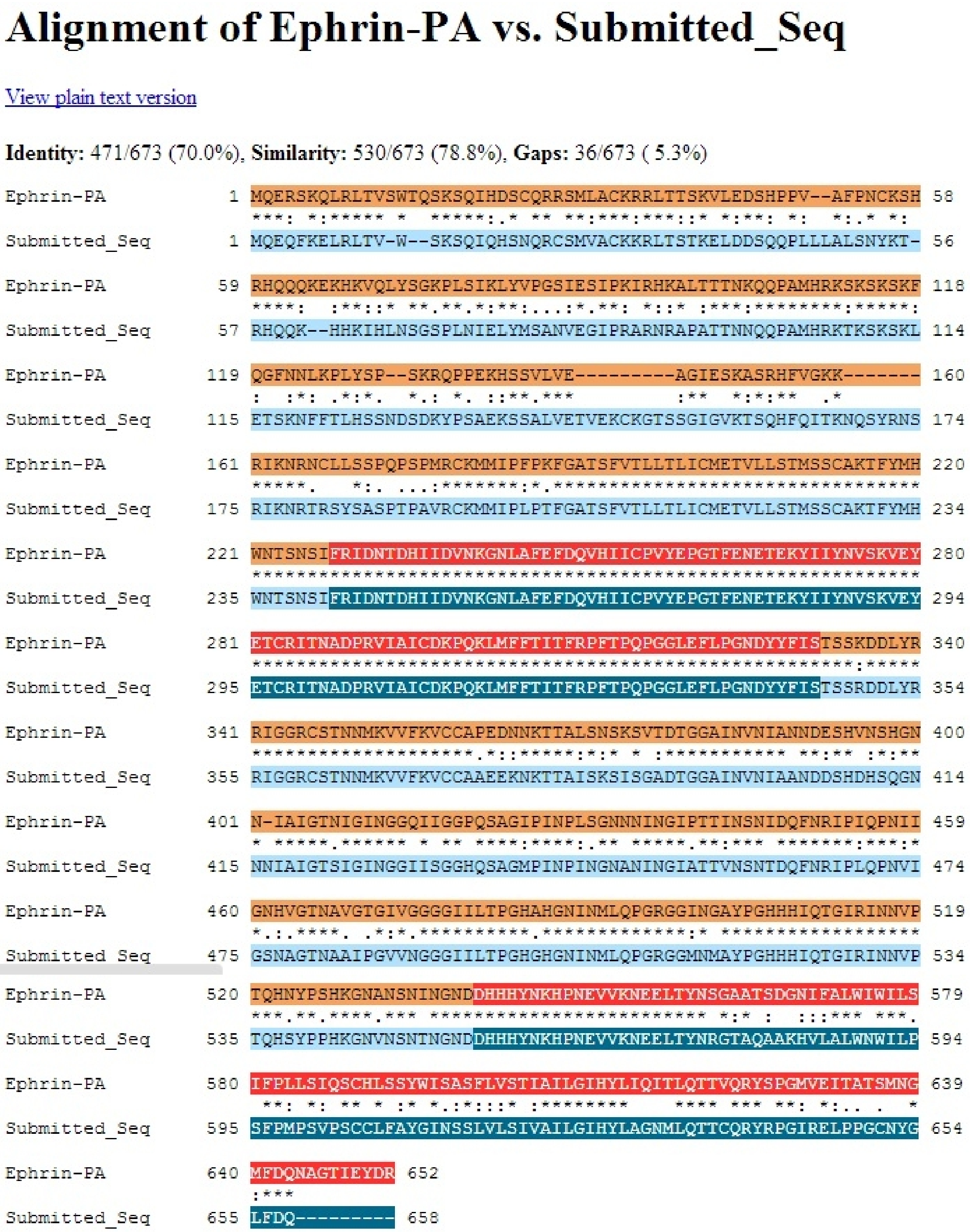


position/search contig34:17,539-21,945

size 4,407 bp.

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

Copy and paste the alignment below:



Alignment of Ephrin-PA vs. Submitted_Seq

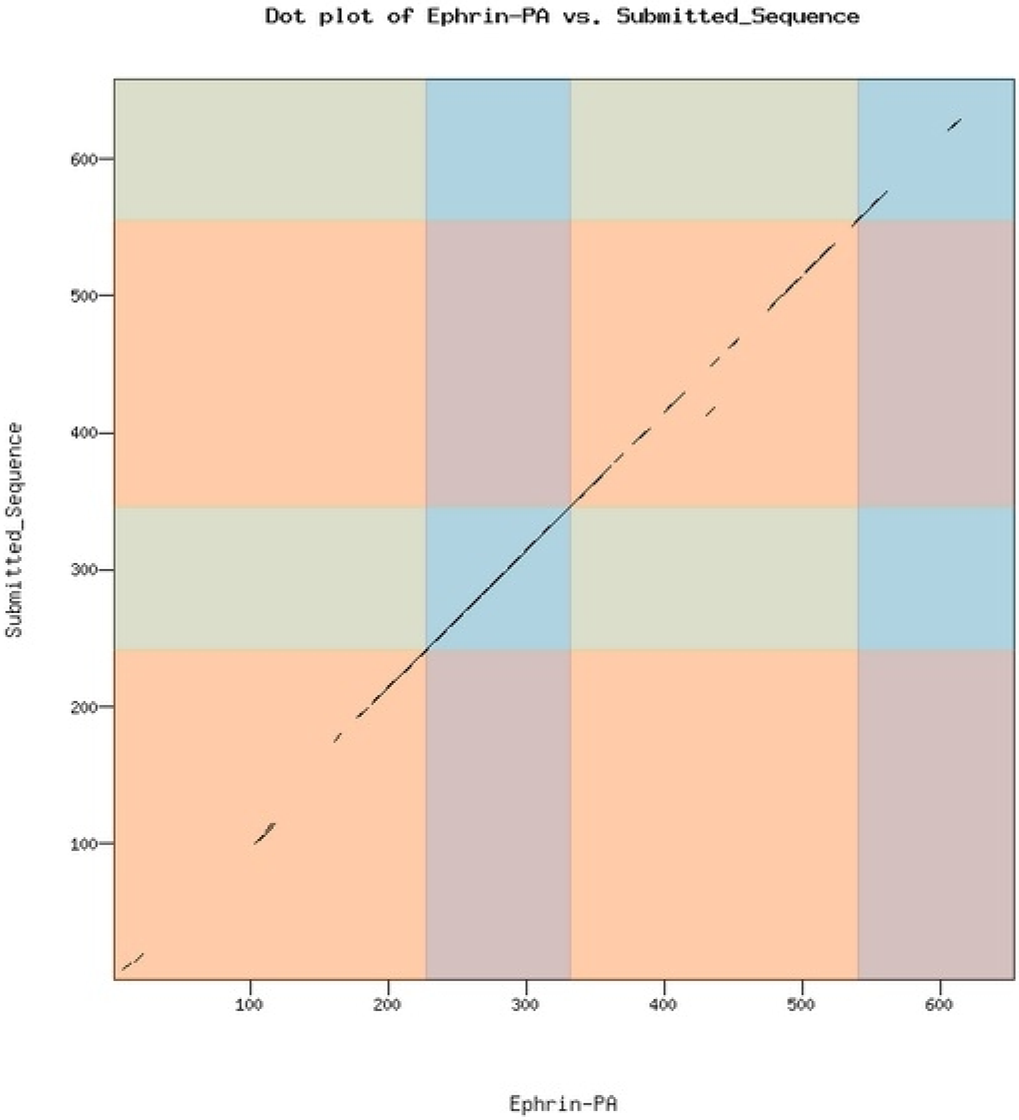
View plain text version

Identity: 471/673 (70.0%), **Similarity:** 530/673 (78.8%), **Gaps:** 36/673 (5.3%)

```
Ephrin-PA           1 MQERSKQLRLTVSWTQSKSQIHDSCQRRSMLACKRRLTTSKVLEDSHPPV--AFPNCKSH 58
                      ***: *:***** *  *****:.* ** **:***:***::* *:**: *:  *:.* *:
Submitted_Seq       1 MQEQFKELRLTV-W--SKSQIQHSNQRCSMVACKKRLTSTKELDDSQQPLLLALSNYKT- 56

Ephrin-PA          59 RHQQQKEKHKVQLYSGKPLSIKLYVPGSIESIPKIRHKALTTTNKQQPAMHRKSKSKSKF 118
                      ****:  :**::* **.**.*:**:...:*.**: *::* :***:********:*****:
Submitted_Seq      57 RHQQK--HHKIHLNSGSPLNIELYMSANVEGIPRARNRAPATTNNQQPAMHRKTKSKSKL 114

Ephrin-PA         119 QGFNNLKPLYSP--SKRQPPEKHSSVLVE---------AGIESKASRHFVGKK------- 160
                      :  :*: .*:*.  *.: *. ::**.***         :**  *:*:**  .*
Submitted_Seq     115 ETSKNFFTLHSSNDSDKYPSAEKSSALVETVEKCKGTSSGIGVKTSQHFQITKNQSYRNS 174

Ephrin-PA         161 RIKNRNCLLSSPQPSPMRCKMMIPFPKFGATSFVTLLTLICMETVLLSTMSSCAKTFYMH 220
                      *****.   *:. ...:*******:*.*********************************
Submitted_Seq     175 RIKNRTRSYSASPTPAVRCKMMIPLPTFGATSFVTLLTLICMETVLLSTMSSCAKTFYMH 234

Ephrin-PA         221 WNTSNSIFRIDNTDHIIDVNKGNLAFEFDQVHIICPVYEPGTFENETEKYIIYNVSKVEY 280
                      ************************************************************
Submitted_Seq     235 WNTSNSIFRIDNTDHIIDVNKGNLAFEFDQVHIICPVYEPGTFENETEKYIIYNVSKVEY 294

Ephrin-PA         281 ETCRITNADPRVIAICDKPQKLMFFTITFRPFTPQPGGLEFLPGNDYYFISTSSKDDLYR 340
                      ******************************************************:*****
Submitted_Seq     295 ETCRITNADPRVIAICDKPQKLMFFTITFRPFTPQPGGLEFLPGNDYYFISTSSRDDLYR 354

Ephrin-PA         341 RIGGRCSTNNMKVVFKVCCAPEDNNKTTALSNSKSVTDTGGAINVNIANNDESHVNSHGN 400
                      ********************.*::*****:*:* * :*********** **:** :*:**
Submitted_Seq     355 RIGGRCSTNNMKVVFKVCCAAEEKNKTTAISKSISGADTGGAINVNIAANDDSHDHSQGN 414

Ephrin-PA         401 N-IAIGTNIGINGGQIIGGPQSAGIPINPLSGNNNINGIPTTINSNIDQFNRIPIQPNII 459
                      * *****.****** * ** ****:****:.** *****.**:*** *******:***:*
Submitted_Seq     415 NNIAIGTSIGINGGIISGGHQSAGMPINPINGNANINGIATTVNSNTDQFNRIPLQPNVI 474

Ephrin-PA         460 GNHVGTNAVGTGIVGGGGIILTPGHAHGNINMLQPGRGGINGAYPGHHHIQTGIRINNVP 519
                      *.:.****. .*:*.**********.*************:* ******************
Submitted_Seq     475 GSNAGTNAAIPGVVNGGGIILTPGHGHGNINMLQPGRGGMNMAYPGHHHIQTGIRINNVP 534

Ephrin-PA         520 TQHNYPSHKGNANSNINGNDDHHHYNKHPNEVVKNEELTYNSGAATSDGNIFALWIWILS 579
                      ***.**.****.*** ************************* *:* :  :::*** ***.
Submitted_Seq     535 TQHSYPPHKGNVNSNTNGNDDHHHYNKHPNEVVKNEELTYNRGTAQAAKHVLALWNWILP 594

Ephrin-PA         580 IFPLLSIQSCHLSSYWISASFLVSTIAILGIHYLIQITLQTTVQRYSPGMVEITATSMNG 639
                       **: *: ** * :* *.:*:::* :********    **** *** **: *:.. .  *
Submitted_Seq     595 SFPMPSVPSCCLFAYGINSSLVLSIVAILGIHYLAGNMLQTTCQRYRPGIRELPPGCNYG 654

Ephrin-PA         640 MFDQNAGTIEYDR 652
                      :***
Submitted_Seq     655 LFDQ--------- 658
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).



Gene report form

Gene name (i.e. *D. mojavensis eyeless*): __*D. biarmipes CG1909*___________
Gene symbol (i.e. dmoj_ey): ____dbia_CG1909____
Approximate location in project (from 5' end to 3' end): __27459-27496__________
Number of isoforms in *D. melanogaster:* _2_______________
Number of isoforms in this project: ____1 New isoform with new intron in CDS 6____________

Complete the following table for all the isoforms in this project:
If you are annotating untranslated regions then all isoforms are unique (by definition)

Name of unique isoform based on coding sequence	**List of isoforms with identical coding sequences**
CG1970-Unamed Isoform	

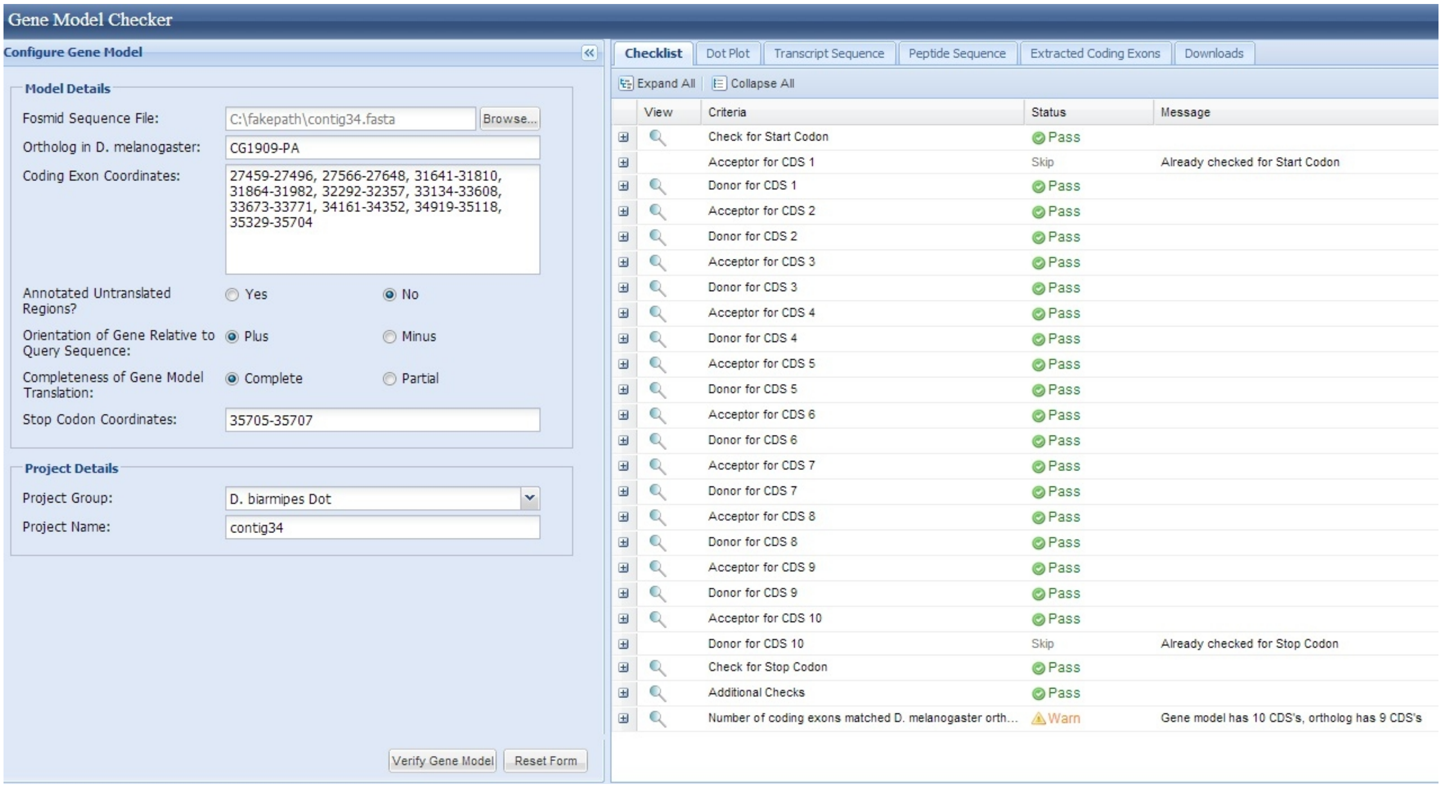
Isoform report form

Complete this report form for each unique isoform listed in the table above (copy and paste to create as many copies of this Isoform Report Form as needed):

Gene-isoform name (i.e. dmoj_ey-PA): ___dbia_*CG1909-Unamed Isoform*_______
Names of the isoforms with identical coding sequences as this isoform
___NA_____________
Is the 5' end of this isoform missing from the end of project: ___No____________
If so, how many exons are missing from the 5' end: ____NA____________
Is the 3' end of this isoform missing from the end of the project: __No___________
If so, how many exons are missing from the 3' end: ___NA___________

1. Gene Model Checker checklist

Enter the coordinates of your final gene model for this isoform into the Gene Model Checker and **paste a screenshot of the checklist results below:**



Model Details

Fosmid Sequence File: C:\fakepath\contig34.fasta

Ortholog in D. melanogaster: CG1909-PA

Coding Exon Coordinates: 27459-27496, 27566-27648, 31641-31810, 31864-31982, 32292-32357, 33134-33608, 33673-33771, 34161-34352, 34919-35118, 35329-35704

Annotated Untranslated Regions? ○ Yes ● No

Orientation of Gene Relative to Query Sequence: ● Plus ○ Minus

Completeness of Gene Model Translation: ● Complete ○ Partial

Stop Codon Coordinates: 35705-35707

Project Details

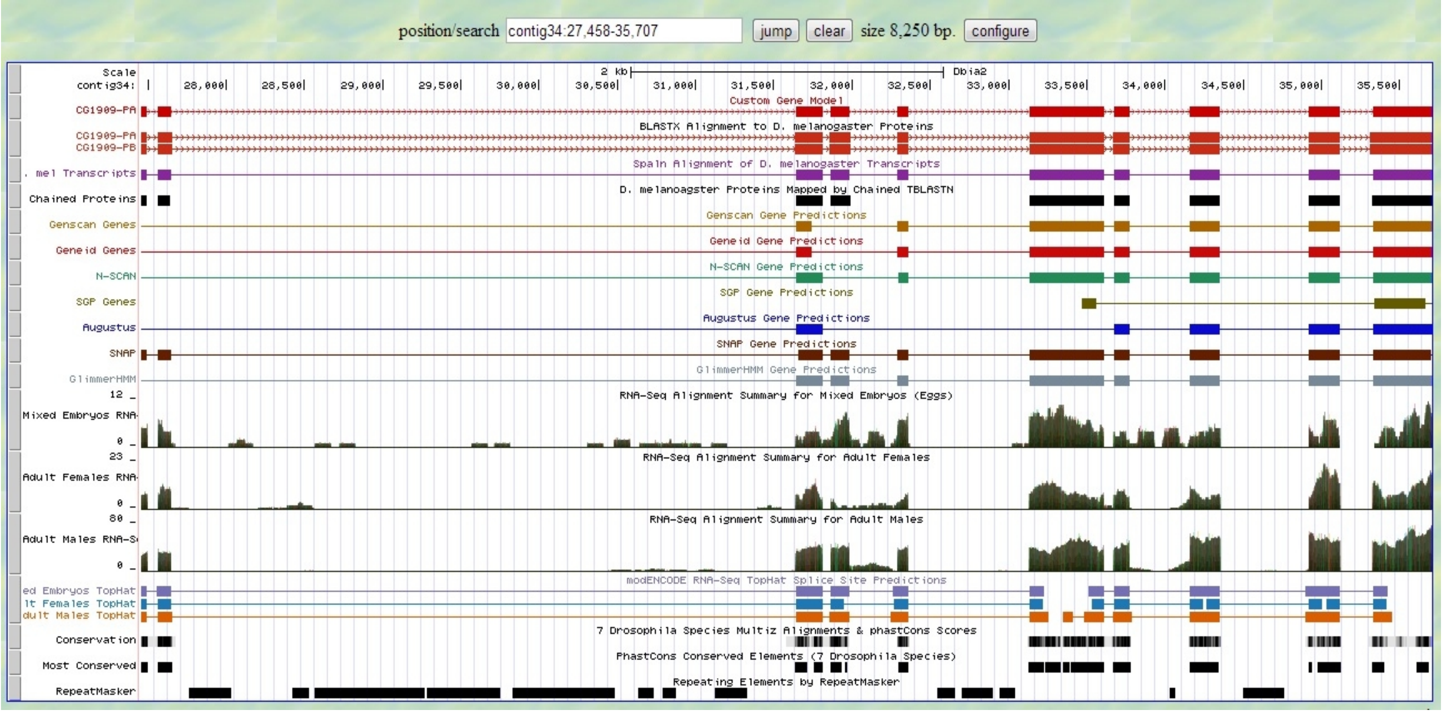
Project Group: D. biarmipes Dot

Project Name: contig34

View	Criteria	Status	Message
🔍	Check for Start Codon	✓ Pass	
	Acceptor for CDS 1	Skip	Already checked for Start Codon
🔍	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	✓ Pass	
🔍	Acceptor for CDS 9	✓ Pass	
🔍	Donor for CDS 9	✓ Pass	
🔍	Acceptor for CDS 10	✓ Pass	
	Donor for CDS 10	Skip	Already checked for Stop Codon
🔍	Check for Stop Codon	✓ Pass	
🔍	Additional Checks	✓ Pass	
🔍	Number of coding exons matched D. melanogaster orth...	⚠ Warn	Gene model has 10 CDS's, ortholog has 9 CDS's

2. View the gene model on the Genome Browser

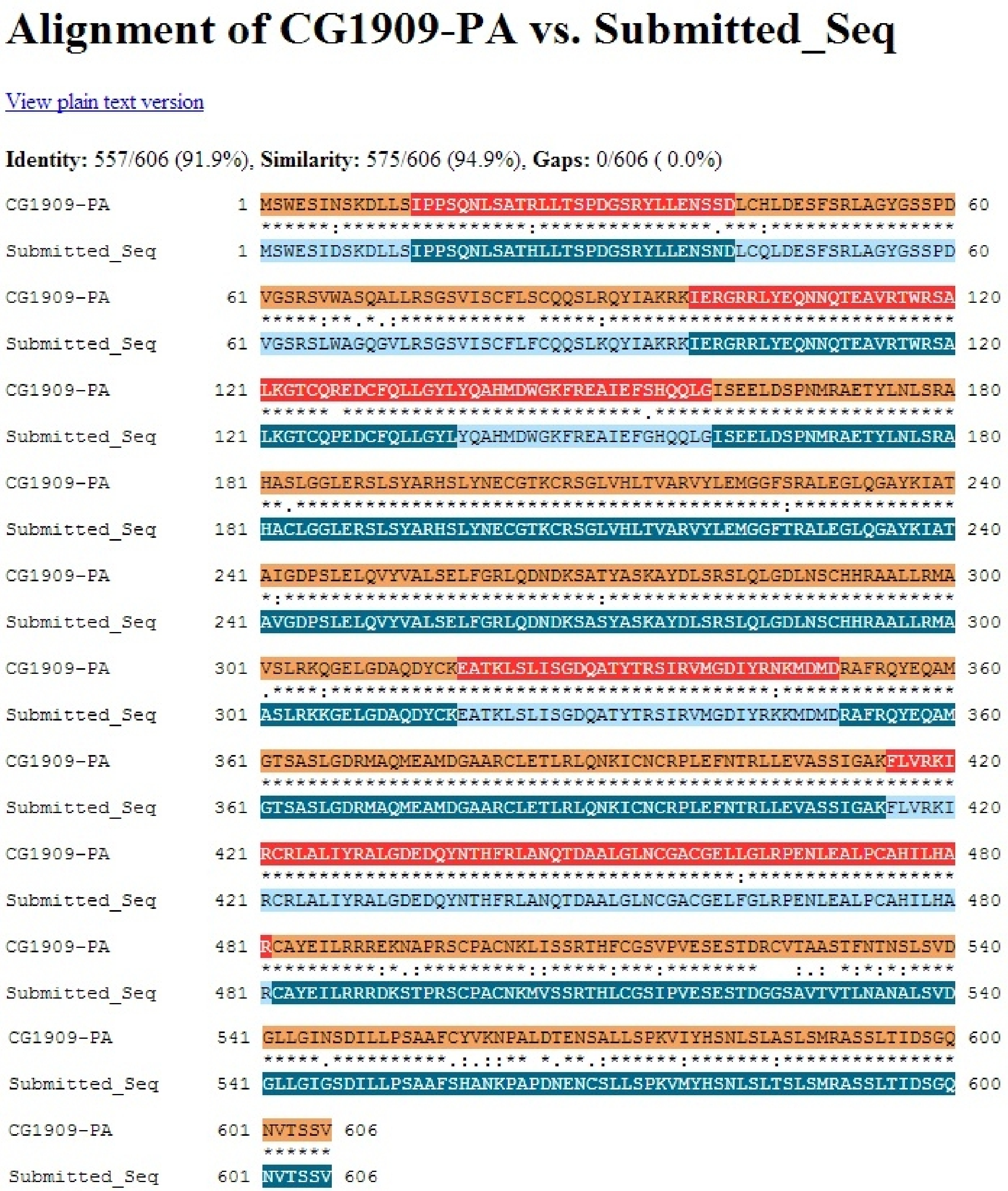
Paste the screenshot of your gene model as shown on the Genome Browser below:



position/search contig34:27,458-35,707 size 8,250 bp.

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

Copy and paste the alignment below:



Alignment of CG1909-PA vs. Submitted_Seq

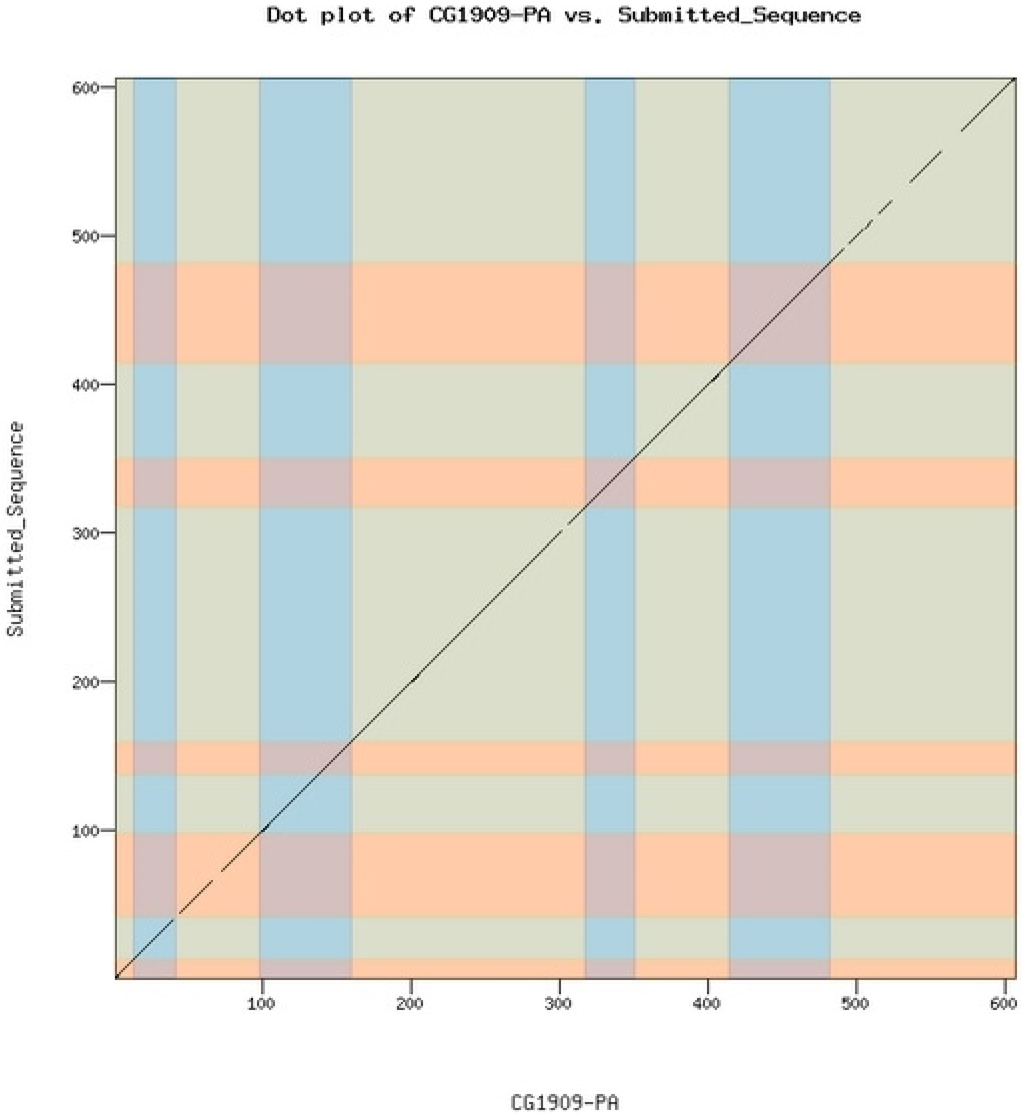
View plain text version

Identity: 557/606 (91.9%), **Similarity:** 575/606 (94.9%), **Gaps:** 0/606 (0.0%)

```
CG1909-PA          1 MSWESINSKDLLSIPPSQNLSATRLLTSPDGSRYLLENSSDLCHLDESFSRLAGYGSSPD 60
                     ******:****************:***************.***:****************
Submitted_Seq      1 MSWESIDSKDLLSIPPSQNLSATHLLTSPDGSRYLLENSNDLCQLDESFSRLAGYGSSPD 60

CG1909-PA         61 VGSRSVWASQALLRSGSVISCFLSCQQSLRQYIAKRKIERGRRLYEQNNQTEAVRTWRSA 120
                     *****:**.*.:*********** *****:******************************
Submitted_Seq     61 VGSRSLWAGQGVLRSGSVISCFLFCQQSLKQYIAKRKIERGRRLYEQNNQTEAVRTWRSA 120

CG1909-PA        121 LKGTCQREDCFQLLGYLYQAHMDWGKFREAIEFSHQQLGISEELDSPNMRAETYLNLSRA 180
                     ****** **************************.**************************
Submitted_Seq    121 LKGTCQPEDCFQLLGYLYQAHMDWGKFREAIEFGHQQLGISEELDSPNMRAETYLNLSRA 180

CG1909-PA        181 HASLGGLERSLSYARHSLYNECGTKCRSGLVHLTVARVYLEMGGFSRALEGLQGAYKIAT 240
                     **.******************************************:**************
Submitted_Seq    181 HACLGGLERSLSYARHSLYNECGTKCRSGLVHLTVARVYLEMGGFTRALEGLQGAYKIAT 240

CG1909-PA        241 AIGDPSLELQVYVALSELFGRLQDNDKSATYASKAYDLSRSLQLGDLNSCHHRAALLRMA 300
                     *:***************************:******************************
Submitted_Seq    241 AVGDPSLELQVYVALSELFGRLQDNDKSASYASKAYDLSRSLQLGDLNSCHHRAALLRMA 300

CG1909-PA        301 VSLRKQGELGDAQDYCKEATKLSLISGDQATYTRSIRVMGDIYRNKMDMDRAFRQYEQAM 360
                     .****:**************************************:***************
Submitted_Seq    301 ASLRKKGELGDAQDYCKEATKLSLISGDQATYTRSIRVMGDIYRKKMDMDRAFRQYEQAM 360

CG1909-PA        361 GTSASLGDRMAQMEAMDGAARCLETLRLQNKICNCRPLEFNTRLLEVASSIGAKFLVRKI 420
                     ************************************************************
Submitted_Seq    361 GTSASLGDRMAQMEAMDGAARCLETLRLQNKICNCRPLEFNTRLLEVASSIGAKFLVRKI 420

CG1909-PA        421 RCRLALIYRALGDEDQYNTHFRLANQTDAALGLNCGACGELLGLRPENLEALPCAHILHA 480
                     *****************************************:******************
Submitted_Seq    421 RCRLALIYRALGDEDQYNTHFRLANQTDAALGLNCGACGELFGLRPENLEALPCAHILHA 480

CG1909-PA        481 RCAYEILRRREKNAPRSCPACNKLISSRTHFCGSVPVESESTDRCVTAASTFNTNSLSVD 540
                     **********:*.:*********::*****:***:********   :.: *:*:*:****
Submitted_Seq    481 RCAYEILRRRDKSTPRSCPACNKMVSSRTHLCGSIPVESESTDGGSAVTVTLNANALSVD 540

CG1909-PA        541 GLLGINSDILLPSAAFCYVKNPALDTENSALLSPKVIYHSNLSLASLSMRASSLTIDSGQ 600
                     *****.**********.:.::** *.**.:******:*******:***************
Submitted_Seq    541 GLLGIGSDILLPSAAFSHANKPAPDNENCSLLSPKVMYHSNLSLTSLSMRASSLTIDSGQ 600

CG1909-PA        601 NVTSSV 606
                     ******
Submitted_Seq    601 NVTSSV 606
```

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

Paste a copy of the dot plot of your submitted model against the putative *D. melanogaster* ortholog (generated by the Gene Model Checker).
Provide an explanation for any anomalies on the dot plot (e.g. large gaps, regions with no sequence similarity).

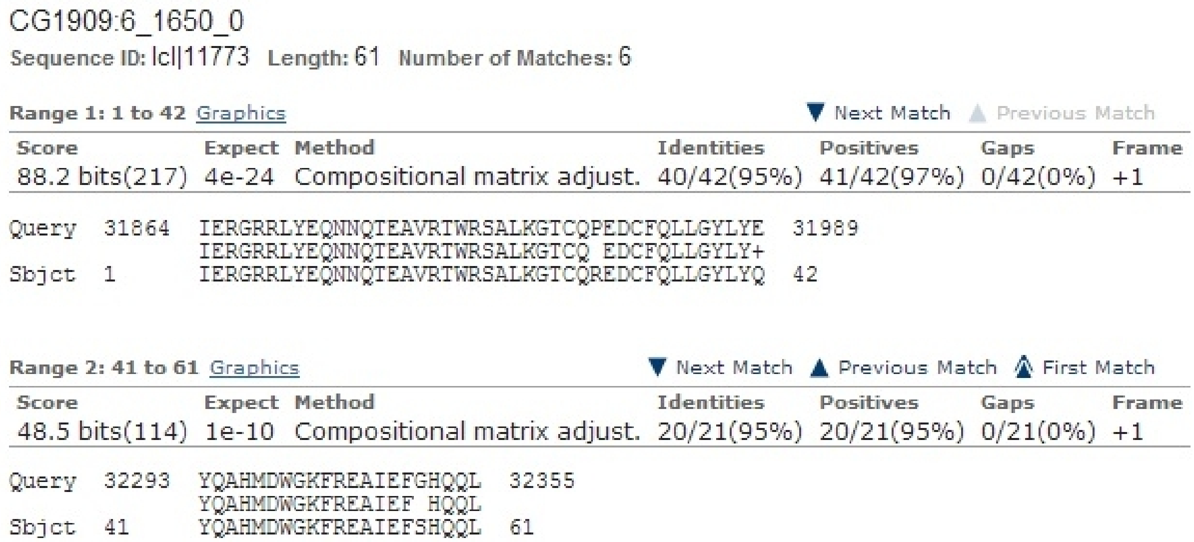


* Support for new intron in 4th exon of *CG1909* CDS 6_1650_0

As shown in the below image, 22 amino acids from CDS 6 in the gene have been separated from the remainder of the exon by 310 base pairs. There are several stop codons in the frame 1 where the proteins of this exon exist in this region meaning that this cannot be a continuous open reading frame. The circled region labeled “1” in the below image contains the initial proteins in the exon and the circled region labeled “3” contains the remaining amino acids from the exon. The region labeled “2” shows several stop codons in the reading frame.

Each of the good coding regions from the exon contains good splice sites allowing for a new intron that contains these stop codons.

BLASTX results of comparing the protein sequence from CDS 6 to the *D. biarmipes* contig 34 show significant hits on the entire protein but it is across two different regions to encompass the entire protein



CG1909:6_1650_0

Sequence ID: lcl|11773 Length: 61 Number of Matches: 6

Range 1: 1 to 42 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
88.2 bits(217)	4e-24	Compositional matrix adjust.	40/42(95%)	41/42(97%)	0/42(0%)	+1

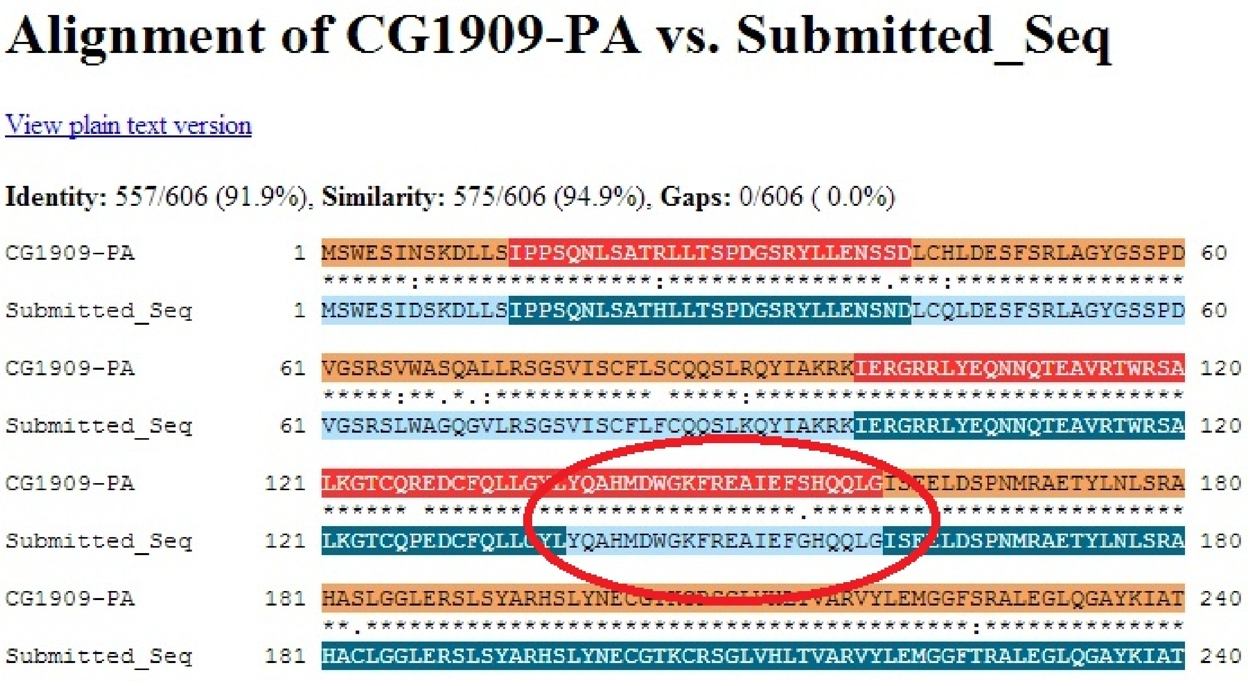
```
Query  31864   IERGRRLYEQNNQTEAVRTIWRSALKGTCQPEDCFQLLGYLYE  31989
               IERGRRLYEQNNQTEAVRTIWRSALKGTCQ EDCFQLLGYLY+
Sbjct  1       IERGRRLYEQNNQTEAVRTIWRSALKGTCQREDCFQLLGYLYQ  42
```

Range 2: 41 to 61 Graphics ▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
48.5 bits(114)	1e-10	Compositional matrix adjust.	20/21(95%)	20/21(95%)	0/21(0%)	+1

```
Query  32293   YQAHMDWGKFREAIEFGHQQL  32355
               YQAHMDWGKFREAIEF HQQL
Sbjct  41      YQAHMDWGKFREAIEFSHQQL  61
```

An image of the alignment with the separated 22 amino acids circled in red is shown below.



Alignment of CG1909-PA vs. Submitted_Seq

View plain text version

Identity: 557/606 (91.9%), **Similarity:** 575/606 (94.9%), **Gaps:** 0/606 (0.0%)

```
CG1909-PA          1 MSWESINSKDLLSIPPSQNLSATRLLTSPDGSRYLLENSSDLCHLDESFSRLAGYGSSPD 60
                     ******:****************:***************.***:****************
Submitted_Seq      1 MSWESIDSKDLLSIPPSQNLSATHLLTSPDGSRYLLENSNDLCQLDESFSRLAGYGSSPD 60

CG1909-PA         61 VGSRSVWASQALLRSGSVISCFLSCQQSLRQYIAKRKIERGRRLYEQNNQTEAVRTWRSA 120
                     *****:**.*.:*********** *****:******************************
Submitted_Seq     61 VGSRSLWAGQGVLRSGSVISCFLFCQQSLKQYIAKRKIERGRRLYEQNNQTEAVRTWRSA 120

CG1909-PA        121 LKGTCQREDCFQLLGYLYQAHMDWGKFREAIEFSHQQLGISFELDSPNMRAETYLNLSRA 180
                     ****** **************************.******** *****************
Submitted_Seq    121 LKGTCQPEDCFQLLGYLYQAHMDWGKFREAIEFGHQQLGISF[illegible]LDSPNMRAETYLNLSRA 180

CG1909-PA        181 HASLGGLERSLSYARHSLYNECGTKCRSGLVHLTVARVYLEMGGFSRALEGLQGAYKIAT 240
                     **.******************************************:**************
Submitted_Seq    181 HACLGGLERSLSYARHSLYNECGTKCRSGLVHLTVARVYLEMGGFTRALEGLQGAYKIAT 240
```

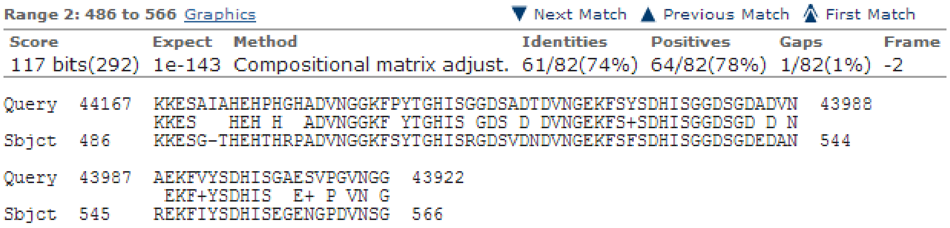
Gene report form

Gene name (i.e. *D. mojavensis eyeless*): __*D. biarmipes onecut*___________
Gene symbol (i.e. dmoj_ey): ____dbia_onecut____
Approximate location in project (from 5’ end to 3’ end): __44645-44211__________
Number of isoforms in *D. melanogaster:* _2_______________
Number of isoforms in this project: Unable to determine due to possible sequencing error__

The gene ortholog for onecut appears to be present on contig34. This gene, however, will not be included in the completed model due to an inability to complete an accurate gene model for this gene.

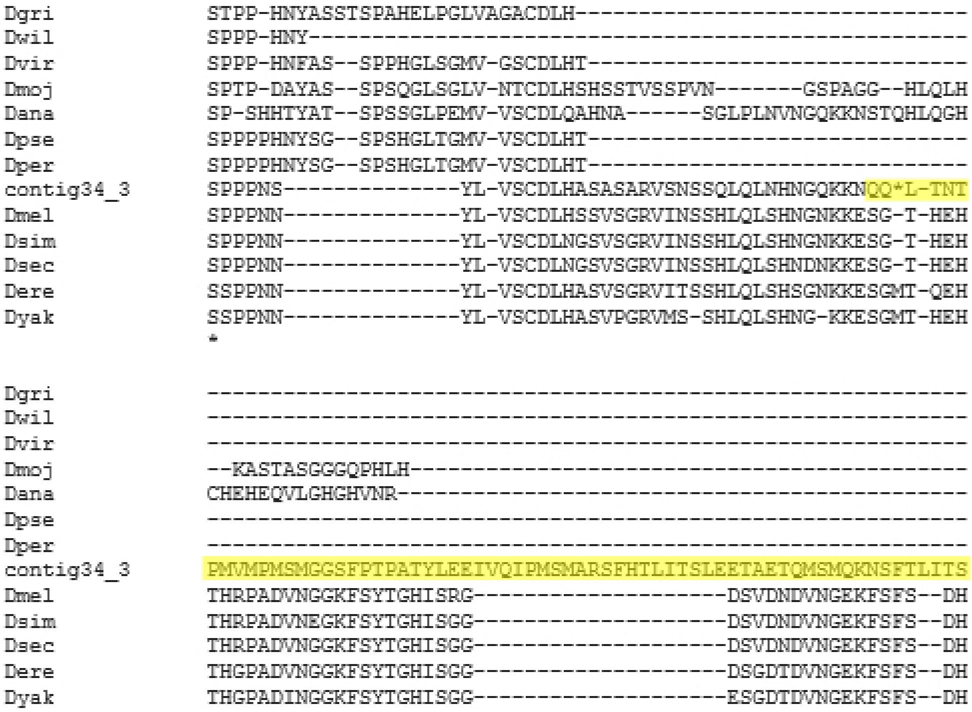
The first exon of *oncecut*, CDS 5_1629_0 shifts into a different frame mid exon. There are no acceptable splices in the region where the frame shifts to account for this frame shift and the region is not large enough to become a legitimate intron. The gene model does not work without the entire exon.

```
Range 1: 1 to 487 Graphics                                    ▼ Next Match  ▲ Previous Match
Score            Expect   Method                          Identities     Positives      Gaps         Frame
404 bits(1037)   1e-143   Compositional matrix adjust.    333/506(66%)   392/506(77%)   30/506(5%)   -3

Query  45645  MESISEIIDHQTFSQDLVEDATDFISVGHNSERQSHPHHEHQEGEPDSGEDLAISLQTMT  45466
              M+S+++IIDHQTFSQ+LVEDA++FI+VGH+SER S    +     P+SG+DL +S+Q + 
Sbjct  1      MDSLNDIIDHQTFSQELVEDASEFITVGHHSERPSQSSQQ-----PNSGQDLTMSMQDII  55

Query  45465  SCPRTDGSSGGKKHRPgsgsgsglvsgsgsdsvvmvidsmGHSNRQTTFQIVSQQMHPR-  45289
              SCP         KHR  + S SG  S SGSDSVVMVID++G  NRQ+ +QIV QQ+  R 
Sbjct  56     SCP--------VKHR--TCSASGSGSASGSDSVVMVIDALGQGNRQSAYQIVPQQLQQRN  105

Query  45288  MPLPFGLLqqnsqntqqcpQRIVSSSPVDFVASDISLDGLTVDS-MSQSVLSQEMAIKQE  45112
              MPLPFGLL+++ Q+ Q    R V++SPVDFV+SDI+LDGLTVD+ +SQ+  SQE A+KQE
Sbjct  106    MPLPFGLLERDRQHMQHG--REVNTSPVDFVSSDINLDGLTVDADVSQTDHSQETAVKQE  163

Query  45111  QKLLIAQSKSVDQGHKRIRMHVEVANVNASLGvdvdemddissddvgcdDEGITLNqhhq  44932
              QKLLI QSKS DQ H+RIRM V+V++VN+ LGV VD+MD+ISSD VGCDDEG+TL+  H 
Sbjct  164    QKLLIVQSKSQDQSHRRIRMLVDVSSVNSGLGVHVDDMDEISSDGVGCDDEGVTLSHQHL  223

Query  44931  qlleqqqQYGLTshhphhqahsqalHGLHHRSAQLEMGLEADHGEVLSVIVHSQDSdkev  44752
                 E+Q  +GLTSHHPH Q H+Q +HGLH RS   EMGL+  HGEVLSVIVHSQDSDKE 
Sbjct  224    LEQEEQ--FGLTSHHPHLQPHTQIIHGLHQRSTHSEMGLDNGHGEVLSVIVHSQDSDKED  281

Query  44751  dgegdadgegdg-dpeddddrDSRSRGQMLSHSSYQTLTSVNDRLSPPGFSQTSYATLTP  44575
                E D        + EDDD+RDSRSR Q+LSHSSYQTLTSVNDRLS PGFSQTSYATLTP
Sbjct  282    CEENDDGDAEGDLENEDDDERDSRSREQLLSHSSYQTLTSVNDRLSSPGFSQTSYATLTP  341

Query  44574  IQPLPPISTMSEKFAYSGHISGGDSgdadvnggegvggvvdggeVTNHSTDAAGTGLIIS  44395
              IQPLPPISTMSEKFAYSGHISGGDSGD DVNG    GGVV+ GEVTN S++A GT  I S
Sbjct  342    IQPLPPISTMSEKFAYSGHISGGDSGDTDVNGDGAGGGVVEVGEVTNQSSEATGTVSISS  401

Query  44394  GNPAS--------SSFPALTMPMgsghlslgvlsgvqSPYSSYEKLSSMISPPPNSYLVS  44239
              GN  S        SSF AL+MP+GSGHL LGVLSGVQSP+SSYEKLSSMISPPPN+YLVS
Sbjct  402    GNATSSVCSNNDCSSFSALSMPIGSGHLGLGVLSGVQSPFSSYEKLSSMISPPPNNYLVS  461

Query  44238  CDLHASASARVSNSSQLQLNHNGQKK  44161
              CDLH+S S RV NSS LQL+HNG KK
Sbjct  462    CDLHSSVSGRVINSSHLQLSHNGNKK  487
```



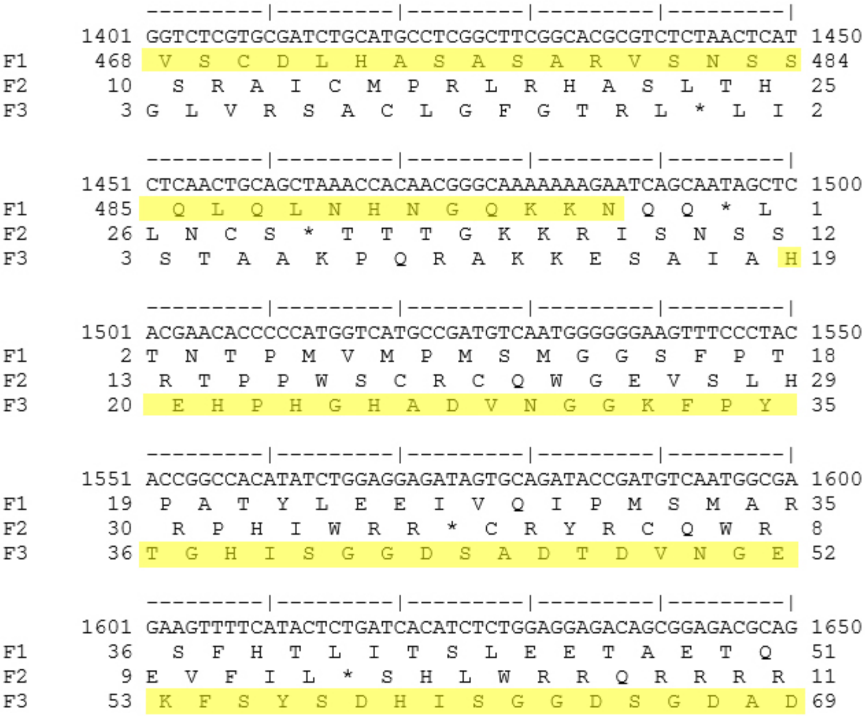
```
Range 2: 486 to 566 Graphics                    ▼ Next Match  ▲ Previous Match  ⩓ First Match
Score           Expect   Method                          Identities   Positives    Gaps        Frame
117 bits(292)   1e-143   Compositional matrix adjust.    61/82(74%)   64/82(78%)   1/82(1%)    -2

Query  44167  KKESAIAHEHPHGHADVNGGKFPYTGHISGGDSADTDVNGEKFSYSDHISGGDSGDADVN  43988
              KKES   HEH H  ADVNGGKF YTGHIS GDS D DVNGEKFS+SDHISGGDSGD D N
Sbjct  486    KKESG-THEHTHRPADVNGGKFSYTGHISRGDSVDNDVNGEKFSFSDHISGGDSGDEDAN  544

Query  43987  AEKFVYSDHISGAESVPGVNGG  43922
               EKF+YSDHIS  E+ P VN G
Sbjct  545    REKFIYSDHISEGENGPDVNSG  566
```

Conservation of the gene and the protein sequence is strong amongst closely related species as indicated by the below multiple alignment. The alignment on *D. biarmipes* is lost in this region and does not align to any other species.



```
Dgri            STPP-HNYASSTSPAHELPGLVAGACDLH-------------------------------
Dwil            SPPP-HNY----------------------------------------------------
Dvir            SPPP-HNFAS--SPPHGLSGMV-GSCDLHT------------------------------
Dmoj            SPTP-DAYAS--SPSQGLSGLV-NTCDLHSHSSTVSSPVN-------GSPAGG--HLQLH
Dana            SP-SHHTYAT--SPSSGLPEMV-VSCDLQAHNA------SGLPLNVNGQKKNSTQHLQGH
Dpse            SPPPPHNYSG--SPSHGLTGMV-VSCDLHT------------------------------
Dper            SPPPPHNYSG--SPSHGLTGMV-VSCDLHT------------------------------
contig34_3      SPPPNS--------------YL-VSCDLHASASARVSNSSQLQLNHNGQKKNQQ*L-TNT
Dmel            SPPPNN--------------YL-VSCDLHSSVSGRVINSSHLQLSHNGNKKESG-T-HEH
Dsim            SPPPNN--------------YL-VSCDLNGSVSGRVINSSHLQLSHNGNKKESG-T-HEH
Dsec            SPPPNN--------------YL-VSCDLNGSVSGRVINSSHLQLSHNDNKKESG-T-HEH
Dere            SSPPNN--------------YL-VSCDLHASVSGRVITSSHLQLSHSGNKKESGMT-QEH
Dyak            SSPPNN--------------YL-VSCDLHASVPGRVMS-SHLQLSHNG-KKESGMT-HEH
                *

Dgri            ------------------------------------------------------------
Dwil            ------------------------------------------------------------
Dvir            ------------------------------------------------------------
Dmoj            --KASTASGGGQPHLH--------------------------------------------
Dana            CHEHEQVLGHGHVNR---------------------------------------------
Dpse            ------------------------------------------------------------
Dper            ------------------------------------------------------------
contig34_3      PMVMPMSMGGSFPTPATYLEEIVQIPMSMARSFHTLITSLEETAETQMSMQKNSFTLITS
Dmel            THRPADVNGGKFSYTGHISRG--------------------DSVDNDVNGEKFSFS--DH
Dsim            THRPADVNEGKFSYTGHISGG--------------------DSVDNDVNGEKFSFS--DH
Dsec            THRPADVNGGKFSYTGHISGG--------------------DSVDNDVNGEKFSFS--DH
Dere            THGPADVNGGKFSYTGHISGG--------------------DSGDTDVNGEKFSFS--DH
Dyak            THGPADINGGKFSYTGHISGG--------------------ESGDTDVNGEKFSFS--DH
```

When looking at the proteins in each reading frame in this region more closely, the location of the shift and the possible error is visible.



```
           ---------|---------|---------|---------|---------|
      1401 GGTCTCGTGCGATCTGCATGCCTCGGCTTCGGCACGCGTCTCTAACTCAT 1450
F1     468  V  S  C  D  L  H  A  S  A  S  A  R  V  S  N  S  S 484
F2      10   S  R  A  I  C  M  P  R  L  R  H  A  S  L  T  H   25
F3       3 G  L  V  R  S  A  C  L  G  F  G  T  R  L  *  L  I  2

           ---------|---------|---------|---------|---------|
      1451 CTCAACTGCAGCTAAACCACAACGGGCAAAAAAAGAATCAGCAATAGCTC 1500
F1     485   Q  L  Q  L  N  H  N  G  Q  K  K  N  Q  Q  *  L   1
F2      26 L  N  C  S  *  T  T  T  G  K  K  R  I  S  N  S  S  12
F3       3  S  T  A  A  K  P  Q  R  A  K  K  E  S  A  I  A  H 19

           ---------|---------|---------|---------|---------|
      1501 ACGAACACCCCCATGGTCATGCCGATGTCAATGGGGGGAAGTTTCCCTAC 1550
F1       2 T  N  T  P  M  V  M  P  M  S  M  G  G  S  F  P  T  18
F2      13  R  T  P  P  W  S  C  R  C  Q  W  G  E  V  S  L  H 29
F3      20   E  H  P  H  G  H  A  D  V  N  G  G  K  F  P  Y   35

           ---------|---------|---------|---------|---------|
      1551 ACCGGCCACATATCTGGAGGAGATAGTGCAGATACCGATGTCAATGGCGA 1600
F1      19  P  A  T  Y  L  E  E  I  V  Q  I  P  M  S  M  A  R 35
F2      30   R  P  H  I  W  R  R  *  C  R  Y  R  C  Q  W  R   8
F3      36 T  G  H  I  S  G  G  D  S  A  D  T  D  V  N  G  E  52

           ---------|---------|---------|---------|---------|
      1601 GAAGTTTTCATACTCTGATCACATCTCTGGAGGAGACAGCGGAGACGCAG 1650
F1      36   S  F  H  T  L  I  T  S  L  E  E  T  A  E  T  Q   51
F2       9 E  V  F  I  L  *  S  H  L  W  R  R  Q  R  R  R  R  11
F3      53  K  F  S  Y  S  D  H  I  S  G  G  D  S  G  D  A  D 69
```

All of this evidence points to a possible sequencing error in this gene.

I was able to locate and construct the model of the remaining exons with no issues and they aligned well.
CDS 5_1629_0: Unable to model
CDS 3_1629_1: (43235-42313)
CDS 2_1629_2: (41459-41281)
CDS 1_1629_0: (39592-39188)
Stop: (39187-39185)

Final Annotation Files Merger Genome Browser View

GFF, FASTA, and PEP merged files are included in submission. Files and screenshot exclude the gene *onecut* due to the issues previously discussed in that genes section of this report.