GEP Annotation Report

Note: For each gene described in this annotation report, you should also prepare the corresponding GFF, transcript and peptide sequence files as part of your submission.

Student name: Peter Selim Siyahhan Julnes
Student email: pjulnes@unm.edu
Faculty Advisor: Dr. Paul Szauter
College/University: University of New Mexico

Project details

Project name: contig29
Project species: *D. Biarmipes*
Date of submission: 7/15/13
Size of project in base pairs: 50,000
Number of genes in project: 1

Does this report cover all genes and all isoforms or is it a partial report? All
If this is a partial report because different students are working on different regions of this sequence, please report the region of the project covered by this report:
from base _____ to base _____

Instructions for project with no genes

If you believe that the project does not contain any genes, please provide the following evidence to support your conclusions:

1. Perform a BLASTX search of the entire contig sequence against the non-redundant (*nr*) protein database. Provide an explanation for any significant (E-value < 1e-5) hits to known genes in the nr database as to why they do not correspond to real genes in the project.

2. For each Genscan prediction, perform a BLASTP search using the predicted amino acid sequence against the protein database (*nr*) using the strategy described above.

3. Examine the gene expression tracks (e.g. cDNA/EST/RNA-Seq) for evidence of transcribed regions that do not correspond to alignments to known *D. melanogaster* proteins. Perform a BLASTX search against the *nr* database using these genomic regions to determine if the region is similar to any known or predicted proteins in the *nr* database.

Complete the following Gene Report Form for each gene in your project. Copy and paste the sections below to create as many copies as needed. Be sure to create enough Isoform Report Forms within your Gene Report Form for all isoforms.

Gene report form

Gene name (i.e. *D. mojavensis eyeless*): Asator
Gene symbol (i.e. dmoj_ey): Asator
Approximate location in project (from 5' end to 3' end): 21100-290
Number of isoforms in *D. melanogaster:* 8
Number of isoforms in this project: 8

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
Asator-PJ	Asator-PI
Asator-PH	Asator-PK
Asator-PD	Asator-PE
Asator-PG	
Asator-PF	

Note: For isoforms with identical coding sequence, you only need to complete the Isoform Report Form for one of these isoforms (i.e. using the name of the isoform listed in the left column of the table above). However, you should generate GFF, transcript, and peptide sequence files for ALL isoforms, irrespective of whether they have identical coding sequences as other isoforms.

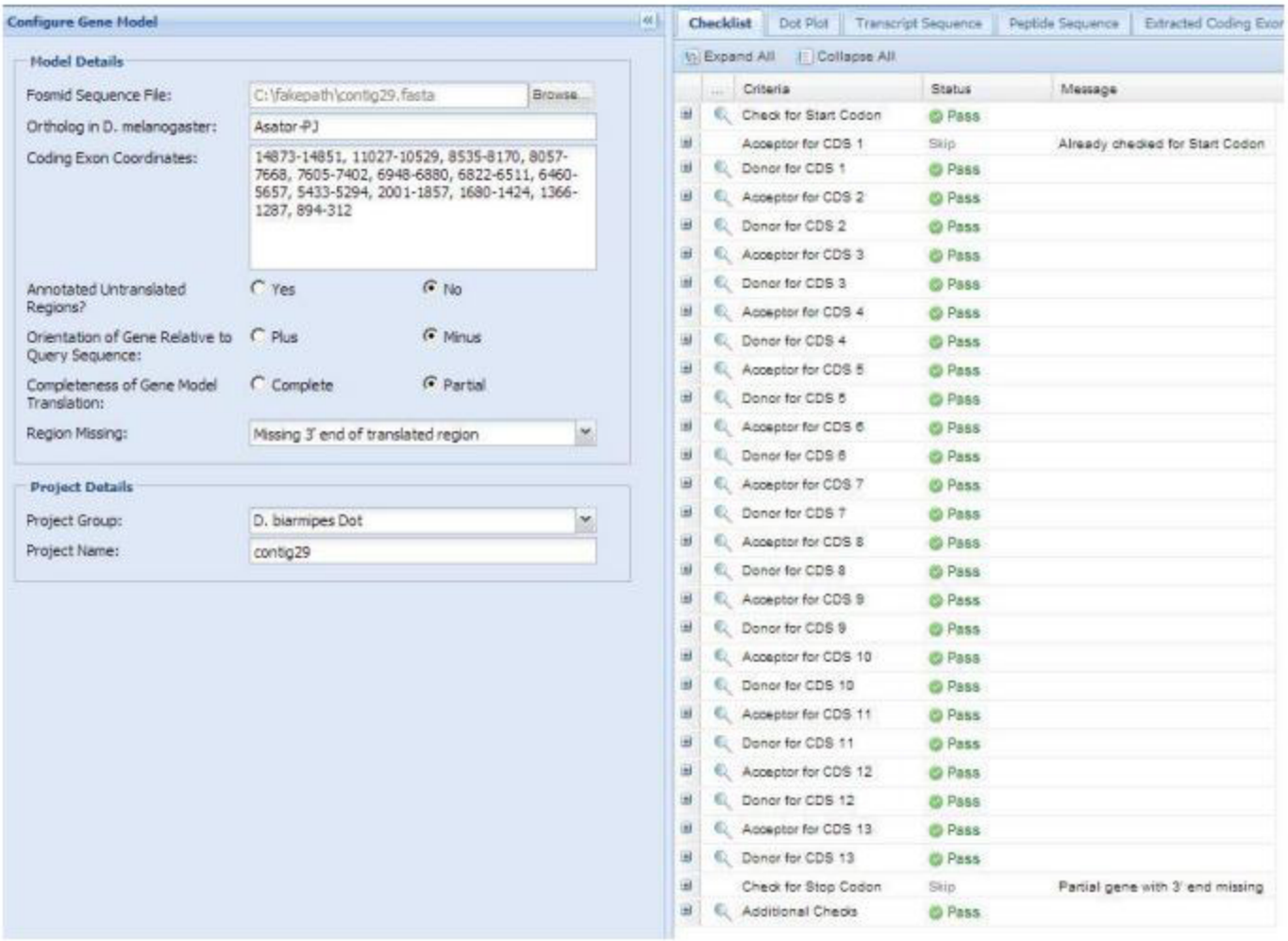
Isoform report form

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): Asator-PJ
Names of the isoforms with identical coding sequences as this isoform
Asator-PI
Is the 5' end of this isoform missing from the end of project: Yes
If so, how many exons are missing from the 5' end: 2
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: _____

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:**

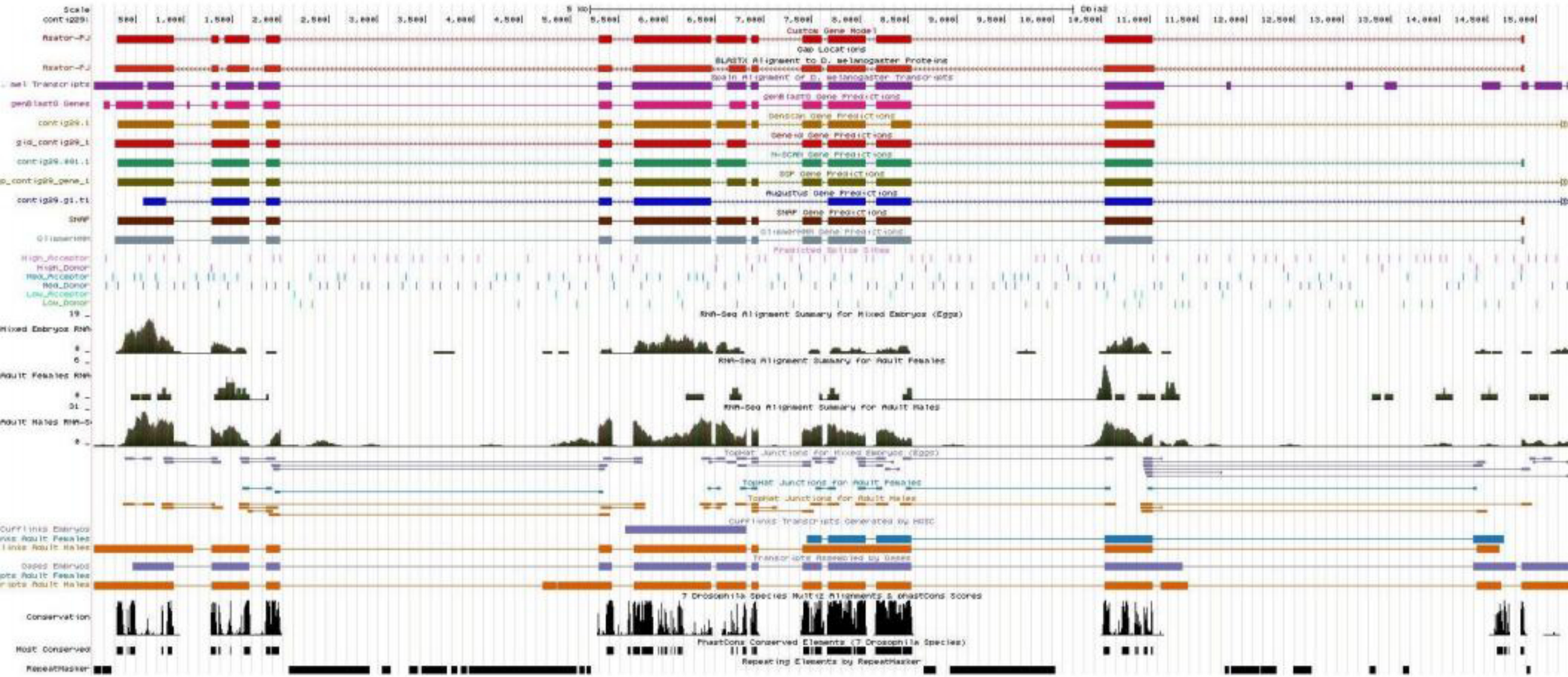


2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker (see page 10 of the Gene Model Checker user guide on how to do this; you can find the guide under "Help" -> "Documentations" -> "Web Framework" on the GEP website at http://gep.wustl.edu). Capture a screenshot of your gene model shown on the Genome Browser for your project; zoom in so that only this isoform is in the screenshot. Include the following evidence tracks in the screenshot if they are available.

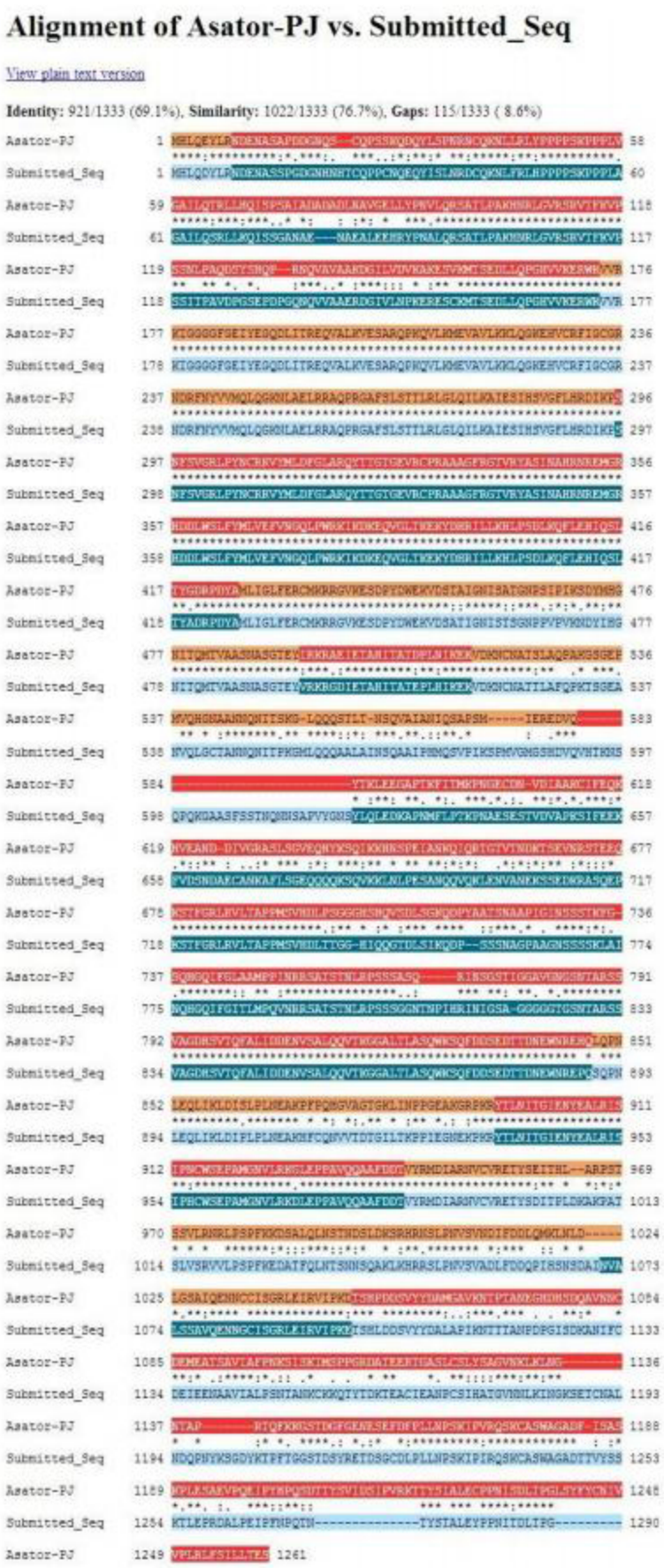
1. A sequence alignment track (D. mel Protein or Other RefSeq)
2. At least one gene prediction track (e.g. Genscan)
3. At least one RNA-Seq track (e.g. RNA-Seq Alignment Summary)
4. A comparative genomics track
   (e.g. Conservation, D. mel. Net Alignment, 3-way, 5-way or 7-way multiz)

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


Custom Gene Model
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
genBlastG Gene Predictions
Genscan Gene Predictions
GeneID Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
Glimmer/GeneMark Gene Predictions
Predicted Splice Sites
RNA-Seq Alignment Summary for Mixed Embryos (Eggs)
RNA-Seq Alignment Summary for Adult Females
RNA-Seq Alignment Summary for Adult Males
TopHat Junctions for Mixed Embryos (Eggs)
TopHat Junctions for Adult Females
TopHat Junctions for Adult Males
Cufflinks Transcripts Generated by Male
Transcripts Assembled by Genes
7 Drosophila Species Multiz Alignments & phastCons Scores
phastCons Conserved Elements (7 Drosophila Species)
Repeating Elements by RepeatMasker

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

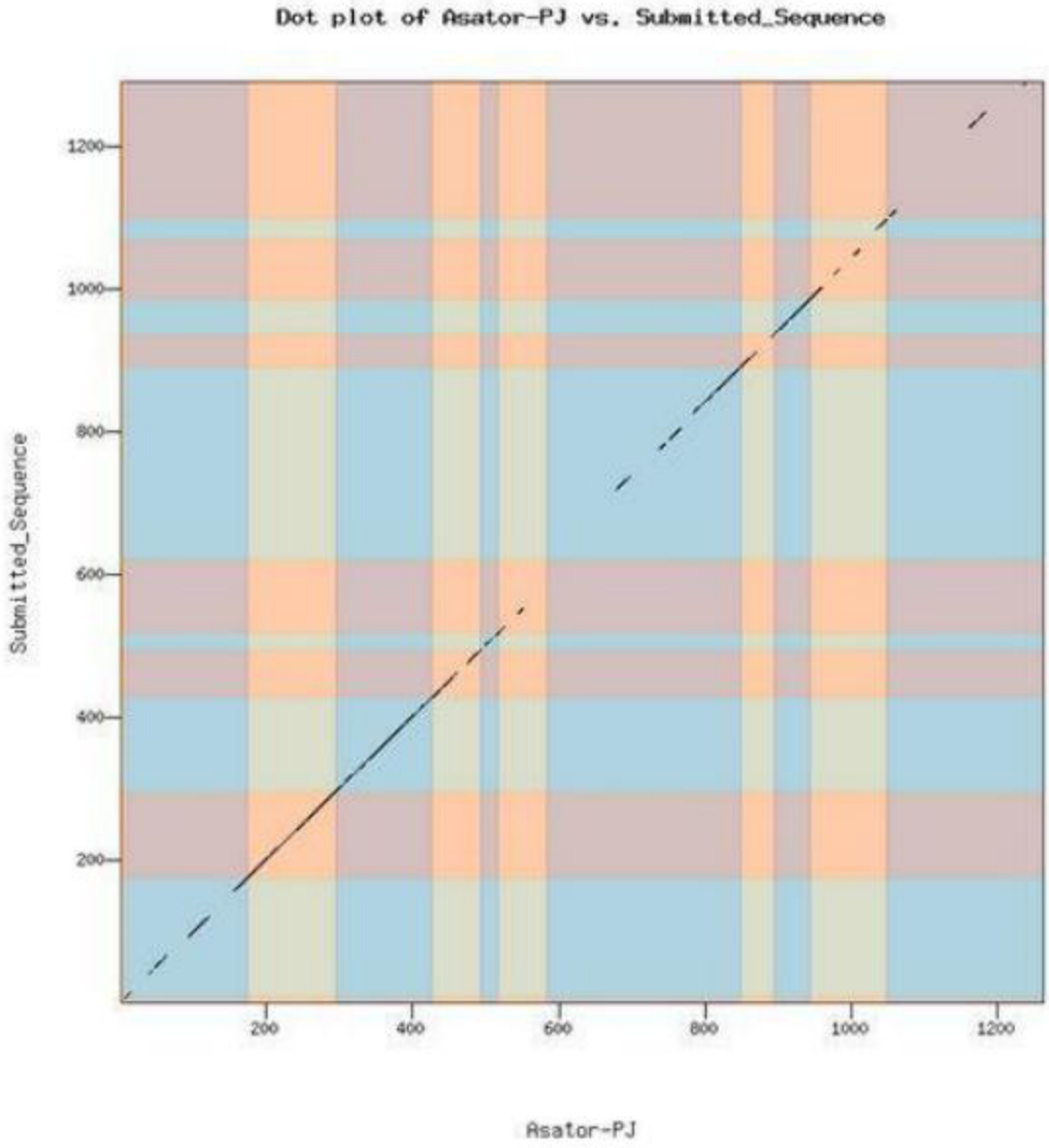
Show an alignment between the protein sequence for your gene model and the protein sequence from the putative *D. melanogaster* ortholog. You can use the protein alignment generated by the Gene Model Checker or you can generate a new alignment using BLAST 2 Sequences (*bl2seq*). **Copy and paste the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



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).

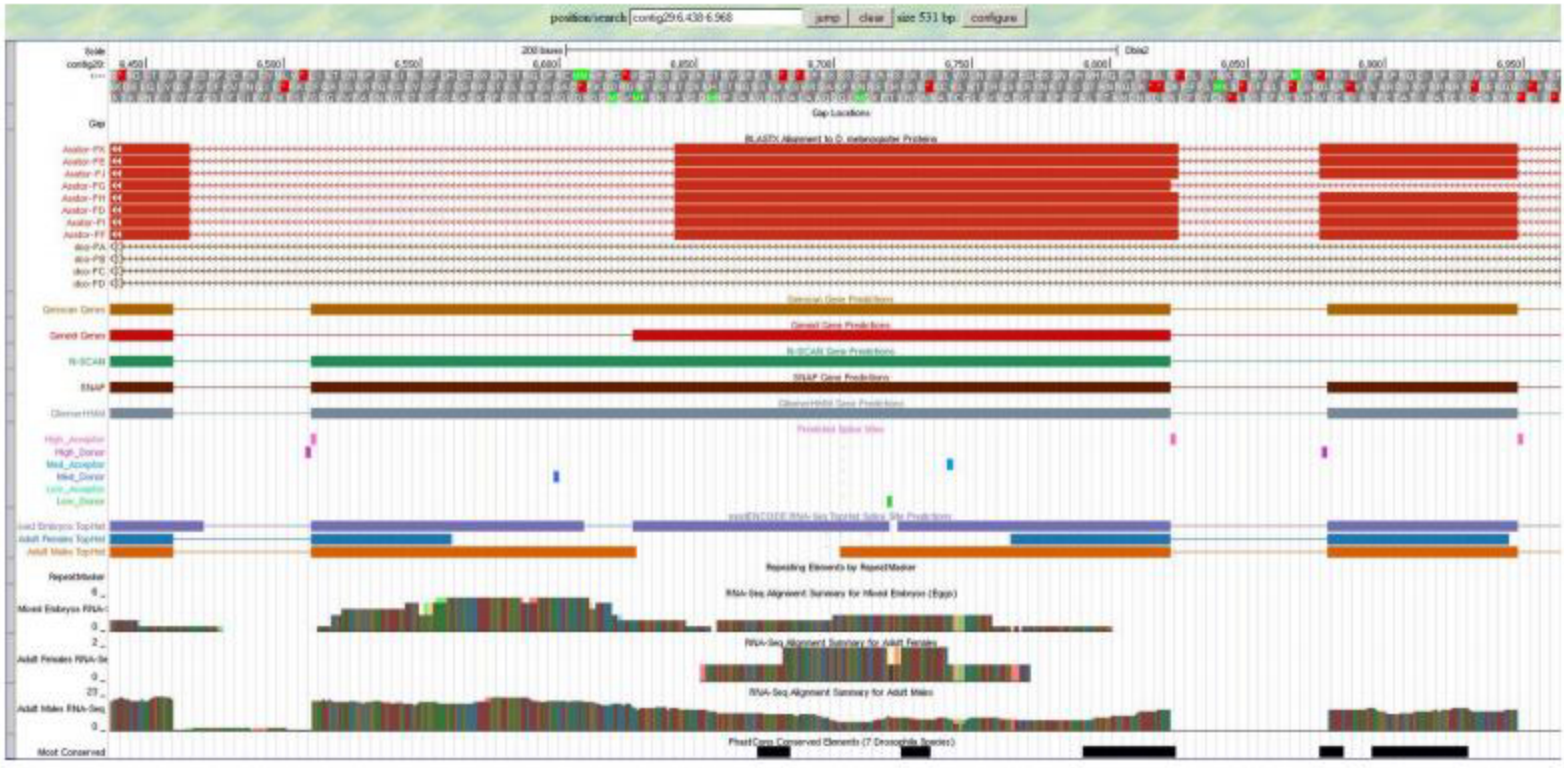
Note: Large <u>vertical and horizontal gap</u> near exon boundaries in the dot plot often indicates that an incorrect splice site might have been picked. Please re-examine these regions and provide a detail justification as to why you have selected this particular set of donor and acceptor sites.



The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in D. melanogaster. An open reading frame the entire way through is present as well as computational evidence of tophat junctions and RNA Seq and a high donor in support of the new splice site.

```
D.Wil       PVDSVLQNVNSGINQNPLPKT---------------LAQ------QQLAAANPNLPSALI
D.Moja      SADCVVQNVNLGNPNNLLKHQQQQQQQQQQQQQQQQQQQQQQQLVLVSNTSY----------
D.Viri      TADCAVQNVNSGI--QNT-----------TLPKQQHQQQQQLVLVANTSIAIPNLPSALI
D.Grim      TAESALQNSGTLNALPKQ-----------QQQQLQQQQQQQLVLVSNTAIAVANLPSALI
D.Anan      LVEAASQYG--INNQNNLAKELLQQ------------QKSCECL--------------L
D.pseud     KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Pers      KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Erec      SGDAIVPHGNTANNQNITTKRLQQQ------------STLT--TNSQV--AIPNIQV---
D.Yak       SGEATIQHGNTANNQNIKTKGLQQQ------------STMT--TNSQV--AIPNIQSAPI
D.mel       SGEPMVQHGNAANNQNITSKGLQQQ------------STL---TNSQV--AIANIQSAP-
D.Sim       SGEALIQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
D.          SGEALTQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
                                :

D.Wil       KSNHENPAAI------VGQIKSLHGHTAIHTKKCQPHVGTGSLSISNQNNSAPVYQYPHA
D.Moja      ------------------------------------------------------------PES
D.Viri      KSST-VGIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNNSAPVY-YPES
D.Grim      KPSTVTGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQYPES
D.Anan      VESML-------------------------------------------------------
D.pseud     KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Pers      KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Erec      ----H---QREDVQ----------------------------------------------YTKL
D.Yak       KSGMT---EREDVQ----------------------------------------------YTKM
D.mel       --SMI---EREDVQ----------------------------------------------YTKL
D.Sim       KSSMI---EREDVQ----------------------------------------------YTKL
D.          KSSMI---EREDVQ----------------------------------------------YTKL
```

Marked above is the region in question in a multiple alignment and the region is shown below in the genome browser.



A loss of intron is also present between the ortholog in melanogaster and the 3' end of the exon in biarmipes (11050-10529).

```
D.Wil    ------------------------------------------------------------
D.Moja   ------------------------------------------------------------
D.Viri   ------------------------------------------------------------
D.Grim   ------------------------------------------------------------
D.Anan   ------------------------------------------------------------
D.pseud  -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Pers   -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Biar   MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEQYISLNRD---CQKNLF
D.Erec   ------------------------------------------------------------
D.Yak    ------------------------------------------------------------
D.mel    ---------MFWHLLCVP-NDENASAPDDG--NQSCQPSSKQDQYLSPNRN---CQKNLL
D.Sim    ------------------------------------------------------------
```

The poorly conserved region contains a small, 8 amino acid exon in melanogaster which is present on the 3' end of an exon (11050-10529) due to a loss of intron. 4 of the 8 amino acids in melanogaster are present on the 3' end of this exon in the order which they were present on melanogaster. There is an open reading frame all the way through and a translated nucleotide blast of melanogaster following a loss of intron vs biarmipes produces the alignment below with a significant e-value of 4*10^-58.

```
Query  1    MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEQYISLNRDCQKNLFRLHppppskpppLA  180
            M + L   NDENAS+P DGN   +CQP   Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct  1    MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPPPPSKPPPLV  58

Query  181  GAILQSRLLKQISSGanaenaealee---HRYPNALQRSATLPAKHNRLGVRSRVTFKVP  351
            GAILQ+RLL QIS  A A+    L       YPN LQRSATLPAKHNRLGVRSRVTFKVP
Sbjct  59   GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNRLGVRSRVTFKVP  118

Query  352  SSITPAVDPGS-EPDPGQNQVVAAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK  522
            SS  PA D  S +P   +NQV  A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct  119  SSNLPAQDSYSHQP---RNQVAVAAKDGILVDVKAKESVKMTSEDLLQPGHVVKERWK  173
```

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): Asator-PH

Names of the isoforms with identical coding sequences as this isoform
Asator-PK

Is the 5' end of this isoform missing from the end of project: Yes

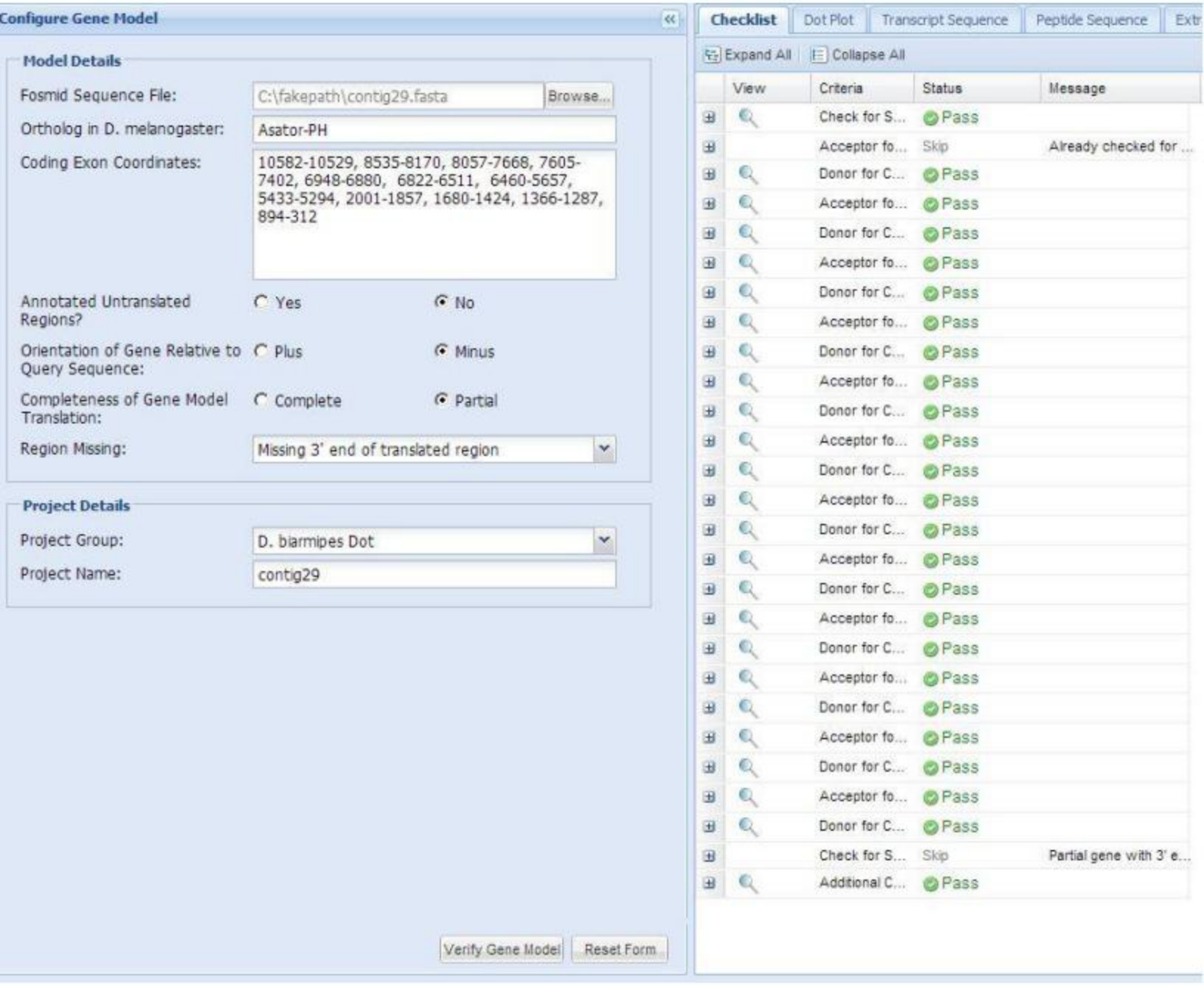
If so, how many exons are missing from the 5' end: 2

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: _____

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:**

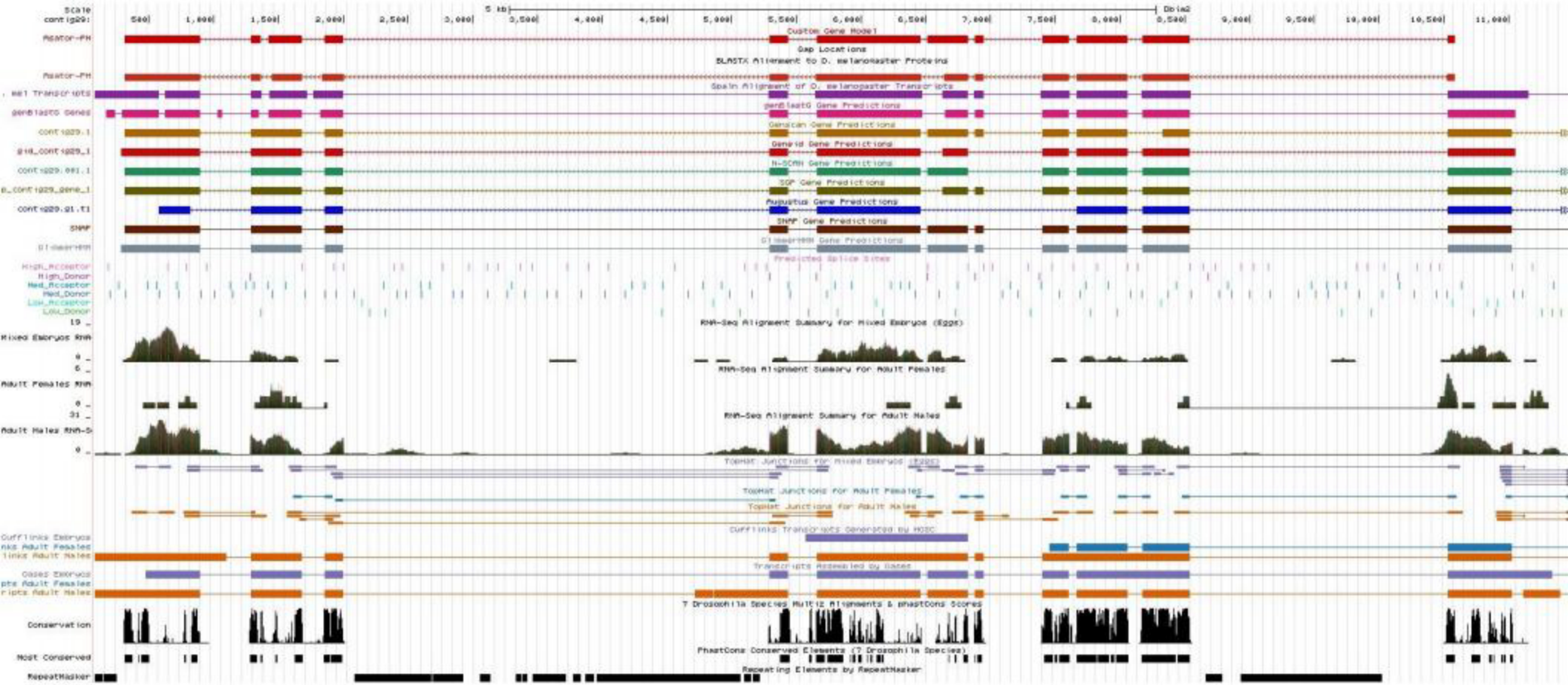


2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker (see page 10 of the Gene Model Checker user guide on how to do this; you can find the guide under "Help" -> "Documentations" -> "Web Framework" on the GEP website at http://gep.wustl.edu). Capture a screenshot of your gene model shown on the Genome Browser for your project; zoom in so that only this isoform is in the screenshot. Include the following evidence tracks in the screenshot if they are available.

1. A sequence alignment track (D. mel Protein or Other RefSeq)
2. At least one gene prediction track (e.g. Genscan)
3. At least one RNA-Seq track (e.g. RNA-Seq Alignment Summary)
4. A comparative genomics track
   (e.g. Conservation, D. mel. Net Alignment, 3-way, 5-way or 7-way multiz)

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


Scale
contig29:
5 kb
Custom Gene Model
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
genBlastG Gene Predictions
Genscan Gene Predictions
Genezid Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
Glimmer3HM Gene Predictions
Predicted Splice Sites
High_Acceptor
High_Donor
Med_Acceptor
Med_Donor
Low_Acceptor
Low_Donor
RNA-Seq Alignment Summary for Mixed Embryos (Eggs)
Mixed Embryos RNA
RNA-Seq Alignment Summary for Adult Females
Adult Females RNA
RNA-Seq Alignment Summary for Adult Males
Adult Males RNA-S
TopHat Junctions for Mixed Embryos (Eggs)
TopHat Junctions for Adult Females
TopHat Junctions for Adult Males
Cufflinks Transcripts Generated by MISC
Transcripts Remodeled by Oases
7 Drosophila Species Multiz Alignments & phastCons Scores
Conservation
PhastCons Conserved Elements (7 Drosophila Species)
Most Conserved
Repeating Elements by RepeatMasker
RepeatMasker

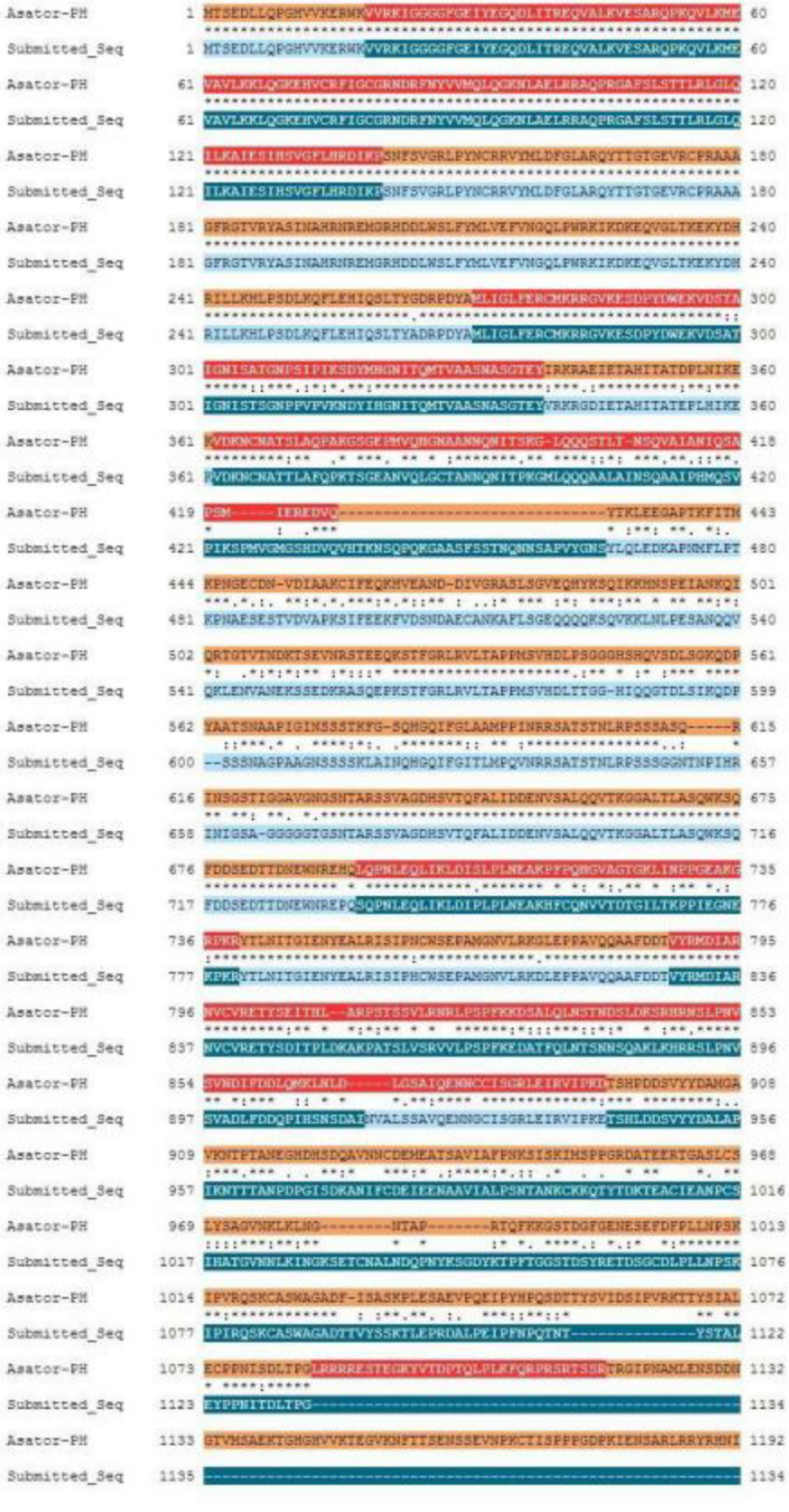
3. Alignment between the submitted model and the *D. melanogaster ortholog*

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 Asator-PH vs. Submitted_Seq

View plain text version

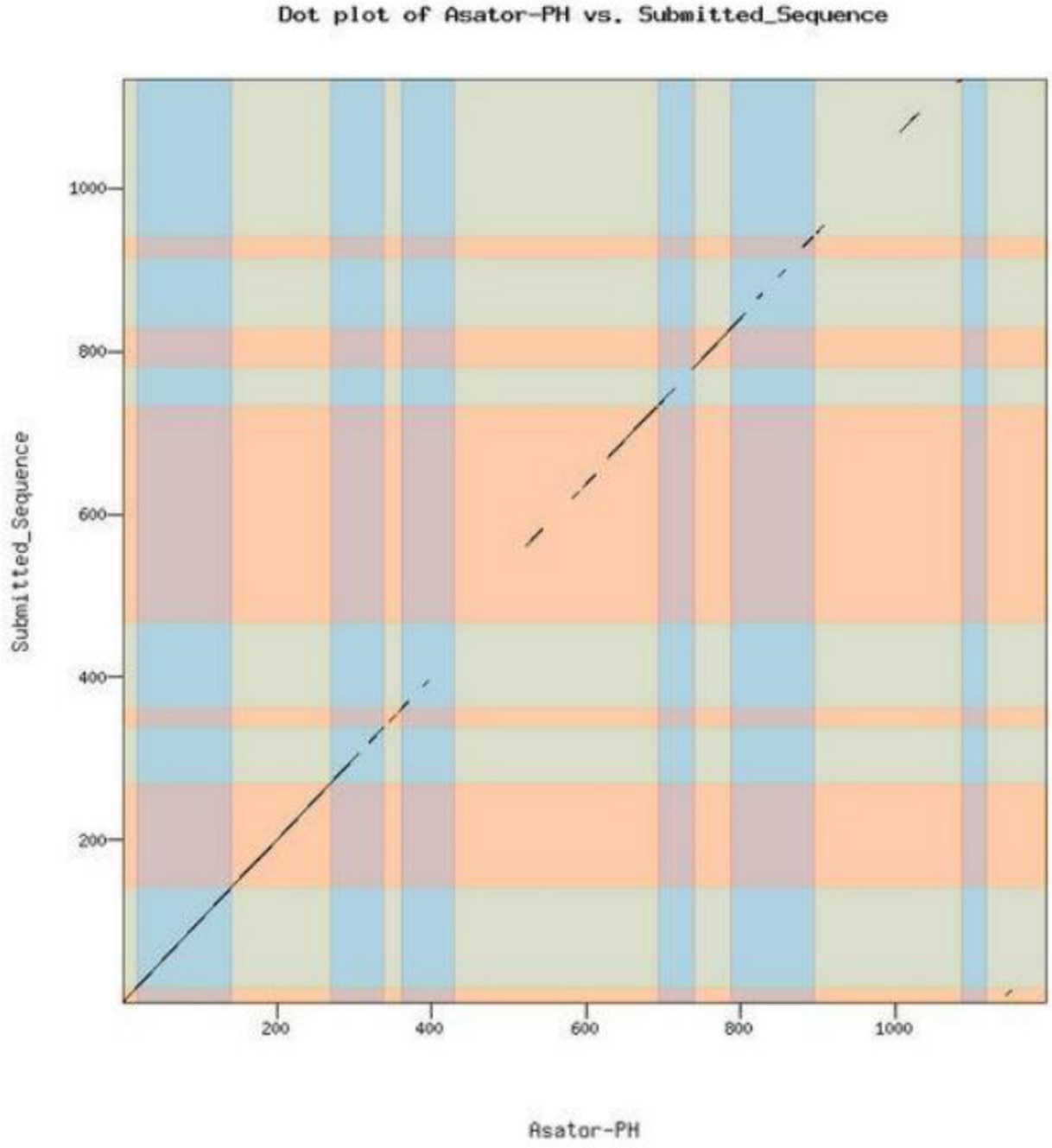
Identity: 818/1261 (64.9%), **Similarity:** 902/1261 (71.5%), **Gaps:** 195/1261 (15.5%)



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).

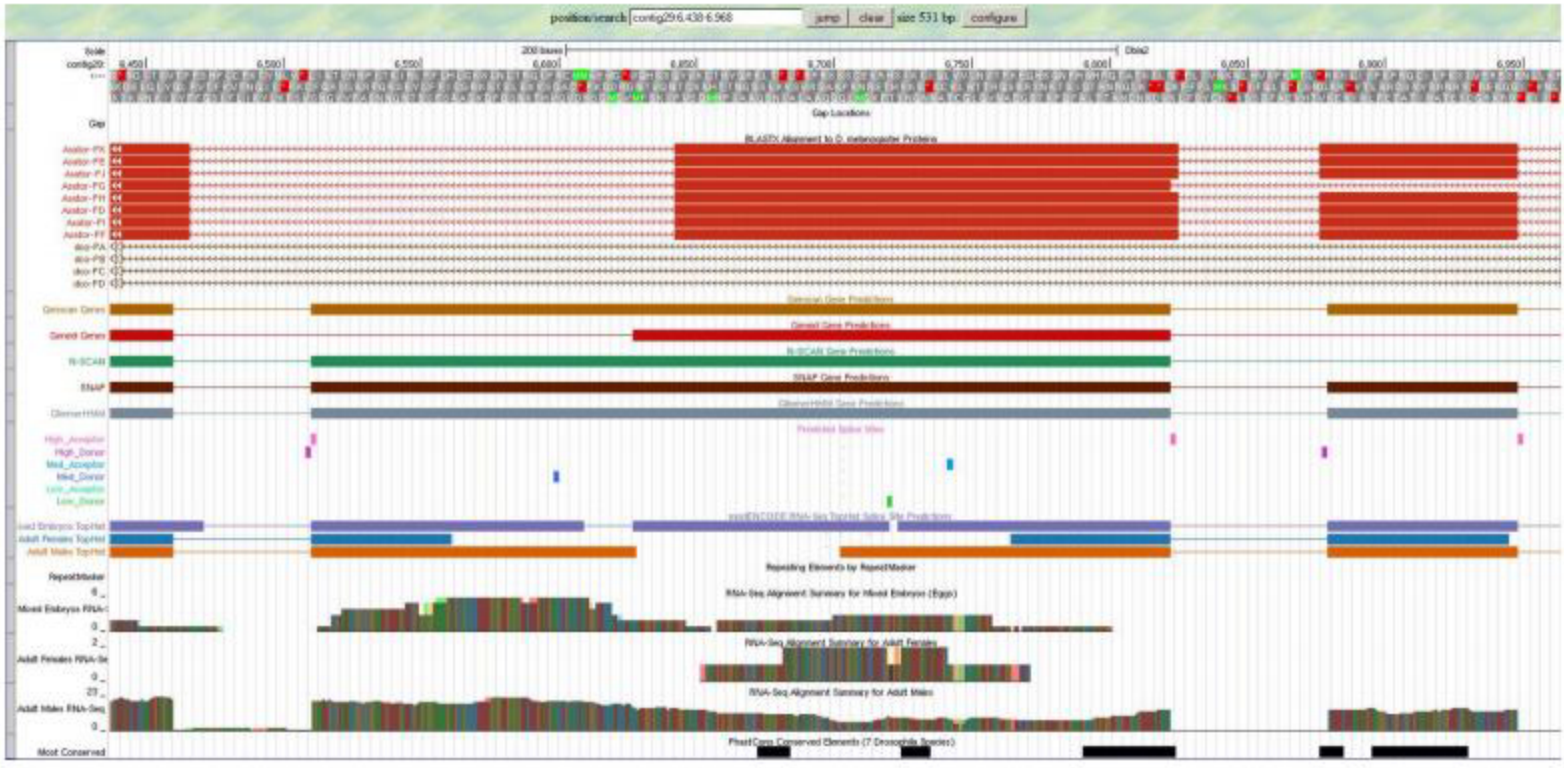
Note: Large vertical and horizontal gap near exon boundaries in the dot plot often indicates that an incorrect splice site might have been picked. Please re-examine these regions and provide a detail justification as to why you have selected this particular set of donor and acceptor sites.



The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in D. melanogaster. An open reading frame the entire way through is present as well as computational evidence of tophat junctions and RNA Seq and a high donor in support of the new splice site.

```
D.Wil       PVDSVLQNVNSGINQNPLPKT---------------LAQ------QQLAAANPNLPSALI
D.Moja      SADCVVQNVNLGNPNNLLKHQQQQQQQQQQQQQQQQQQQQQQQLVLVSNTSY----------
D.Viri      TADCAVQNVNSGI--QNT-----------TLPKQQHQQQQQLVLVANTSIAIPNLPSALI
D.Grim      TAESALQNSGTLNALPKQ-----------QQQQLQQQQQQQLVLVSNTAIAVANLPSALI
D.Anan      LVEAASQYG--INNQNNLAKELLQQ------------QKSCECL--------------L
D.pseud     KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Pers      KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Erec      SGDAIVPHGNTANNQNITTKRLQQQ------------STLT--TNSQV--AIPNIQV---
D.Yak       SGEATIQHGNTANNQNIKTKGLQQQ------------STMT--TNSQV--AIPNIQSAPI
D.mel       SGEPMVQHGNAANNQNITSKGLQQQ------------STL---TNSQV--AIANIQSAP-
D.Sim       SGEALIQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
D.          SGEALTQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
                                :

D.Wil       KSNHENPAAI------VGQIKSLHGHTAIHTKKCQPHVGTGSLSISNQNNSAPVYQYPHA
D.Moja      ------------------------------------------------------------PES
D.Viri      KSST-VGIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNNSAPVY-YPES
D.Grim      KPSTVTGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQYPES
D.Anan      VESML-------------------------------------------------------
D.pseud     KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Pers      KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Erec      ----H---QREDVQ----------------------------------------------YTKL
D.Yak       KSGMT---EREDVQ----------------------------------------------YTKM
D.mel       --SMI---EREDVQ----------------------------------------------YTKL
D.Sim       KSSMI---EREDVQ----------------------------------------------YTKL
D.          KSSMI---EREDVQ----------------------------------------------YTKL
```

Marked above is the region in question in a multiple alignment and the region is shown below in the genome browser.



A loss of intron is also present between the ortholog in melanogaster and the 3' end of the exon in biarmipes (11050-10529).

```
D.Wil    ------------------------------------------------------------
D.Moja   ------------------------------------------------------------
D.Viri   ------------------------------------------------------------
D.Grim   ------------------------------------------------------------
D.Anan   ------------------------------------------------------------
D.pseud  -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Pers   -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Biar   MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEQYISLNRD---CQKNLF
D.Erec   ------------------------------------------------------------
D.Yak    ------------------------------------------------------------
D.mel    ---------MFWHLLCVP-NDENASAPDDG--NQSCQPSSKQDQYLSPNRN---CQKNLL
D.Sim    ------------------------------------------------------------
```

The poorly conserved region contains a small, 8 amino acid exon in melanogaster which is present on the 3' end of an exon (11050-10529) due to a loss of intron. 4 of the 8 amino acids in melanogaster are present on the 3' end of this exon in the order which they were present on melanogaster. There is an open reading frame all the way through and a translated nucleotide blast of melanogaster following a loss of intron vs biarmipes produces the alignment below with a significant e-value of 4*10^-58.

```
Query  1    MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEQYISLNRDCQKNLFRLHppppskpppLA  180
            M + L   NDENAS+P DGN   +CQP   Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct  1    MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPPPPSKPPPLV  58

Query  181  GAILQSRLLKQISSGanaenaealee---HRYPNALQRSATLPAKHNRLGVRSRVTFKVP  351
            GAILQ+RLL QIS  A A+    L       YPN LQRSATLPAKHNRLGVRSRVTFKVP
Sbjct  59   GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNRLGVRSRVTFKVP  118

Query  352  SSITPAVDPGS-EPDPGQNQVVAAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK  522
            SS  PA D  S +P   +NQV  A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct  119  SSNLPAQDSYSHQP---RNQVAVAAKDGILVDVKAKESVKMTSEDLLQPGHVVKERWK  173
```

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): Asator-PD

Names of the isoforms with identical coding sequences as this isoform
Asator-PE

Is the 5' end of this isoform missing from the end of project: Yes

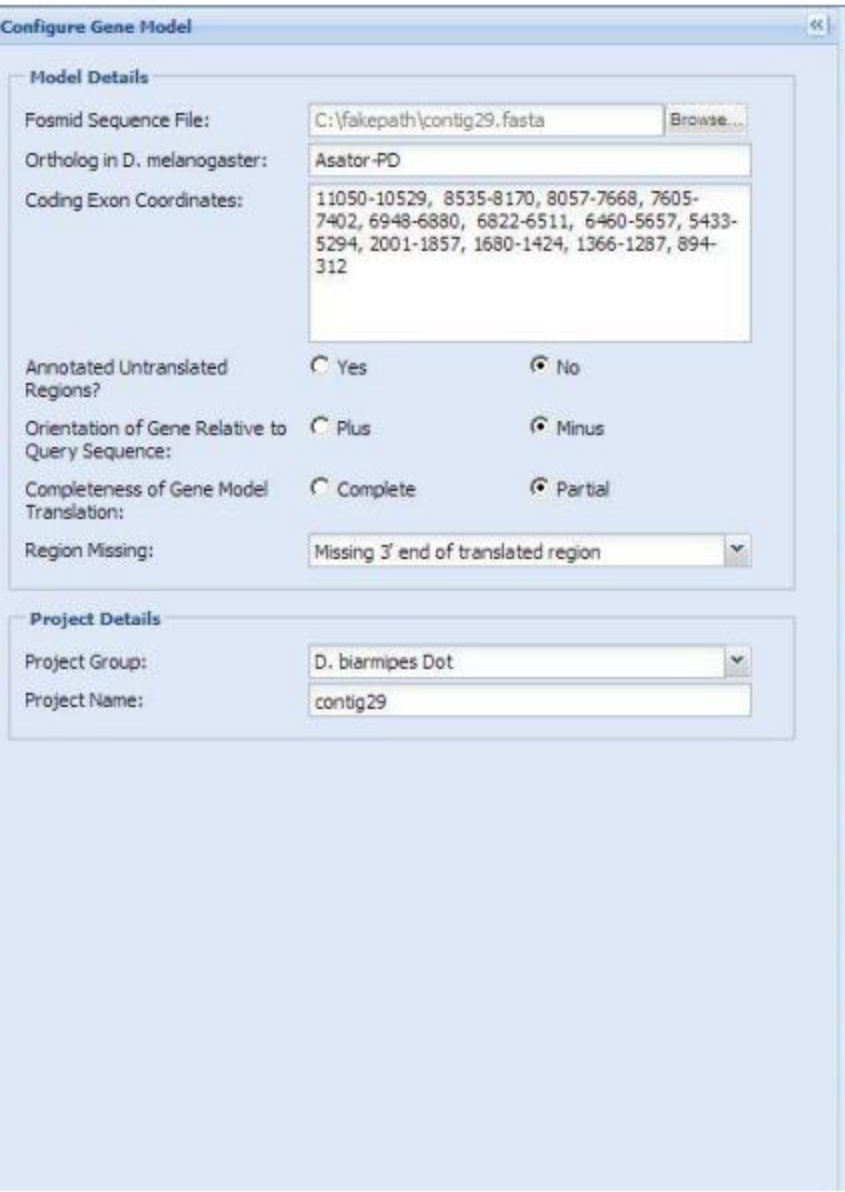
If so, how many exons are missing from the 5' end: 2

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: _____

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:**

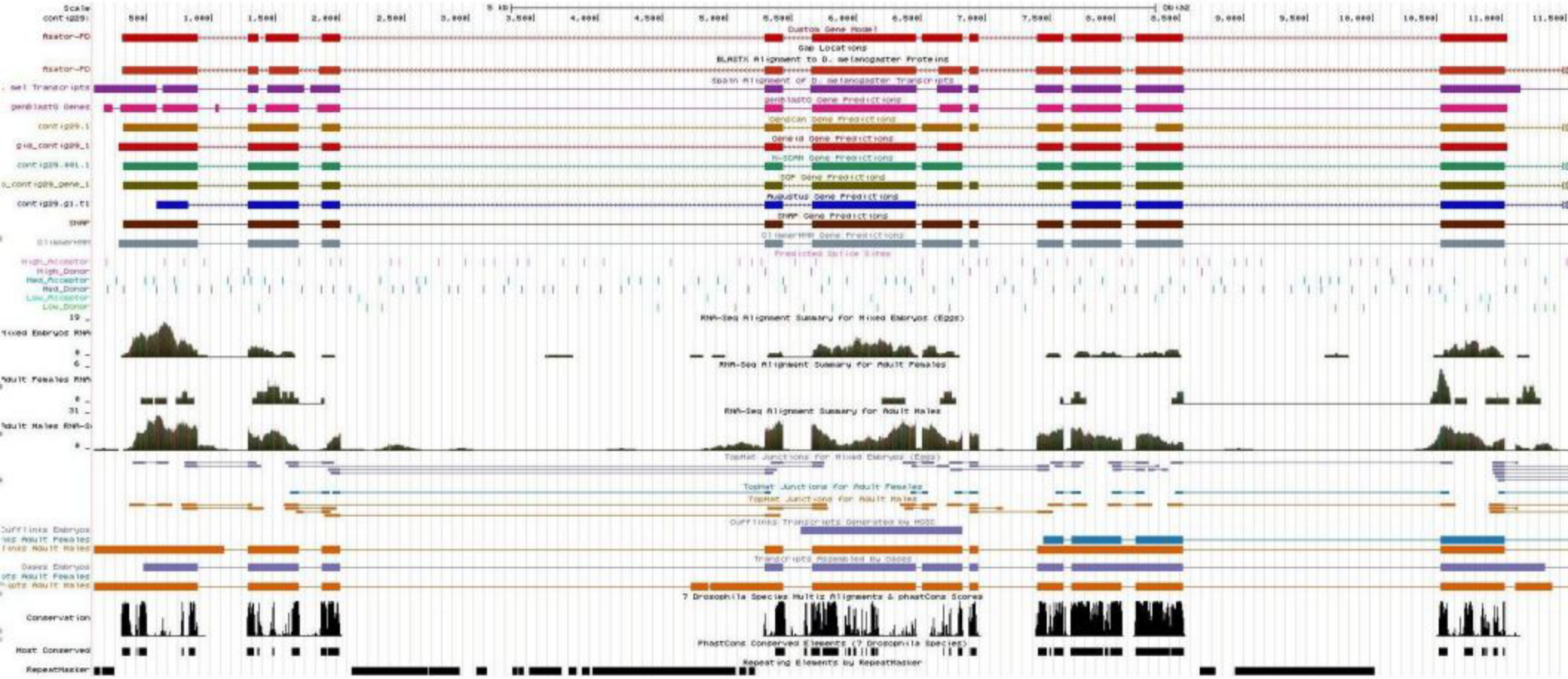


2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker (see page 10 of the Gene Model Checker user guide on how to do this; you can find the guide under "Help" -> "Documentations" -> "Web Framework" on the GEP website at http://gep.wustl.edu). Capture a screenshot of your gene model shown on the Genome Browser for your project; zoom in so that only this isoform is in the screenshot. Include the following evidence tracks in the screenshot if they are available.

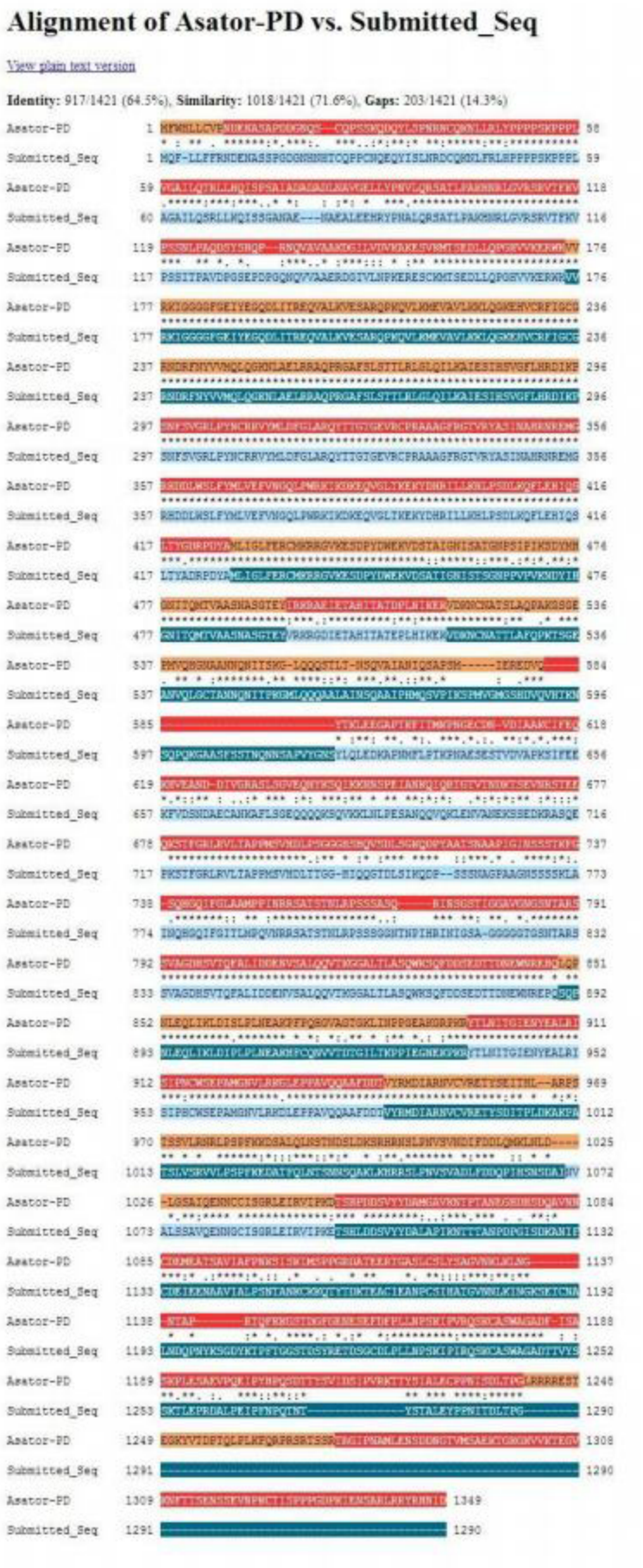
1. A sequence alignment track (D. mel Protein or Other RefSeq)
2. At least one gene prediction track (e.g. Genscan)
3. At least one RNA-Seq track (e.g. RNA-Seq Alignment Summary)
4. A comparative genomics track
   (e.g. Conservation, D. mel. Net Alignment, 3-way, 5-way or 7-way multiz)

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


Scale
contig29:
5 kb
Chr3R
Custom Gene Model
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
geneid Gene Predictions
Genscan Gene Predictions
GeneID Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
Predicted Splice Sites
RNA-Seq Alignment Summary for Mixed Embryos (E225)
RNA-Seq Alignment Summary for Adult Females
RNA-Seq Alignment Summary for Adult Males
TopHat Junctions for Mixed Embryos (Eggs)
TopHat Junctions for Adult Females
TopHat Junctions for Adult Males
Cufflinks Transcripts Generated by HOSC
Transcripts Assembled by Oases
7 Drosophila Species Multiz Alignments & phastCons Scores
PhastCons Conserved Elements (7 Drosophila Species)
Repeating Elements by RepeatMasker
Conservation
Most Conserved
RepeatMasker

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

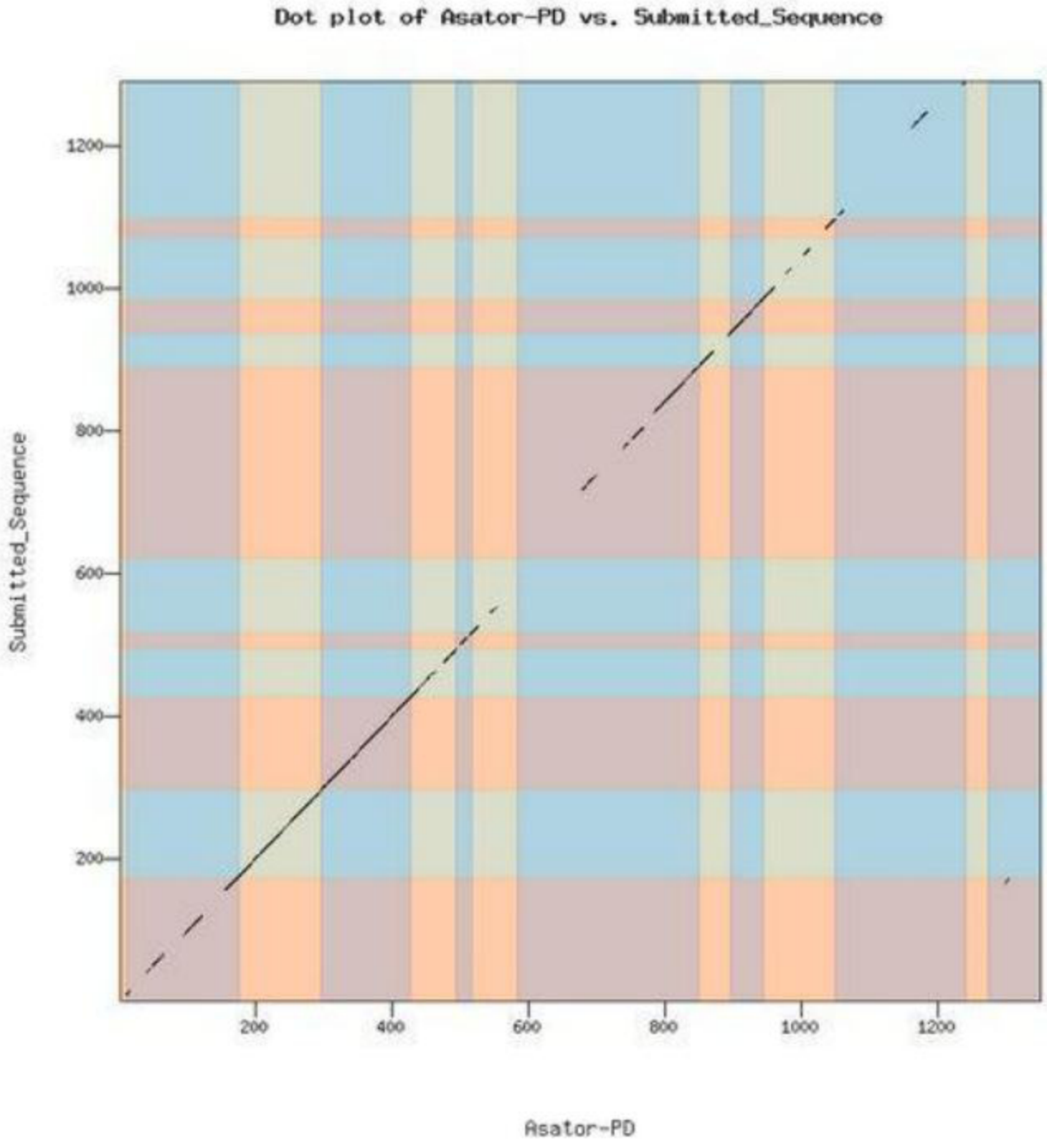
Show an alignment between the protein sequence for your gene model and the protein sequence from the putative *D. melanogaster* ortholog. You can use the protein alignment generated by the Gene Model Checker or you can generate a new alignment using BLAST 2 Sequences (*bl2seq*). **Copy and paste the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



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).

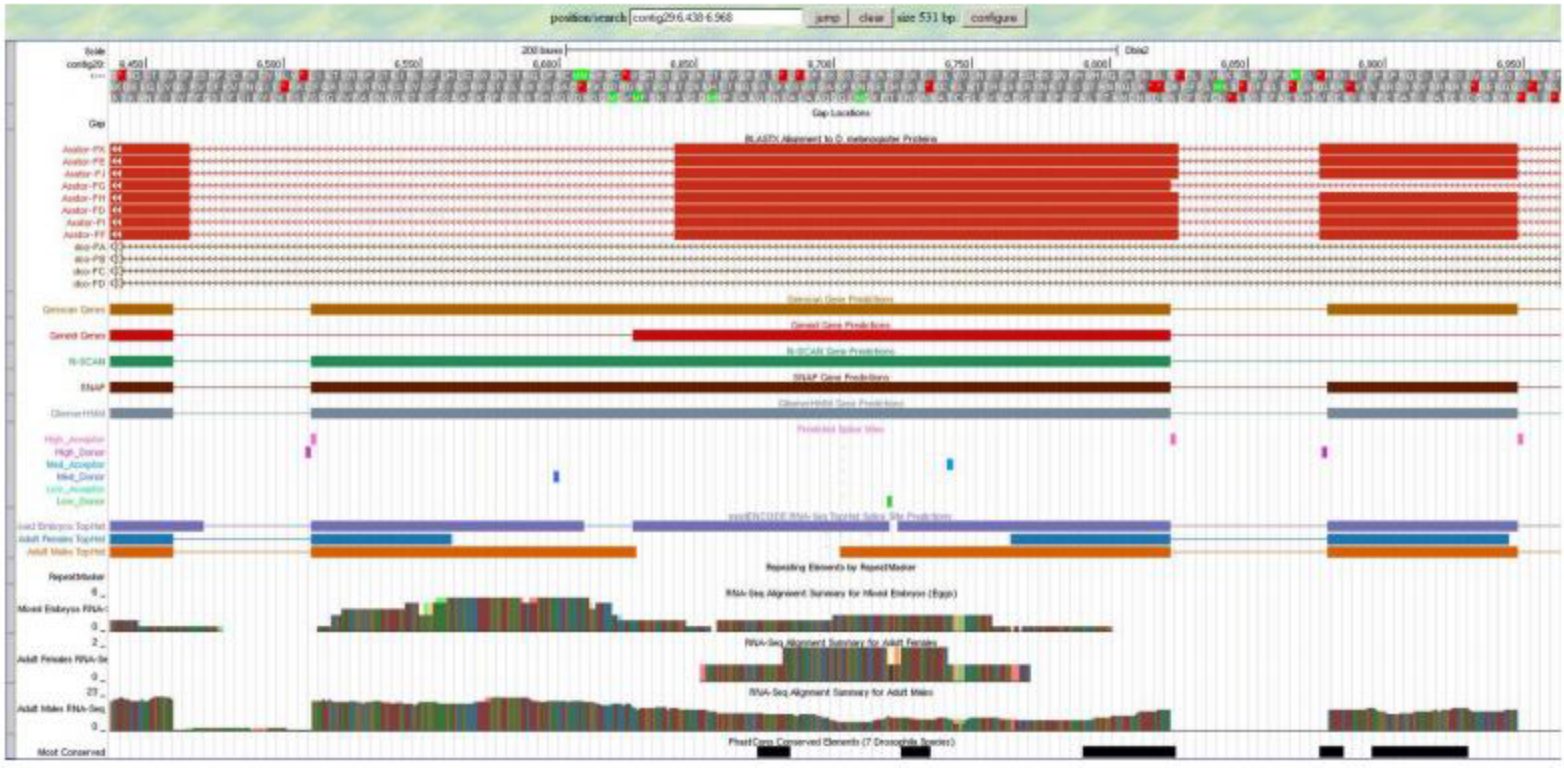
Note: Large vertical and horizontal gap near exon boundaries in the dot plot often indicates that an incorrect splice site might have been picked. Please re-examine these regions and provide a detail justification as to why you have selected this particular set of donor and acceptor sites.



The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in D. melanogaster. An open reading frame the entire way through is present as well as computational evidence of tophat junctions and RNA Seq and a high donor in support of the new splice site.

```
D.Wil       PVDSVLQNVNSGINQNPLPKT---------------LAQ------QQLAAANPNLPSALI
D.Moja      SADCVVQNVNLGNPNNLLKHQQQQQQQQQQQQQQQQQQQQQQQLVLVSNTSY----------
D.Viri      TADCAVQNVNSGI--QNT-----------TLPKQQHQQQQQLVLVANTSIAIPNLPSALI
D.Grim      TAESALQNSGTLNALPKQ-----------QQQQLQQQQQQQLVLVSNTAIAVANLPSALI
D.Anan      LVEAASQYG--INNQNNLAKELLQQ------------QKSCECL--------------L
D.pseud     KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Pers      KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Erec      SGDAIVPHGNTANNQNITTKRLQQQ------------STLT--TNSQV--AIPNIQV---
D.Yak       SGEATIQHGNTANNQNIKTKGLQQQ------------STMT--TNSQV--AIPNIQSAPI
D.mel       SGEPMVQHGNAANNQNITSKGLQQQ------------STL---TNSQV--AIANIQSAP-
D.Sim       SGEALIQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
D.          SGEALTQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
                                :

D.Wil       KSNHENPAAI------VGQIKSLHGHTAIHTKKCQPHVGTGSLSISNQNNSAPVYQYPHA
D.Moja      ------------------------------------------------------------PES
D.Viri      KSST-VGIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNNSAPVY-YPES
D.Grim      KPSTVTGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQYPES
D.Anan      VESML-------------------------------------------------------
D.pseud     KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Pers      KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Erec      ----H---QREDVQ----------------------------------------------YTKL
D.Yak       KSGMT---EREDVQ----------------------------------------------YTKM
D.mel       --SMI---EREDVQ----------------------------------------------YTKL
D.Sim       KSSMI---EREDVQ----------------------------------------------YTKL
D.          KSSMI---EREDVQ----------------------------------------------YTKL
```

Marked above is the region in question in a multiple alignment and the region is shown below in the genome browser.



A loss of intron is also present between the ortholog in melanogaster and the 3' end of the exon in biarmipes (11050-10529).

```
D.Wil    ------------------------------------------------------------
D.Moja   ------------------------------------------------------------
D.Viri   ------------------------------------------------------------
D.Grim   ------------------------------------------------------------
D.Anan   ------------------------------------------------------------
D.pseud  -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Pers   -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Biar   MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEQYISLNRD---CQKNLF
D.Erec   ------------------------------------------------------------
D.Yak    ------------------------------------------------------------
D.mel    ---------MFWHLLCVP-NDENASAPDDG--NQSCQPSSKQDQYLSPNRN---CQKNLL
D.Sim    ------------------------------------------------------------
```

The poorly conserved region contains a small, 8 amino acid exon in melanogaster which is present on the 3' end of an exon (11050-10529) due to a loss of intron. 4 of the 8 amino acids in melanogaster are present on the 3' end of this exon in the order which they were present on melanogaster. There is an open reading frame all the way through and a translated nucleotide blast of melanogaster following a loss of intron vs biarmipes produces the alignment below with a significant e-value of 4*10^-58.

```
Query  1    MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEQYISLNRDCQKNLFRLHppppskpppLA  180
            M + L   NDENAS+P DGN   +CQP   Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct  1    MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPPPPSKPPPLV  58

Query  181  GAILQSRLLKQISSGanaenaealee---HRYPNALQRSATLPAKHNRLGVRSRVTFKVP  351
            GAILQ+RLL QIS  A A+    L       YPN LQRSATLPAKHNRLGVRSRVTFKVP
Sbjct  59   GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNRLGVRSRVTFKVP  118

Query  352  SSITPAVDPGS-EPDPGQNQVVAAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK  522
            SS  PA D  S +P   +NQV  A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct  119  SSNLPAQDSYSHQP---RNQVAVAAKDGILVDVKAKESVKMTSEDLLQPGHVVKERWK  173
```

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): Asator-PF

Names of the isoforms with identical coding sequences as this isoform

_____

Is the 5' end of this isoform missing from the end of project: No

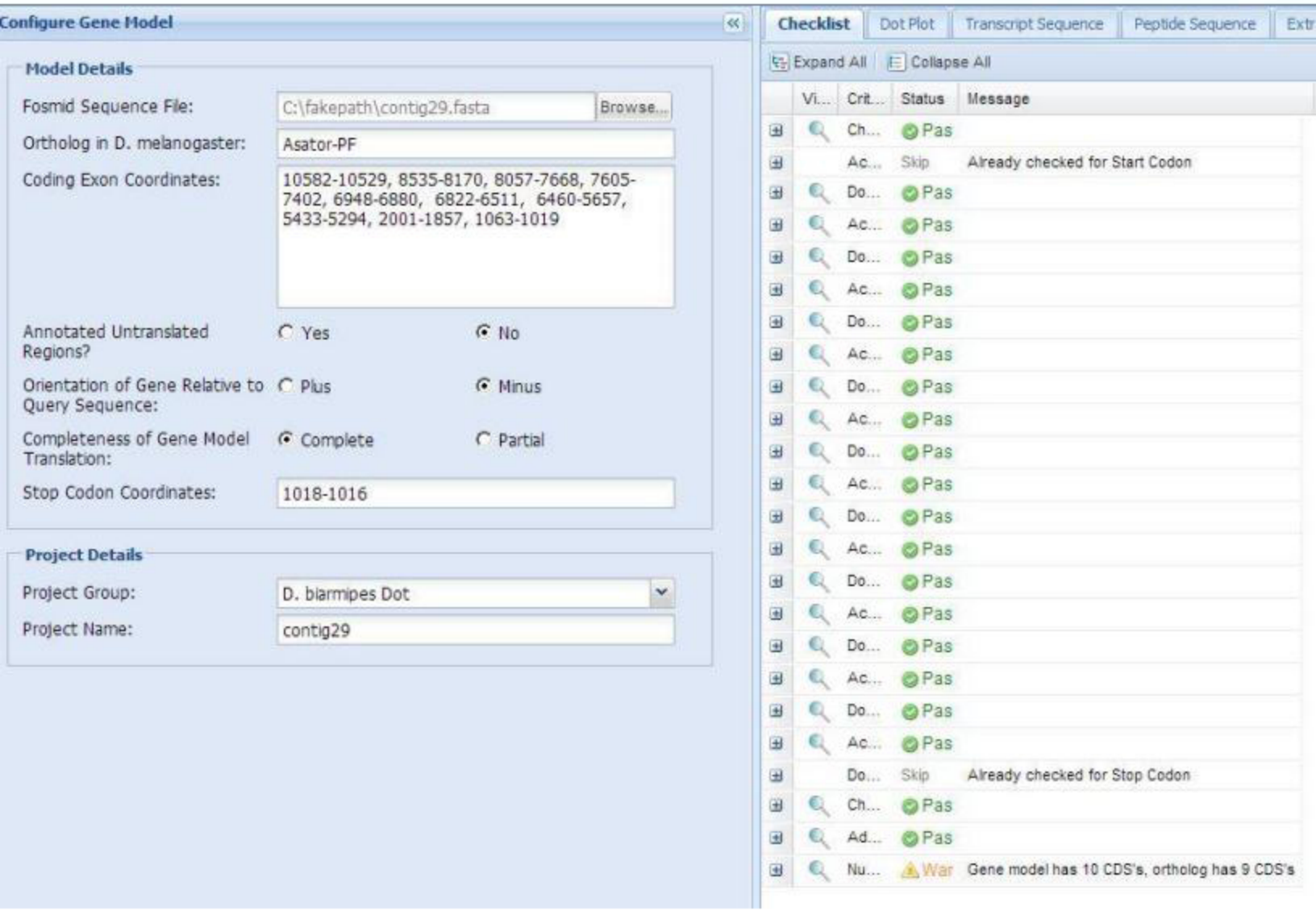
If so, how many exons are missing from the 5' end: _____

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: _____

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:**

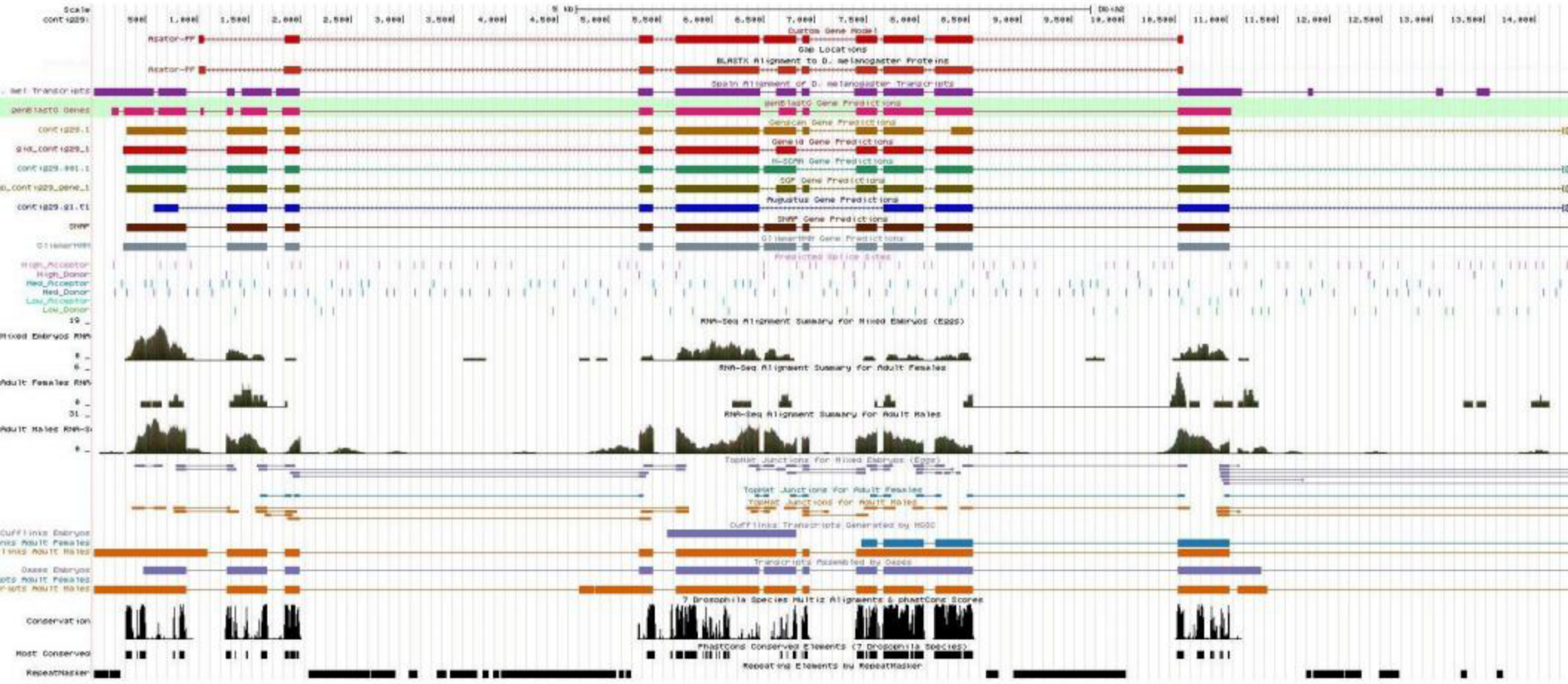


2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker (see page 10 of the Gene Model Checker user guide on how to do this; you can find the guide under "Help" -> "Documentations" -> "Web Framework" on the GEP website at http://gep.wustl.edu). Capture a screenshot of your gene model shown on the Genome Browser for your project; zoom in so that only this isoform is in the screenshot. Include the following evidence tracks in the screenshot if they are available.

1. A sequence alignment track (D. mel Protein or Other RefSeq)
2. At least one gene prediction track (e.g. Genscan)
3. At least one RNA-Seq track (e.g. RNA-Seq Alignment Summary)
4. A comparative genomics track
   (e.g. Conservation, D. mel. Net Alignment, 3-way, 5-way or 7-way multiz)

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


Custom Gene Model
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
genBlastG Gene Predictions
Genscan Gene Predictions
Geneid Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
Predicted Splice Sites
RNA-Seq Alignment Summary for Mixed Embryos (ES2S)
RNA-Seq Alignment Summary for Adult Females
RNA-Seq Alignment Summary for Adult Males
Cufflinks Transcripts Generated by HQSC
Transcripts Assembled by Oases
7 Drosophila Species Multiz Alignments & phastCons Scores
PhastCons Conserved Elements (7 Drosophila Species)
Repeating Elements by RepeatMasker

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 Asator-PF vs. Submitted_Seq

View plain text version

Identity: 649/856 (75.8%), **Similarity:** 700/856 (81.8%), **Gaps:** 58/856 (6.8%)

```
Asator-PF          1 MTSEDLLQPGHVVKERWRVVRKIGGGGFGEIYEGQDLITREQVALKVESARQPKQVLKME  60
                     ************************************************************
Submitted_Seq      1 MTSEDLLQPGHVVKERWRVVRKIGGGGFGEIYEGQDLITREQVALKVESARQPKQVLKME  60

Asator-PF         61 VAVLKKLQGKEHVCRFIGCGRNDRFNYVVMQLQGKNLAELRRAQPRGAFSLSTTLRLGLQ  120
                     ************************************************************
Submitted_Seq     61 VAVLKKLQGKEHVCRFIGCGRNDRFNYVVMQLQGKNLAELRRAQPRGAFSLSTTLRLGLQ  120

Asator-PF        121 ILKAIESIHSVGFLHRDIKPSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA  180
                     ************************************************************
Submitted_Seq    121 ILKAIESIHSVGFLHRDIKPSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA  180

Asator-PF        181 GFRGTVRYASINAHRNREMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKEQVGLTKEKYDH  240
                     ************************************************************
Submitted_Seq    181 GFRGTVRYASINAHRNREMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKEQVGLTKEKYDH  240

Asator-PF        241 RILLKHLPSDLKQFLEHIQSLTYGDRPDYAMLIGLFERCMKRRGVKESDPYDWEKVDSTA  300
                     ***********************.************************************::
Submitted_Seq    241 RILLKHLPSDLKQFLEHIQSLTYADRPDYAMLIGLFERCMKRRGVKESDPYDWEKVDSAT  300

Asator-PF        301 IGNISATGNPSIPIKSDYMHGNITQMTVAASNASGIEYIRKRAEIETAHITATDPLNIKE  360
                     *****::***.:*:*.**:**********************:***.:**********:**:***
Submitted_Seq    301 IGNISTSGNPPVPVKNDYIHGNITQMTVAASNASGTEYVRKRGDIETAHITATEPLHIKE  360

Asator-PF        361 KVDKNCNATSLAQPAKGSGEPMVQHGNAANNQNITSKG-LQQQSTLT-NSQVAIANIQSA  418
                     **********:**  .* ***. ** * :*******.** ****::*: ***.**.::**.
Submitted_Seq    361 RVDKNCNATTLAFQPKTSGEANVQLGCTANNQNITPKGMLQQQAALAINSQAAIPHMQSV  420

Asator-PF        419 PSM------IEREDVQ------------------------------YTKLEEGAPTKFITM  443
                     *       :  .***                               * :**: **. *:.
Submitted_Seq    421 PIKSPMVGMGSHDVQVHTKNSQPQKGAASFSSTNQNNSAPVYGNSYLQLEDKAPNMFLPT  480

Asator-PF        444 KPNGECDN-VDIAAKCIFEQKHVEAND-DIVGRASLSGVEQHYKSQIKKHNSPEIANKQI  501
                     ***.*.:. **:*.*.***:*.*::** : ..:* *** :*: ***:** * ** **:*:
Submitted_Seq    481 KPNAESESTVDVAPKSIFEEKFVDSNDAECANKAFLSGEQQQQKSQVKKLNLPESANQQV  540

Asator-PF        502 QRTGTVTNDKTSEVNRSTEEQKSTFGRLRVLTAPPMSVHDLPSGGGHSHQVSDLSGKQDP  561
                     *:  .*:*:*:** :*::** ********************.:** * :* :*** ****
Submitted_Seq    541 QKLENVANEKSSEDKRASQEPKSTFGRLRVLTAPPMSVHDLTTGG-HIQQGTDLSIKQDP  599

Asator-PF        562 YAATSNAAPIGINSSSTKFG-SQHGQIFGLAAMPPINRRSATSTNLRPSSSASQ-----R  615
                       ::***.* . ****:*:. .*******:: ** :***************..:     *
Submitted_Seq    600 --SSSNAGPAAGNSSSSKLAINQHGQIFGITLMPQVNRRSATSTNLRPSSSGGNTNPIHR  657

Asator-PF        616 INSGSTIGGAVGNGSNTARSSVAGDHSVTQFALIDDENVSALQQVTKGGALTLASQWKSQ  675
                     ** **: **. *.**************************************************
Submitted_Seq    658 INIGSA-GGGGGTGSNTARSSVAGDHSVTQFALIDDENVSALQQVTKGGALTLASQWKSQ  716

Asator-PF        676 FDDSEDTTDNEWNREHQLQPNLEQLIKLDISLPLNEAKPFPQHGVAGTGKLINPPGEAKG  735
                     *************** * ************.******* * *: *:.** * :** *.:
Submitted_Seq    717 FDDSEDTTDNEWNREPQSQPNLEQLIKLDIPLPLNEAKHFCQNVVTDTGILTKPPIEGNE  776

Asator-PF        736 RPKRYTLNITGIENYEALRISIPNCWSEPAMGNVLRKGLEPPAVQQAAFDDTVSILNKSL  795
                     :**********************:*************.************** *: *
Submitted_Seq    777 KPKRYTLNITGIENYEALRISIPHCWSEPAMGNVLRKDLEPPAVQQAAFDDTNSLNN---  833

Asator-PF        796 PVIIYIYMHAVTFFCF 811
```

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 Asator-PF vs. Submitted_Seq

View plain text version

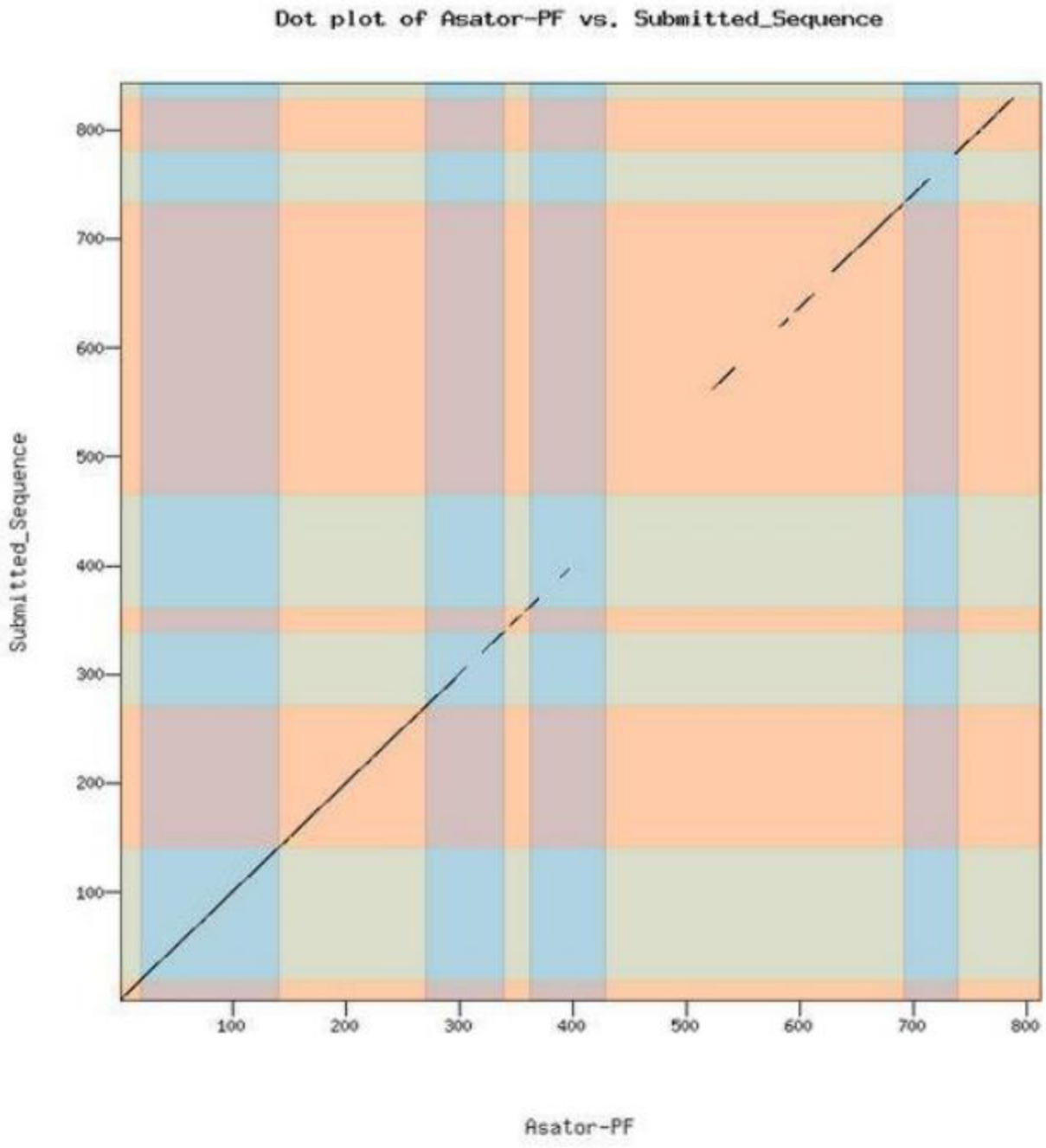
Identity: 649/856 (75.8%), **Similarity:** 700/856 (81.8%), **Gaps:** 58/856 (6.8%)

```
Asator-PF          1 MTSEDLLQPGHVVKERWKVVRKIGGGGFGEIYEGQDLITREQVALKVESARQPKQVLKME 60
                     ************************************************************
Submitted_Seq      1 MTSEDLLQPGHVVKERWKVVRKIGGGGFGEIYEGQDLITREQVALKVESARQPKQVLKME 60

Asator-PF         61 VAVLKKLQGKEHVCRFIGCGRNDRFNYVVMQLQGKNLAELRRAQPRGAFSLSTTLRLGLQ 120
                     ************************************************************
Submitted_Seq     61 VAVLKKLQGKEHVCRFIGCGRNDRFNYVVMQLQGKNLAELRRAQPRGAFSLSTTLRLGLQ 120

Asator-PF        121 ILKAIESIHSVGFLHRDIKPSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA 180
                     ************************************************************
Submitted_Seq    121 ILKAIESIHSVGFLHRDIKPSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA 180

Asator-PF        181 GFRGTVRYASINAHRNREMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKEQVGLTKEKYDH 240
                     ************************************************************
Submitted_Seq    181 GFRGTVRYASINAHRNREMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKEQVGLTKEKYDH 240

Asator-PF        241 RILLKHLPSDLKQFLEHIQSLTYGDRPDYAMLIGLFERCMKRRGVKESDPYDWEKVDSTA 300
                     **********************.*************************************::
Submitted_Seq    241 RILLKHLPSDLKQFLEHIQSLTYADRPDYAMLIGLFERCMKRRGVKESDPYDWEKVDSAT 300

Asator-PF        301 IGNISATGNPSIPIKSDYMHGNITQMTVAASNASGTEYIRKRAEIETAHITATDPLNIKE 360
                     *****::***.:*:*.**:**********************:***.:*********:**:***
Submitted_Seq    301 IGNISTSGNPPVPVKNDYIHGNITQMTVAASNASGTEYVRKRGDIETAHITATEPLHIKE 360

Asator-PF        361 KVDKNCNATSLAQPAKGSGEPMVQHGNAANNQNITSKG-LQQQSTLT-NSQVAIANIQSA 418
                     *********:**  .* ***. ** * :********.** ****::*: ***.**.::**.
Submitted_Seq    361 KVDKNCNATTLAFQPKTSGEANVQLGCTANNQNITPKGMLQQQAALAINSQAAIPHMQSV 420

Asator-PF        419 PSM------IEREDVQ-----------------------------YTKLEEGAPIKFITM 443
                     *       :  .***                             * :**: **. *:.
Submitted_Seq    421 PIKSPMVGMGSHDVQVHTKNSQPQKGAASFSSTNQNNSAPVYGNSYLQLEDKAPNMFLPT 480

Asator-PF        444 KPNGECDN-VDIAAKCIFEQKHVEAND-DIVGRASLSGVEQHYKSQIKKHNSPEIANKQI 501
                     ***.*.:. **:*.*.***:*.*::** : ..:* *** :*: ***:** * ** **:*:
Submitted_Seq    481 KPNAESESTVDVAPKSIFEEKFVDSNDAECANKAFLSGEQQQQKSQVKKLNLPESANQQV 540

Asator-PF        502 QRTGTVTNDKTSEVNRSTEEQKSTFGRLRVLTAPPMSVHDLPSGGGHSHQVSDLSGKQDP 561
                     *:  .*:*:*:** :*::::* **********************.:** * :* :*** ****
Submitted_Seq    541 QKLENVANEKSSEDKRASQEPKSTFGRLRVLTAPPMSVHDLITGG-HIQQGTDLSIKQDP 599

Asator-PF        562 YAATSNAAPIGINSSSTKFG-SQHGQIFGLAAMPPINRRSATSTNLRPSSSASQ-----R 615
                       ::***.* . ****:*:. .*******:: ** :****************..:     *
Submitted_Seq    600 --SSSNAGPAAGNSSSSKLAINQHGQIFGITLMPQVNRRSATSTNLRPSSSGGNTNPIHR 657

Asator-PF        616 INSGSTIGGAVGNGSNTARSSVAGDHSVTQFALIDDENVSALQQVTKGGALTLASQWKSQ 675
                     ** **: **. *.**********************************************
Submitted_Seq    658 INIGSA-GGGGGTGSNTARSSVAGDHSVTQFALIDDENVSALQQVTKGGALTLASQWKSQ 716

Asator-PF        676 FDDSEDTTDNEWNREHQLQPNLEQLIKLDISLPLNEAKPFPQHGVAGTGKLINPPGEAKG 735
                     *************** * ************.******* * *: *:.** * :** *.:
Submitted_Seq    717 FDDSEDTTDNEWNREPQSQPNLEQLIKLDIPLPLNEAKHFCQNVVTDTGILTKPPIEGNE 776

Asator-PF        736 RPKRYTLNITGIENYEALRISIPNCWSEPAMGNVLRKGLEPPAVQQAAFDDTVSILNKSL 795
                     :**********************:*************.********************* *: *
Submitted_Seq    777 KPKRYTLNITGIENYEALRISIPHCWSEPAMGNVLRKDLEPPAVQQAAFDDTNSLNN--- 833

Asator-PF        796 PVIIYIYMHAVTFFCF 811
```

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).

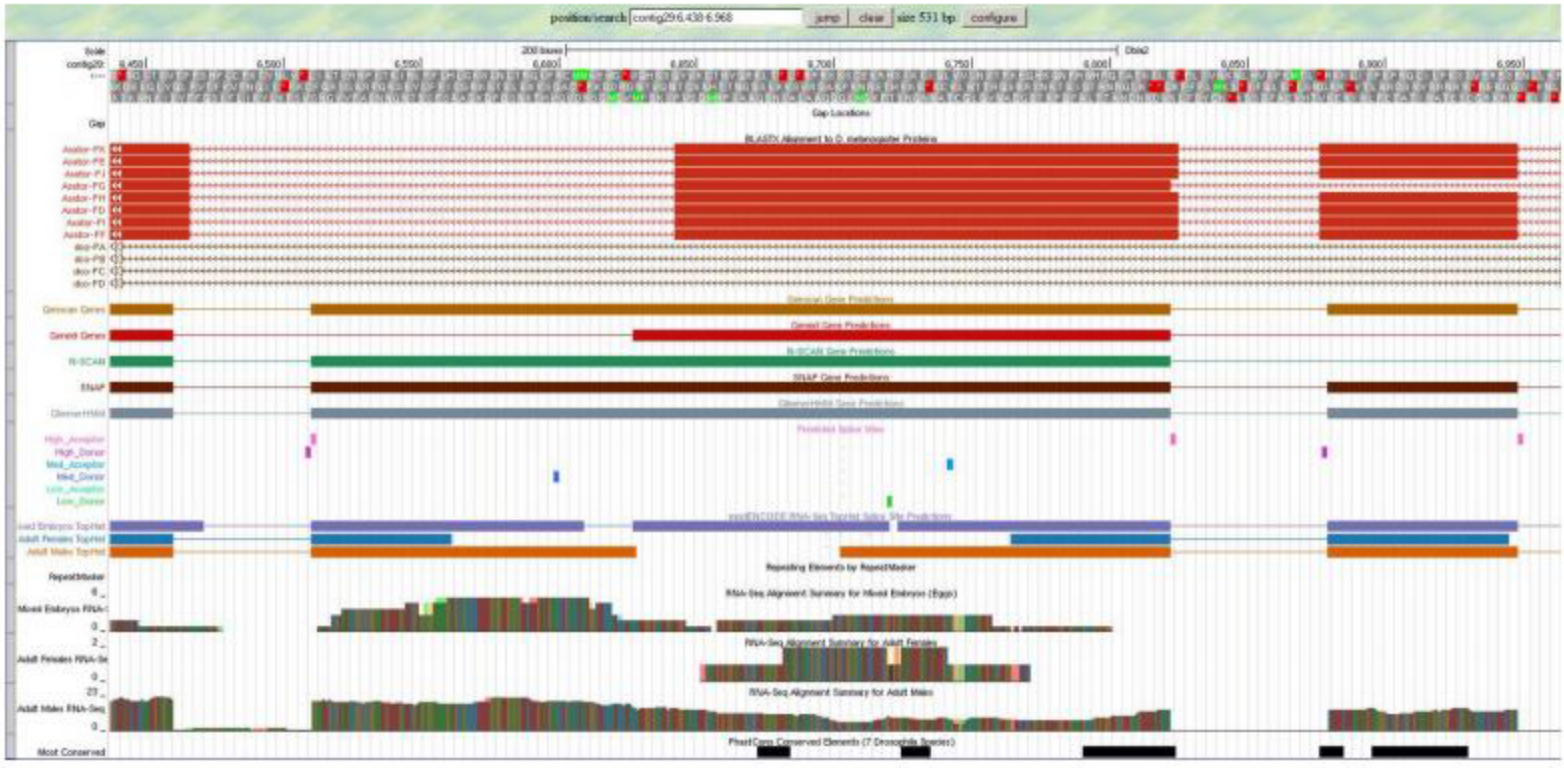
Note: Large vertical and horizontal gap near exon boundaries in the dot plot often indicates that an incorrect splice site might have been picked. Please re-examine these regions and provide a detail justification as to why you have selected this particular set of donor and acceptor sites.



The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in D. melanogaster. An open reading frame the entire way through is present as well as computational evidence of tophat junctions and RNA Seq and a high donor in support of the new splice site.

```
D.Wil       PVDSVLQNVNSGINQNPLPKT---------------LAQ------QQLAAANPNLPSALI
D.Moja      SADCVVQNVNLGNPNNLLKHQQQQQQQQQQQQQQQQQQQQQQQLVKVSNTSY----------
D.Viri      TADCAVQNVNSGI--QNT-----------TLPKQQHQQQQQLVLVANTSIAIPNLPSALI
D.Grim      TAESALQNSGTLNALPKQ-----------QQQQLQQQQQQQLYLVSNTAIAVANLPSALI
D.Anan      LVEAASQYG--INNQNNLAKELLQQ------------QKSCECL--------------L
D.pseud     KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Pers      KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Erec      SGDAIVPHGNTANNQNITTKRLQQQ------------STLT--TNSQV--AIPNIQV---
D.Yak       SGEATIQHGNTANNQNIKTKGLQQQ------------STMT--TNSQV--AIPNIQSAPI
D.mel       SGEPMVQHGNAANNQNITSKGLQQQ------------STL---TNSQV--AIANIQSAP-
D.Sim       SGEALIQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
D.          SGEALTQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
                                :

D.Wil       KSNHENPAAI------VGQIKSLHGHTAIHTKKCQPHVGTGSLSISNQNNSAPVYQYPHA
D.Moja      ------------------------------------------------------------PES
D.Viri      KSST-VGIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNNSAPVY-YPES
D.Grim      KPSTVTGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQYPES
D.Anan      VESML-------------------------------------------------------
D.pseud     KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Pers      KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Erec      ----H---QREDVQ----------------------------------------------YTKL
D.Yak       KSGMT---EREDVQ----------------------------------------------YTKM
D.mel       --SMI---EREDVQ----------------------------------------------YTKL
D.Sim       KSSMI---EREDVQ----------------------------------------------YTKL
D.          KSSMI---EREDVQ----------------------------------------------YTKL
```

Marked above is the region in question in a multiple alignment and the region is shown below in the genome browser.



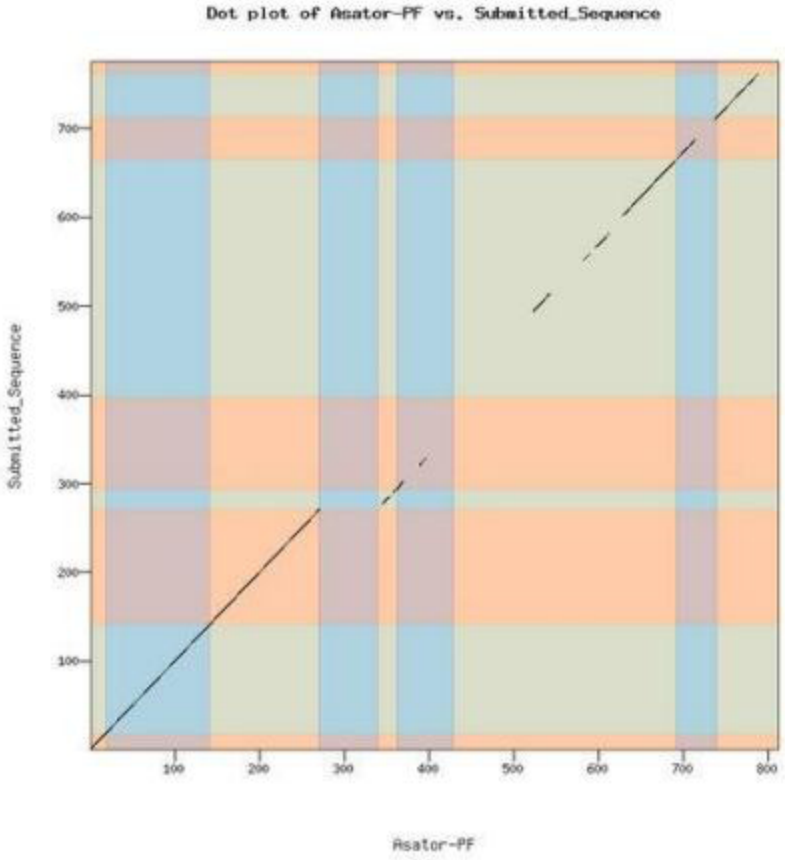
A loss of intron is also present between the ortholog in melanogaster and the 3' end of the exon in biarmipes (11050-10529).

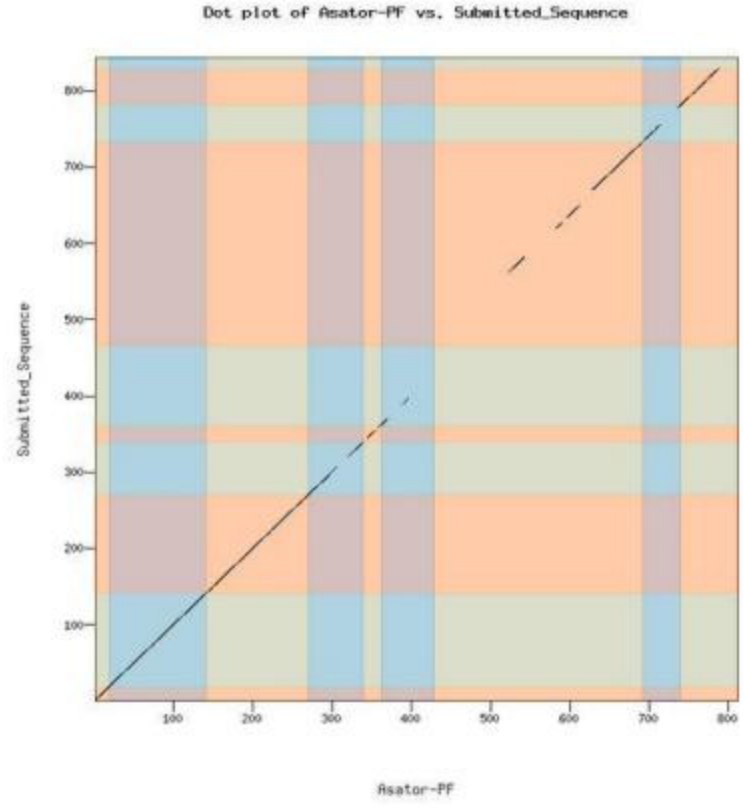
```
D.Wil    ------------------------------------------------------------
D.Moja   ------------------------------------------------------------
D.Viri   ------------------------------------------------------------
D.Grim   ------------------------------------------------------------
D.Anan   ------------------------------------------------------------
D.pseud  -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Pers   -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Biar   MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEQYISLNRD---CQKNLF
D.Erec   ------------------------------------------------------------
D.Yak    ------------------------------------------------------------
D.mel    ---------MFWHLLCVP-NDENASAPDDG--NQSCQPSSKQDQYLSPNRN---CQKNLL
D.Sim    ------------------------------------------------------------
```

The poorly conserved region contains a small, 8 amino acid exon in melanogaster which is present on the 3' end of an exon (11050-10529) due to a loss of intron. 4 of the 8 amino acids in melanogaster are present on the 3' end of this exon in the order which they were present on melanogaster. There is an open reading frame all the way through and a translated nucleotide blast of melanogaster following a loss of intron vs biarmipes produces the alignment below with a significant e-value of 4*10^-58.

```
Query  1    MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEQYISLNRDCQKNLFRLHppppskpppLA  180
            M + L   NDENAS+P DGN   +CQP   Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct  1    MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPPPPSKPPPLV  58

Query  181  GAILQSRLLKQISSGanaenaealee---HRYPNALQRSATLPAKHNRLGVRSRVTFKVP  351
            GAILQ+RLL QIS  A A+    L       YPN LQRSATLPAKHNRLGVRSRVTFKVP
Sbjct  59   GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNRLGVRSRVTFKVP  118

Query  352  SSITPAVDPGS-EPDPGQNQVVAAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK  522
            SS  PA D  S +P   +NQV  A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct  119  SSNLPAQDSYSHQP---RNQVAVAAKDGILVDVKAKESVKMTSEDLLQPGHVVKERWK  173
```

The gene model checker and gene record finder for Asator posit that Asator-PF has 9 CD's in the melanogaster ortholog. In the genome browser, however, the isoform contains 10 coding regions. Dot plots of the isoform containing only those in the record finder produces the heavily offset graph directly below. Adding the coding region of 7605-7402 produces the lower graph which fits the plot better as well as contains the correct amount of exons on the genome browser with the exon well established and common amongst the other isoforms.





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): Asator-PG

Names of the isoforms with identical coding sequences as this isoform

_____

Is the 5' end of this isoform missing from the end of project: No

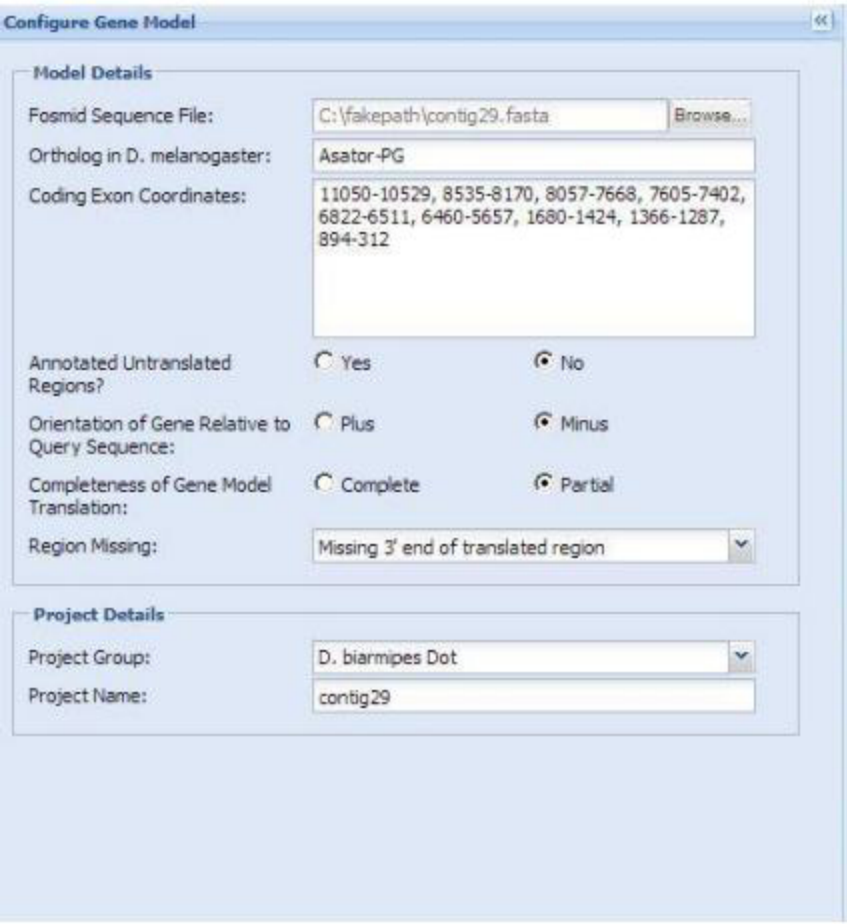
If so, how many exons are missing from the 5' end: _____

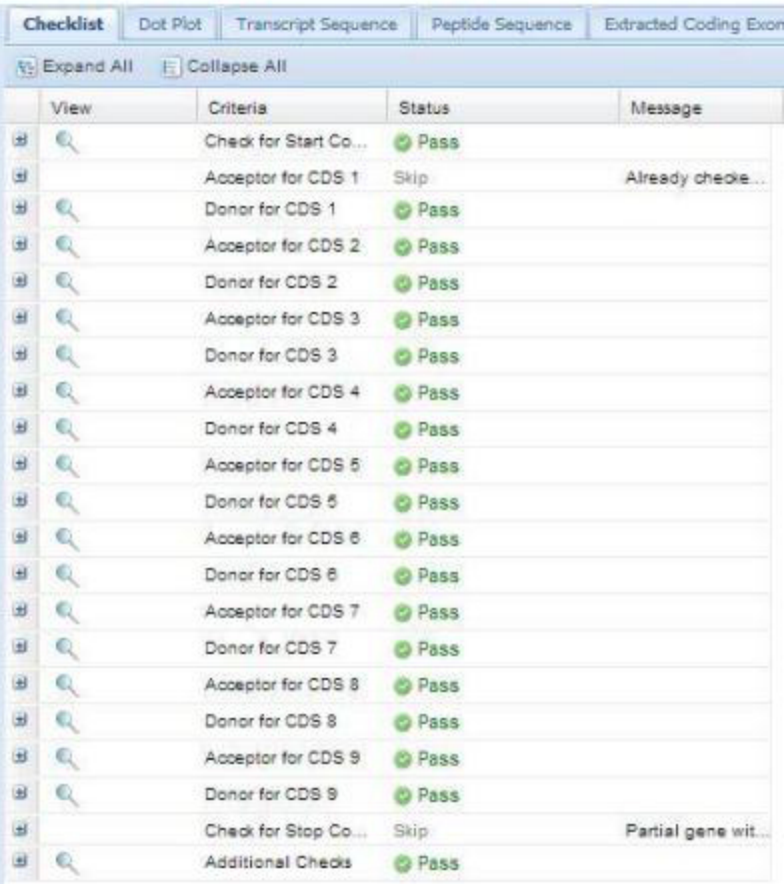
Is the 3' end of this isoform missing from the end of the project: No

If so, how many exons are missing from the 3' end: _____

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:**



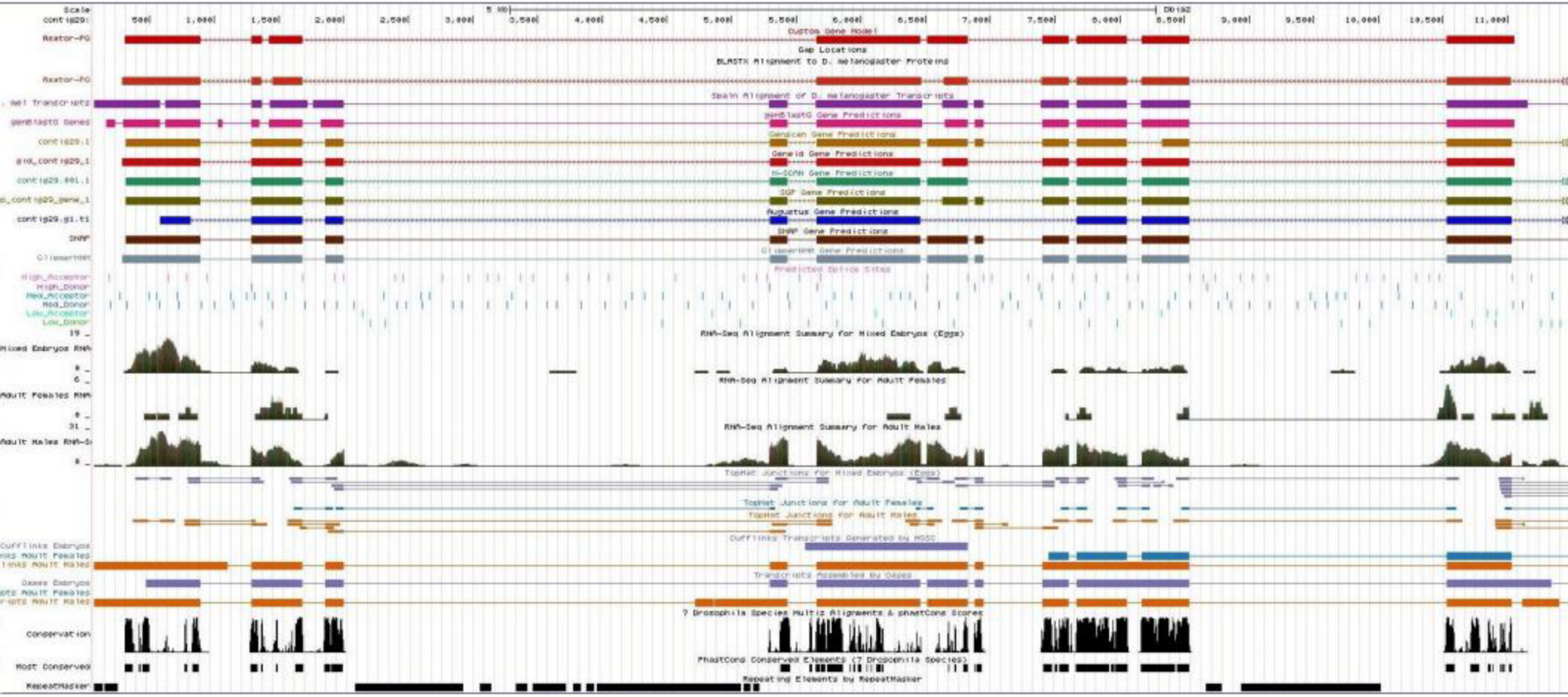


2. View the gene model on the Genome Browser

Using the custom track feature from the Gene Model Checker (see page 10 of the Gene Model Checker user guide on how to do this; you can find the guide under "Help" -> "Documentations" -> "Web Framework" on the GEP website at http://gep.wustl.edu). Capture a screenshot of your gene model shown on the Genome Browser for your project; zoom in so that only this isoform is in the screenshot. Include the following evidence tracks in the screenshot if they are available.

1. A sequence alignment track (D. mel Protein or Other RefSeq)
2. At least one gene prediction track (e.g. Genscan)
3. At least one RNA-Seq track (e.g. RNA-Seq Alignment Summary)
4. A comparative genomics track
   (e.g. Conservation, D. mel. Net Alignment, 3-way, 5-way or 7-way multiz)

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


Scale
contig29:
Augustus Gene Model
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
genBlastG Gene Predictions
Genscan Gene Predictions
Geneid Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
Oligoarray Gene Predictions
Predicted Splice Sites
RNA-Seq Alignment Summary for Mixed Embryos (Eggs)
RNA-Seq Alignment Summary for Adult Females
RNA-Seq Alignment Summary for Adult Males
TopHat Junctions for Mixed Embryos (Eggs)
Tophat Junctions for Adult Females
TopHat Junctions for Adult Males
Cufflinks Transcripts Generated by HGSC
Transcripts Assembled by Oases
7 Drosophila Species Multiz Alignments & phastCons Scores
PhastCons Conserved Elements (7 Drosophila Species)
Repeating Elements by RepeatMasker
Conservation
Most Conserved
RepeatMasker

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 Asator-PG vs. Submitted_Seq

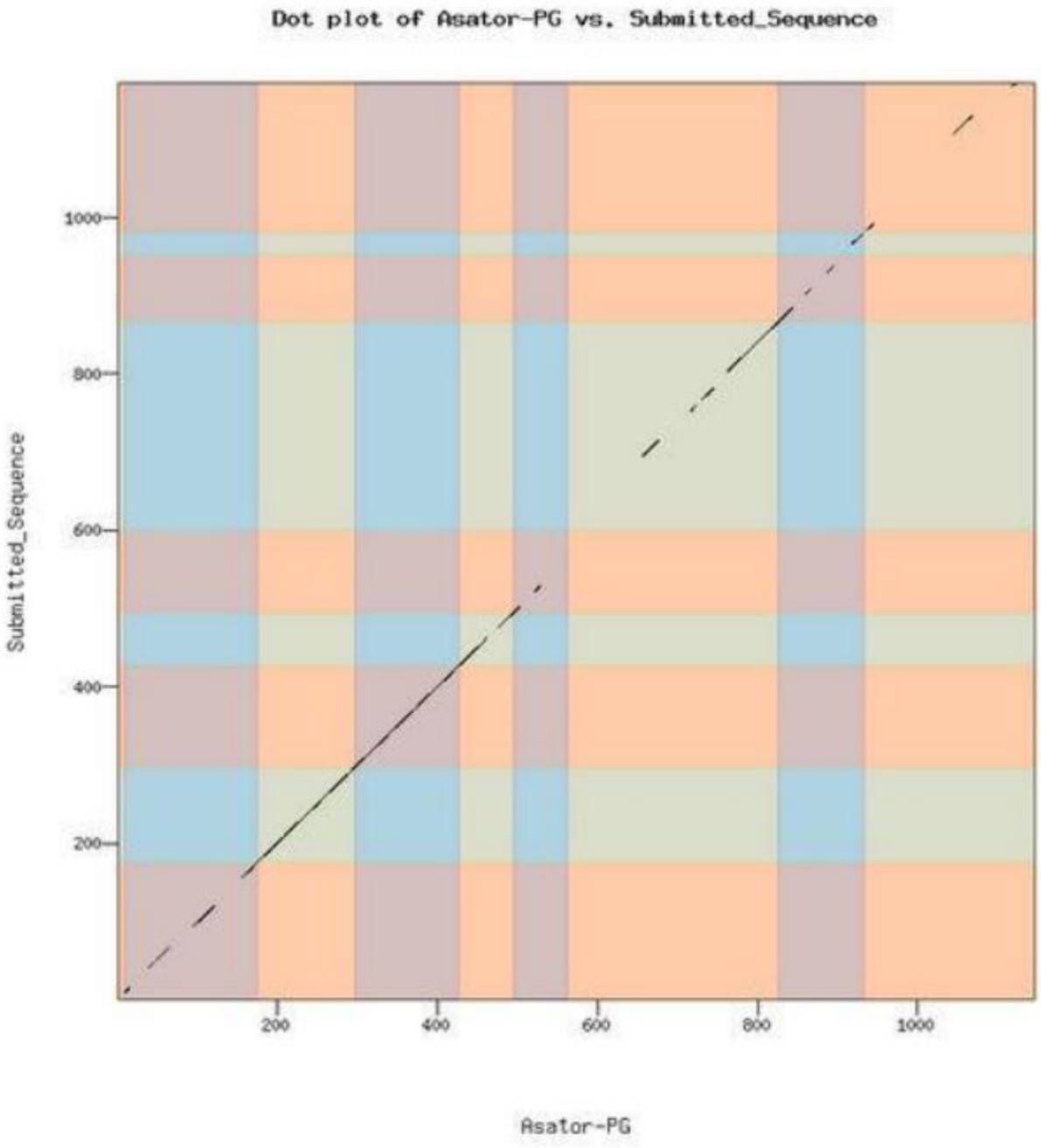
View plain text version

Identity: 822/1216 (67.6%), **Similarity:** 916/1216 (75.3%), **Gaps:** 116/1216 (9.5%)

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).

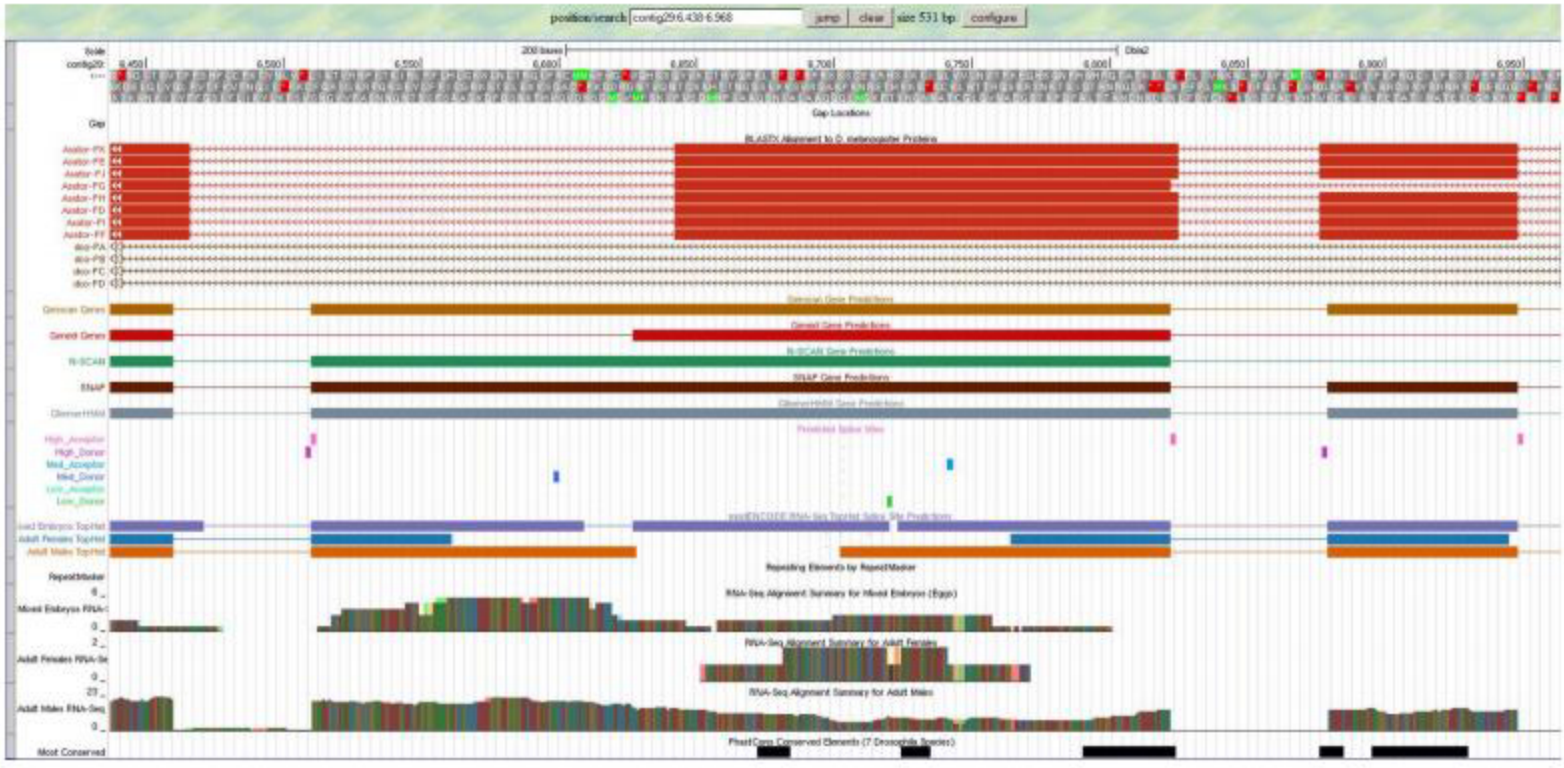
Note: Large vertical and horizontal gap near exon boundaries in the dot plot often indicates that an incorrect splice site might have been picked. Please re-examine these regions and provide a detail justification as to why you have selected this particular set of donor and acceptor sites.



The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in D. melanogaster. An open reading frame the entire way through is present as well as computational evidence of tophat junctions and RNA Seq and a high donor in support of the new splice site.

```
D.Wil       PVDSVLQNVNSGINQNPLPKT---------------LAQ------QQLAAANPNLPSALI
D.Moja      SADCVVQNVNLGNPNNLLKHQQQQQQQQQQQQQQQQQQQQQQQLVLVSNTSY----------
D.Viri      TADCAVQNVNSGI--QNT-----------TLPKQQHQQQQQLVLVANTSIAIPNLPSALI
D.Grim      TAESALQNSGTLNALPKQ-----------QQQQLQQQQQQQLVLVSNTAIAVANLPSALI
D.Anan      LVEAASQYG--INNQNNLAKELLQQ------------QKSCECL--------------L
D.pseud     KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Pers      KCETSLQHANSANNQNTLPKALLQQ------------QHPPQSANSNVAMATPNLPNALI
D.Erec      SGDAIVPHGNTANNQNITTKRLQQQ------------STLT--TNSQV--AIPNIQV---
D.Yak       SGEATIQHGNTANNQNIKTKGLQQQ------------STMT--TNSQV--AIPNIQSAPI
D.mel       SGEPMVQHGNAANNQNITSKGLQQQ------------STL---TNSQV--AIANIQSAP-
D.Sim       SGEALIQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
D.          SGEALTQHGNTANNQNITSKGLQQP------------STL---TNSQV--AIPNIQSAPI
                                :

D.Wil       KSNHENPAAI------VGQIKSLHGHTAIHTKKCQPHVGTGSLSISNQNNSAPVYQYPHA
D.Moja      ------------------------------------------------------------PES
D.Viri      KSST-VGIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNNSAPVY-YPES
D.Grim      KPSTVTGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQYPES
D.Anan      VESML-------------------------------------------------------
D.pseud     KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Pers      KSSMV-GVGNHEAP----------------------------------------------YPNL
D.Erec      ----H---QREDVQ----------------------------------------------YTKL
D.Yak       KSGMT---EREDVQ----------------------------------------------YTKM
D.mel       --SMI---EREDVQ----------------------------------------------YTKL
D.Sim       KSSMI---EREDVQ----------------------------------------------YTKL
D.          KSSMI---EREDVQ----------------------------------------------YTKL
```

Marked above is the region in question in a multiple alignment and the region is shown below in the genome browser.



A loss of intron is also present between the ortholog in melanogaster and the 3' end of the exon in biarmipes (11050-10529).

```
D.Wil    ------------------------------------------------------------
D.Moja   ------------------------------------------------------------
D.Viri   ------------------------------------------------------------
D.Grim   ------------------------------------------------------------
D.Anan   ------------------------------------------------------------
D.pseud  -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Pers   -----------MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNKVSSSSKNNI
D.Biar   MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEQYISLNRD---CQKNLF
D.Erec   ------------------------------------------------------------
D.Yak    ------------------------------------------------------------
D.mel    ---------MFWHLLCVP-NDENASAPDDG--NQSCQPSSKQDQYLSPNRN---CQKNLL
D.Sim    ------------------------------------------------------------
```

The poorly conserved region contains a small, 8 amino acid exon in melanogaster which is present on the 3' end of an exon (11050-10529) due to a loss of intron. 4 of the 8 amino acids in melanogaster are present on the 3' end of this exon in the order which they were present on melanogaster. There is an open reading frame all the way through and a translated nucleotide blast of melanogaster following a loss of intron vs biarmipes produces the alignment below with a significant e-value of 4*10^-58.

```
Query  1    MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEQYISLNRDCQKNLFRLHppppskpppLA  180
            M + L   NDENAS+P DGN   +CQP   Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct  1    MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPPPPSKPPPLV  58

Query  181  GAILQSRLLKQISSGanaenaealee---HRYPNALQRSATLPAKHNRLGVRSRVTFKVP  351
            GAILQ+RLL QIS  A A+    L       YPN LQRSATLPAKHNRLGVRSRVTFKVP
Sbjct  59   GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNRLGVRSRVTFKVP  118

Query  352  SSITPAVDPGS-EPDPGQNQVVAAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK  522
            SS  PA D  S +P   +NQV  A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct  119  SSNLPAQDSYSHQP---RNQVAVAAKDGILVDVKAKESVKMTSEDLLQPGHVVKERWK  173
```

Preparing the project for submission

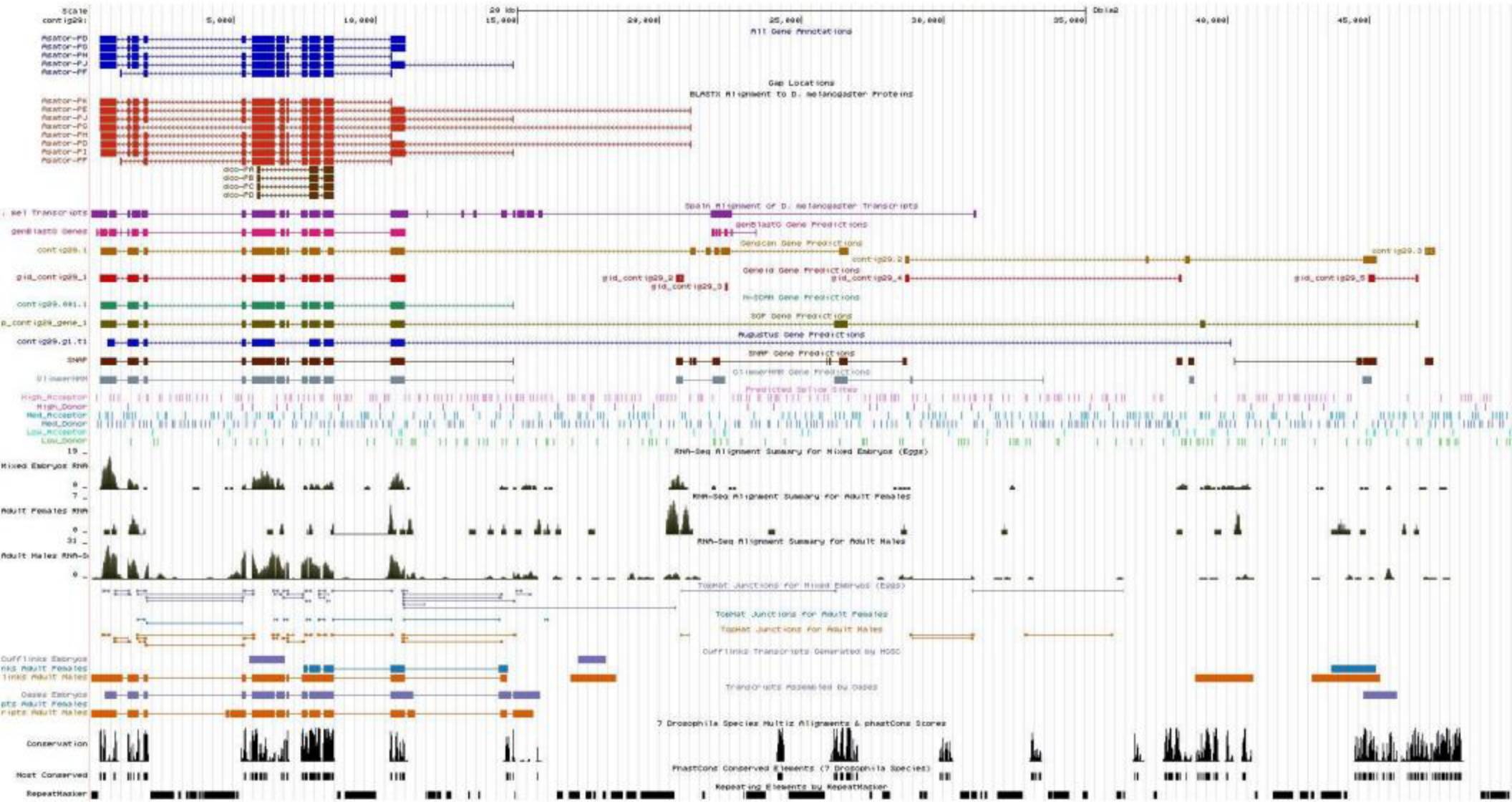
For each project, you should prepare the project GFF, transcripts and peptide sequence files (for **ALL** isoforms) along with this report. You can combine the individual files generated by the Gene Model Checker into a single file using the Annotation Files Merger.

The Annotation Files Merger also allows you to view all the gene models in the combined GFF file within the Genome Browser. Please refer to the Annotation Files Merger User Guide for detail instructions on how to view the combined GFF file on the Genome Browser (you can find the user guide under "Help" -> "Documentations" -> "Web Framework" on the GEP website at http://gep.wustl.edu).

Paste a screenshot (generated by the Annotation Files Merger) with all the gene models you have annotated in this project.

Have you annotated all the genes?

For each region of the project with gene predictions that do not overlap with putative orthologs identified in the BLASTX track, perform a BLASTP search using the predicted amino acid sequence against the non-redundant protein database (*nr*). **Provide a screenshot of the search results.** Provide an explanation for any significant (E-value < 1e-5) hits to known genes in the *nr* database and why you believe these hits do not correspond to real genes in your project.


All Gene Annotations
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
Spaln Alignment of D. melanogaster Transcripts
genBlastG Gene Predictions
Genscan Gene Predictions
Geneid Gene Predictions
N-SCAN Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
GlimmerHMM Gene Predictions
Predicted Splice Sites
RNA-Seq Alignment Summary for Mixed Embryos (Eggs)
RNA-Seq Alignment Summary for Adult Females
RNA-Seq Alignment Summary for Adult Males
TopHat Junctions for Mixed Embryos (Eggs)
TopHat Junctions for Adult Females
TopHat Junctions for Adult Males
Cufflinks Transcripts Generated by MOSC
Transcripts Modelled by Oases
7 Drosophila Species Multiz Alignments & phastCons Scores
PhastCons Conserved Elements (7 Drosophila Species)
Repeating Elements by RepeatMasker