

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 ($E\text{-value} < 1e-5$) 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:**

The screenshot shows the 'Configure Gene Model' window with the following details:

- Model Details:**
 - Fosmid Sequence File: C:\fakepath\contig29.fasta
 - Ortholog in *D. melanogaster*: Asator-PJ
 - Coding Exon Coordinates: 14873-14851, 11027-10529, 8535-8170, 8057-7668, 7605-7402, 6948-6880, 6822-6511, 6460-5657, 5433-5294, 2001-1857, 1680-1424, 1366-1287, 894-312
 - Annotated Untranslated Regions? ☐ Yes ☒ No
 - Orientation of Gene Relative to Query Sequence: ☐ Plus ☒ Minus
 - Completeness of Gene Model Translation: ☐ Complete ☒ Partial
 - Region Missing: Missing 3' end of translated region
- Project Details:**
 - Project Group: D. biarmipes Dot
 - Project Name: contig29

The **Checklist** tab is active, showing the following results:

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	Pass	
Acceptor for CDS 11	Pass	
Donor for CDS 11	Pass	
Acceptor for CDS 12	Pass	
Donor for CDS 12	Pass	
Acceptor for CDS 13	Pass	
Donor for CDS 13	Pass	
Check for Stop Codon	Skip	Partial gene with 3' end missing
Additional Checks	Pass	

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:

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-PJ vs. Submitted_Seq

[View plain text version](#)

Identity: 921/1333 (69.1%), Similarity: 1022/1333 (76.7%), Gaps: 115/1333 (8.6%)

```
Asator-PJ      1  MHLQYVLRNENASAPDGGQS  QPSSMDQYLSNRRGQNTLSIYPPDPKPPPV  58
Submitted_Seq  1  MHLQYVLRNENASAPDGGHRTQPPCNGEQYISLRDCQMLFWLHPPDPKPPPLA  60

Asator-PJ      59  GALLVETLRLQI SPQATADANALNAGGLLYPVVLRQSATLPAKIMRLGVSRVTFPVV  118
Submitted_Seq  61  GALLQSLRLAQISSGANAK  NAKALKKHRYPRALQSATLPAKIMRLGVSRVTFPVV  117

Asator-PJ      119  SSMLPAQDGYENQ  RQMVVAARDGSLVVAQKLVVQCSLQPGHVVKSRWVA  176
Submitted_Seq  118  SSITPAVDGSEFPDQGVVAARDGIVLHPKESQCHTSELQPGHVVKSRWVA  177

Asator-PJ      177  KIGGGGFSIYESQDLITREQVALKVESARQPMQVLMHEVAVLKLQGHVCRFIGQR  236
Submitted_Seq  178  KIGGGGFSIYESQDLITREQVALKVESARQPMQVLMHEVAVLKLQGHVCRFIGQR  237

Asator-PJ      237  NDRFNYYMQLQGNLAELRAQPRGAFSLTTLRLGLQLKATSIHSGFLSRDIKE  296
Submitted_Seq  238  NDRFNYYMQLQGNLAELRAQPRGAFSLTTLRLGLQLKATSIHSGFLSRDIKE  297

Asator-PJ      297  SFVGRLLPYCKRYVIMLDFGLARQITTSIGVBCPRAAGFSTVYASINAHNRKQWR  356
Submitted_Seq  298  SFVGRLLPYCKRYVIMLDFGLARQITTSIGVBCPRAAGFSTVYASINAHNRKQWR  357

Asator-PJ      357  NDLMSLFVMEVFVQQLNWKIKWDQVGLTVEKYDRILLMLPSDLQGLENTQSI  416
Submitted_Seq  358  NDLMSLFVMEVFVQQLNWKIKWDQVGLTVEKYDRILLMLPSDLQGLENTQSI  417

Asator-PJ      417  TYGRPPHYA  LIGLFCMKRRGVKESDFYDWEKVSATIGHISTSGHPPVPVWQYIING  476
Submitted_Seq  418  TYGRPPHYA  LIGLFCMKRRGVKESDFYDWEKVSATIGHISTSGHPPVPVWQYIING  477

Asator-PJ      477  HITQMTVAASGASGVVKKRQDILITATIPMLNKKVQKQCHATSILAQSGSGRF  536
Submitted_Seq  478  HITQMTVAASGASGVVKKRQDILITATIPMLNKKVQKQCHATSILAQSGSGRF  537

Asator-PJ      537  NVQSHAAKNIQITSMG  LQQQSTLT  NSQVAIANIQSAPSM  IEREDV  593
Submitted_Seq  538  NVQLGCTANQNIITPKMLQQQALAINSQAIPIKQSVFISKQVMSHUVQVHTKIS  597

Asator-PJ      594  YTGALGAPTATITMPPRRCOM  VDLAARCTEQ  618
Submitted_Seq  598  QFQKDAASFSSINQHSAPVYQNSYLQLEKQAPNPLTKPNAKSESTVGVAPKSIPEK  657

Asator-PJ      619  RVEAND  DTVGRASISCEVQHYKSDIKKHRSPELAKNDQRTGVINDKTSEVNSSTEE  677
Submitted_Seq  658  PVDSDHACANKAFISHQDQKQKQVQKMLPESANQVQVKKVAVNEKSTLWASQV  717

Asator-PJ      678  KSTFGRRLVLTAFPMQVHLLPSGGHSHVPSLSGMDPLAATNAAPLGKSSSTNY  736
Submitted_Seq  718  KSTFGRRLVLTAFPMQVHLLTGG  HIQQGTULSIMDIP  SSSNAGPAAGSSSSKGLAI  774

Asator-PJ      737  QWQQITFLAMPPINBSATSTNLPSSASQ  KTSQSTIGGAVMGSNTARS  791
Submitted_Seq  775  QWQQITGITLMPQVWBSATSTNLPSSSGGNTPIERINIGSA  GGGGTGSGNTARS  833

Asator-PJ      792  VAGHSVTFQALIDENVSALQVTRGGALLASQWKSQFDSKDTTUNEMWREVSQPN  851
Submitted_Seq  834  VAGHSVTFQALIDENVSALQVTRGGALLASQWKSQFDSKDTTUNEMWREVSQPN  893

Asator-PJ      852  LKQLIKLIDSLMEARFPFQSVAGTGLINFPKANGRPPVTTMTITGKNTYALRL  911
Submitted_Seq  894  LEQLIKLIDFLPMKAKHFCQSVVDTGILTTPPIEGHKKPS  TTMTITGKNTYALRL  953

Asator-PJ      912  IPRCKRFPAMGVLSKGLPFAVQQAIDQVTRMDIARVGVRETYSEITHL  ARPST  969
Submitted_Seq  954  IPRCKRFPAMGVLSKGLPFAVQQAIDQVTRMDIARVGVRETYSDITFLDQKAT  1013

Asator-PJ      970  KSVLPHLSPFPGKSAQLNSTHSLKSRHNSLPYVSVNDIFDQLCKHL  1024
Submitted_Seq  1014  SINRVVLPSPFKEDATPQHTSNISQAKLKHASLPIVSVADLFDQPTSHNSGAIWV  1073

Asator-PJ      1025  LGSATQENCCISGRLEIRVLPKTSMPKQSVYVMMKAVKPTAMRGHNSQAVNK  1084
Submitted_Seq  1074  LSSAVQKNQCSIGRLEIRVLPKTSMLDSVYVYDALAFIKHTITANPDPSISDRANIF  1133

Asator-PJ      1085  DEMATSAVLAEPKSTIKNIPKRNMTAEKTHASLSLYAGVKKIKAG  1136
Submitted_Seq  1134  DEIEKAAVIALPSHTAKCQQTITDKTEACTEANKPSIHATGVNLIKNGKSETCHAL  1193

Asator-PJ      1137  STAP  KIQPKNSTINGKATSEFDFLNPSKI PVQSKCASWAGAD  LSAS  1188
Submitted_Seq  1194  HQQPHYKSGDYKIFFTGGSTDSYRETDGCDLFLNPSKIPIRQSKCASWAGADTVYSS  1253

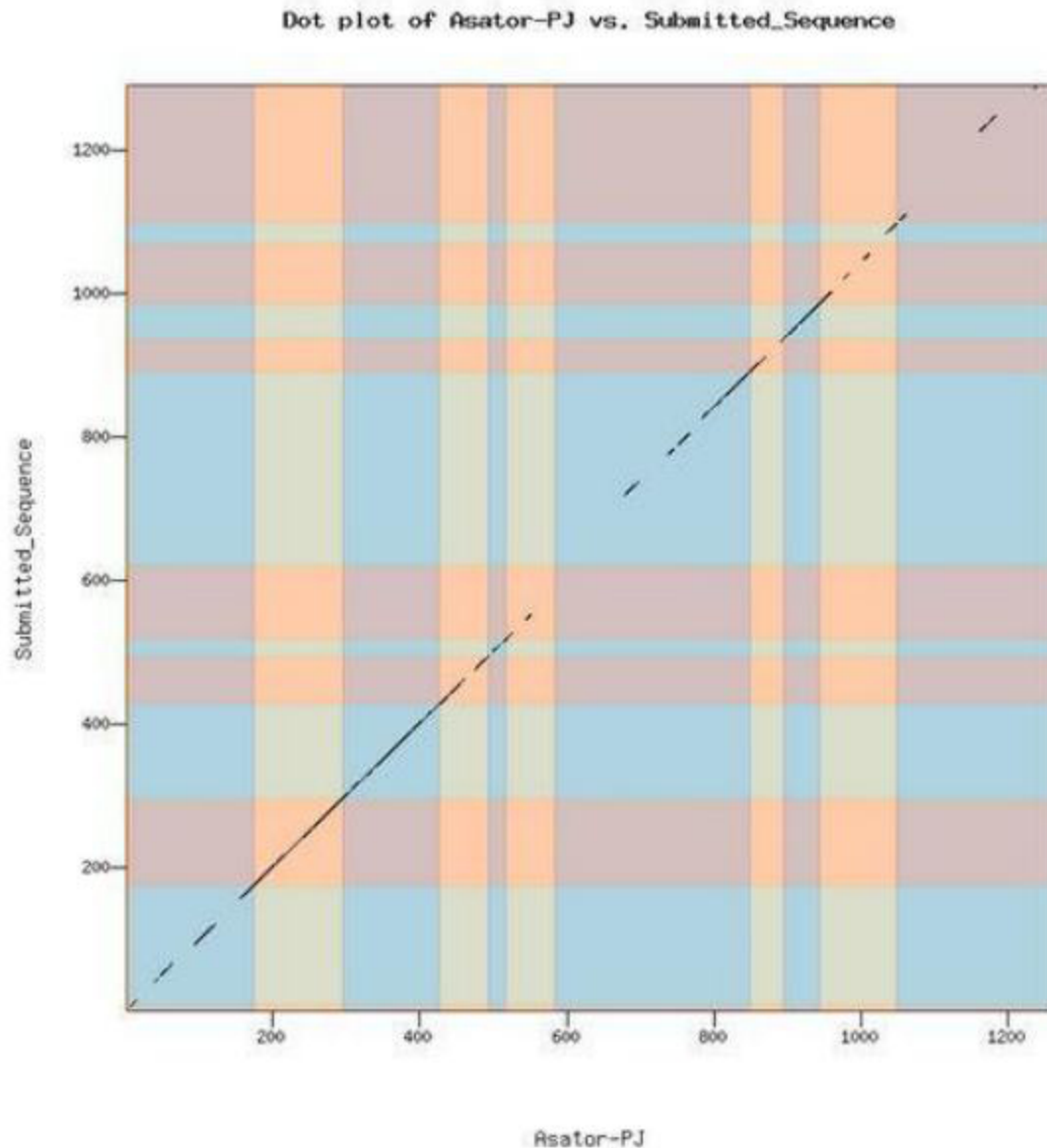
Asator-PJ      1189  KPLASAVPQELPMDQSTHYSVIGIPVAKTYSIALECPMLSDUTGLSTYCNV  1248
Submitted_Seq  1254  KTLFPRDALPEIPNPDQH  TYSTALEYFPHTITLFG  1290

Asator-PJ      1249  VPLRFSTLLTES  1261
Submitted_Seq  1291  1290
```

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.



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---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Pers -----MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Biar MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEYISLNRD---CQKNLF
D.Erec -----
D.Yak -----
D.mel -----MFWHLLCVF--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 MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEYISLNRDCQKNLFRLLHppppskpppLA 180
M + L NDENAS+P DGN +CQP Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct 1 MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPpppskpppLV 58

Query 181 GAILQSRLKQISSGanaenaealee---HRYPNALQRSATLPAKHNLGVRSRVTFKVP 351
GAILQ+RLL QIS A A+ L YPN LQRSATLPAKHNLGVRSRVTFKVP
Sbjct 59 GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNLGVRSRVTFKVP 118

Query 352 SSITPAVDPGS-EPDPGQNQVVAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK 522
SS PA D S +P +NQV A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct 119 SSNLPAQDSYSHQP---RNQVAVAAKDGIIVDVKAKESVKMTSEDLLQPGHVVKERWK 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:**

Configure Gene Model

Model Details

Fosmid Sequence File:

Ortholog in D. melanogaster:

Coding Exon Coordinates:

Annotated Untranslated Regions?
☐ Yes ☒ No

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

Completeness of Gene Model Translation:
☐ Complete ☒ Partial

Region Missing:

Project Details

Project Group:

Project Name:

Checklist

Dot Plot

Transcript Sequence

Peptide Sequence

Extra

Expand All

Collapse All

View	Criteria	Status	Message
	Check for S...	Pass	
	Acceptor fo...	Skip	Already checked for ...
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Acceptor fo...	Pass	
	Donor for C...	Pass	
	Check for S...	Skip	Partial gene with 3' e...
	Additional C...	Pass	

Verify Gene Model

Reset Form...

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:

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%)

```
Asator-PH      1  HTSEDLQPGSVVKKRMVVRKIGGGGFGIITYGQDLITREQVAKVVSARQPKQVLRME 60
Submitted_Seq  1  HTSEDLQPGSVVKKRMVVRKIGGGGFGIITYGQDLITREQVAKVVSARQPKQVLRME 60

Asator-PH      61  VAVLKQLQKKEHVCFICGGRNDRFNVVMQLQGNLAFLELRRAQPRGAFSLSTTLALGLQ 120
Submitted_Seq  61  VAVLKQLQKKEHVCFICGGRNDRFNVVMQLQGNLAFLELRRAQPRGAFSLSTTLALGLQ 120

Asator-PH      121  ILKAIESINSVGLHRDIKPSNFSVGLPYNCRRVYMLDFGLARQYTTGTGEVRCFRAAA 180
Submitted_Seq  121  ILKAIESINSVGLHRDIKPSNFSVGLPYNCRRVYMLDFGLARQYTTGTGEVRCFRAAA 180

Asator-PH      181  GFRSTVRYASINAHNRNEMGRHDLNLSFYMLVEFVNGQLPWRKIKDKEQVGLTEKEDYH 240
Submitted_Seq  181  GFRSTVRYASINAHNRNEMGRHDLNLSFYMLVEFVNGQLPWRKIKDKEQVGLTEKEDYH 240

Asator-PH      241  RILLKHLPSDLQFLEHIQSLTYGDRPDYAMLIGLPERCKRRGVKESDPYDWEKVDSTA 300
Submitted_Seq  241  RILLKHLPSDLQFLEHIQSLTYGDRPDYAMLIGLPERCKRRGVKESDPYDWEKVDSTA 300

Asator-PH      301  IGNISATGMPISIPKSDVMHGNITQMTVAASNASGTYVRKPAEIEIAHITATDPLNIE 360
Submitted_Seq  301  IGNISATGMPISIPKSDVMHGNITQMTVAASNASGTYVRKPAEIEIAHITATDPLNIE 360

Asator-PH      361  FVDKNCNATSLAQFARGSEPMVQNGMAANNQNTIRG-LQQQSTLT-NSQVALAMIQSA 418
Submitted_Seq  361  FVDKNCNATSLAQFARGSEPMVQNGMAANNQNTIRG-LQQQSTLT-NSQVALAMIQSA 420

Asator-PH      419  DSM-----IEREDVQ-----YIKLEEGAPTKFITM 443
Submitted_Seq  421  PIKSPMVQSGSHDQVHTKNSQPQKGAASPSSTNQNNSAPVYGN--YLQLEDKAPNIFLT 480

Asator-PH      444  KPHSECDN-VDIAAKCIFEQKHYEAND-DIVGRASLSGVEQHYKSIKKHNSFEIANKQI 501
Submitted_Seq  481  KPHAESESTVDVAFKSIPEEFVDSNDACAKAFLSGEGQQKSKQVKKLNPESANQQV 540

Asator-PH      502  QRTGTVINDKISEVNRSTEEQKSTFGRLRVLTAPFNSVHDLPSGGHSHQVSDLSQKQDI 561
Submitted_Seq  541  QKLENVANEKSSDKRASQEPKSTFGRLRVLTAPFNSVHDLITGG-HIQGSTOLSIKQDI 599

Asator-PH      562  YAATSNAAPIGINSSTKFG-SQHSQIFGLAAMPFINRSATSTNLRPSSSSAQ-----R 615
Submitted_Seq  600  --SSSNAGPAAGNSSSSKLAINHQGQIFGITIMPQVNRASATSTNLRPSSSSGNTNPIHR 657

Asator-PH      616  INSGSTIGGAVNGSHTARSSVAGDSVITQFALIDENVSALQGVTKGGALTLASQKWSQ 675
Submitted_Seq  658  INIGSA-GGGGSGSHTARSSVAGDSVITQFALIDENVSALQGVTKGGALTLASQKWSQ 716

Asator-PH      676  FDDSEDTIDNENGRHQLQPNLEQLIKLDISLPIAKPPQNGVAGTIGKIINPGEAKG 735
Submitted_Seq  717  FDDSEDTIDNENGRHQLQPNLEQLIKLDISLPIAKPPQNGVAGTIGKIINPGEAKG 776

Asator-PH      736  RPERVTLNITGIENYALRISIPHCNSEFAMGNVLRKLEPPAVQQAADFDRVYRMDIAR 795
Submitted_Seq  777  RPERVTLNITGIENYALRISIPHCNSEFAMGNVLRKLEPPAVQQAADFDRVYRMDIAR 836

Asator-PH      796  NVCVRETTYIITHL--ARPSTSSVLRNRLSPFKKDSALQLNSTNDSLOESRRRNHLPNV 853
Submitted_Seq  837  NVCVRETTYSDITPLQAKPATSLVSRVLPSPFKKDATFQNTSNNSQAKLKHRSILPNV 896

Asator-PH      854  SYNDIFDQLQKMLD-----LQSAIQKNCISGLEIRVIFKTSHPDSDVYDAMGA 908
Submitted_Seq  897  SVADLFDDQPIHNSDAISVALSSAVQENSGCISGLEIRVIFKTSHPDSDVYDAMGA 956

Asator-PH      909  YKNTPTAMGHDMSDQAVNNCEMEATSAVIAPFKSISKHNSPPGRDATEESTGASLCS 968
Submitted_Seq  957  HKNTTTANPDGIDSKANIFCDEIEFNAAVIALPNTANKCKKQTYTDKTEACIFANPCS 1016

Asator-PH      969  LYSAGVNLKLMG-----HTAP-----RTQFKKSGTDGFGENSEFDFLLNPSH 1013
Submitted_Seq  1017  IHATGVNKLKINGSETCNALNDQENYKSGDYKTPFTGGSDTSYRETDGCDLFLNPSK 1076

Asator-PH      1014  IPVRQSKCAGAGADF-ISAKPLESAEVPQEIFYHPQSDTTYSVIDSIFVRKITYSIAL 1072
Submitted_Seq  1077  IPIRQSKCAGAGADTTVYSSKILEPRDALPEIPFNQNT-----YSTAL 1122

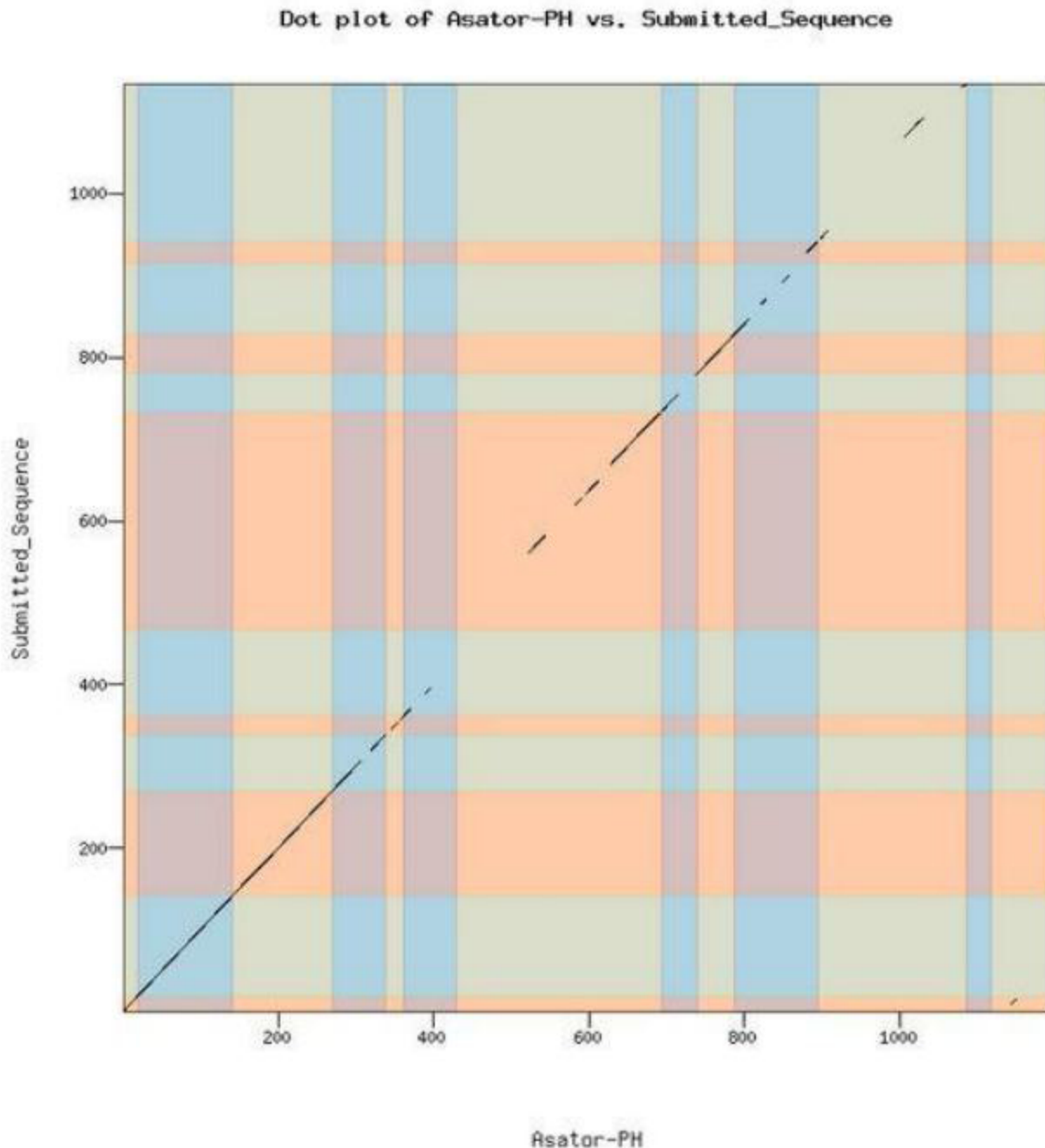
Asator-PH      1073  ECPFNISDLTPQLARRRESTEGYVTDPTQLPLKQAPRSRTSSRTGIFMAKLENSDDH 1132
Submitted_Seq  1123  EYFPNITDLTPG----- 1134

Asator-PH      1133  GTVMSAEKTHGSHVVKTEGVKNFTISNSSEVNPKCTISFPFGDKIKHSARLARVYRHH 1192
Submitted_Seq  1135  ----- 1134
```

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.



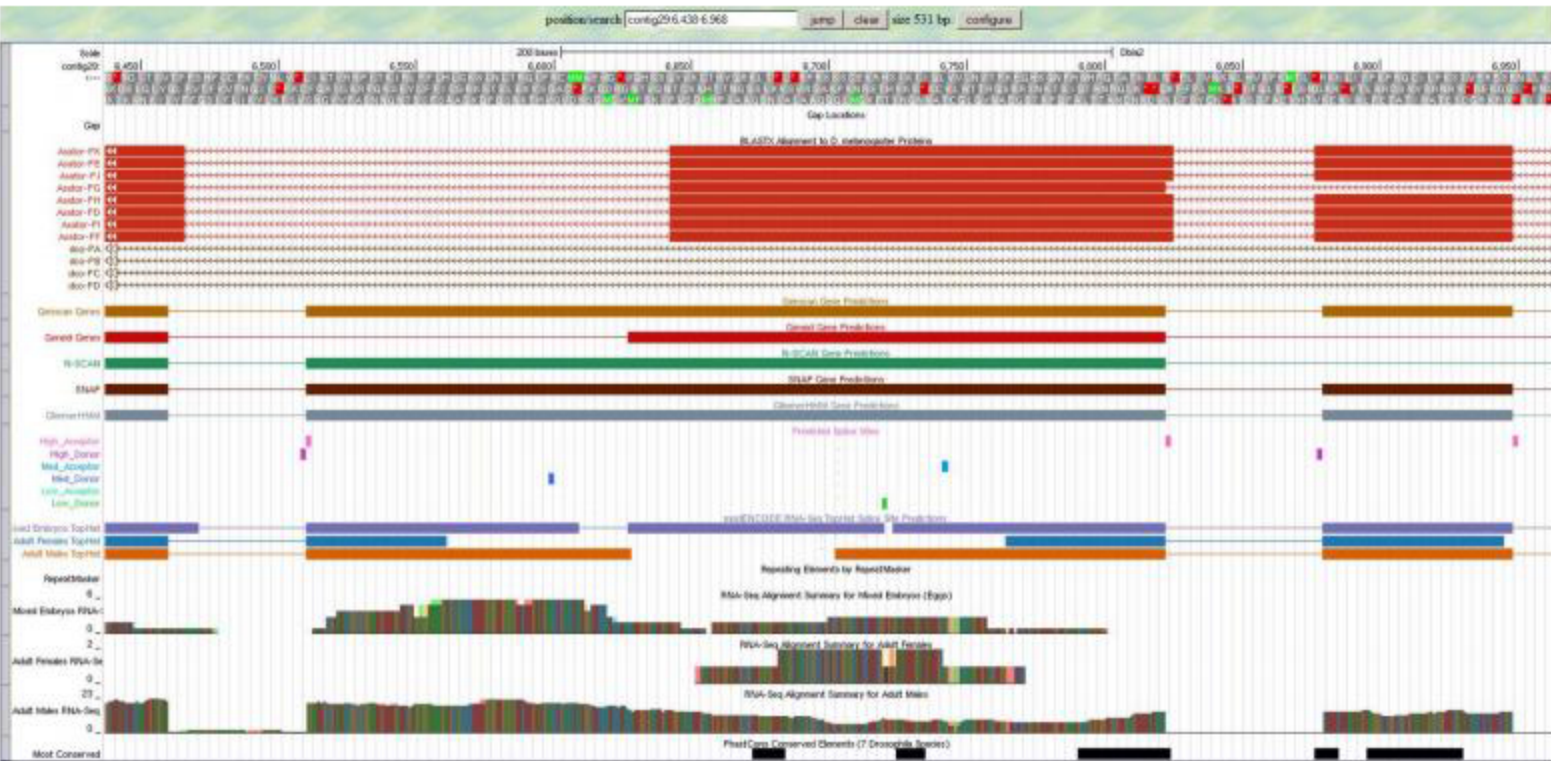
The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in *D. melanogaster*. An open reading frame the entire way through is present as well as computational evidence of tophat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

```

D.Wil  FVDSVLQNVNSGINQNP LPKT-----LAQ-----QQLAAANPNLPSALI
D.Moja SADCVVQNVNVLGNPNLLKHQXXXXXXXXXXXXXXXXXXXXQQLVIVSNTSY-----
D.Viri  TADCAVQNVNNGI--QNT-----TLPKQHQHQHQHQV L VANTSI AIPNLPSALI
D.Grim  TAESALQNSGTLNALPKQ-----QQQQLQHQHQHQV L VNTAIAVANLPSALI
D.Anan  LVEAASQYG--INNQNNLAKELLQ-----QKSCECL-----L
D.pseud KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Pers  KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Erec  SGDAIVPHGNTANNQNITTKRLQ-----STLT--TNSQV--AIPNIQV---
D.Yak   SGEATIQHNTANNQNIKTGLQ-----STMT--TNSQV--AIPNIQSAPI
D.mel   SGEPMVQHGNNAANNQNTSKGLQ-----STL--TNSQV--AIANIQSAP-
D.Sim   SGEALIQHNTANNQNTSKGLQ-----STL--TNSQV--AIPNIQSAPI
D.      SGEALTQHNTANNQNTSKGLQ-----STL--TNSQV--AIPNIQSAPI
:
D.Wil  KSNHENPAAI-----VGQIKSLHGHTAIHTKKCPHVGTGSLISISNQNNSAPVYQYPHA
D.Moja
D.Viri  KSST-VGI VNEHVS GSTHNAQLKST----HGKRCQPQT VSGSYTASNQNNSAPVY-YPES
D.Grim  KPSTVIGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQY P ES
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---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Pers -----MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Biar MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEYISLNRD---CQKNLF
D.Erec -----
D.Yak -----
D.mel -----MFWHLLCVF--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 MQFLLFFRNNDENASSPGDGNHNHTCQPPCNQEYISLNRDCQKNLFRLLHppppskpppLA 180
M + L NDENAS+P DGN +CQP Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct 1 MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPpppskpppLV 58

Query 181 GAILQSRLKQISSGanaenaealee---HRYPNALQRSATLPAKHNLGVRSRVTFKVP 351
GAILQ+RLL QIS A A+ L YPN LQRSATLPAKHNLGVRSRVTFKVP
Sbjct 59 GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNLGVRSRVTFKVP 118

Query 352 SSITPAVDPGS-EPDPGQNQVVAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK 522
SS PA D S +P +NQV A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct 119 SSNLPAQDSYSHQP---RNQVAVAAKDGIIVDVKAKESVKMTSEDLLQPGHVVKERWK 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:**

Configure Gene Model

Model Details

Fosmid Sequence File:

Ortholog in D. melanogaster:

Coding Exon Coordinates:

Annotated Untranslated Regions? ☐ Yes ☒ No

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

Completeness of Gene Model Translation: ☐ Complete ☒ Partial

Region Missing:

Project Details

Project Group:

Project Name:

Checklist | Dot Plot | Transcript Sequence | Peptide Sequence | Extracted Coding Exons

Expand All | Collapse All

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	Pass	
Acceptor for CDS 11	Pass	
Donor for CDS 11	Pass	
Acceptor for CDS 12	Pass	
Donor for CDS 12	Pass	
Check for Stop Codon	Skip	Partial gene with 3' end missing
Additional Checks	Pass	

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:

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-PD vs. Submitted_Seq

[View plain text version](#)

Identity: 917/1421 (64.5%), Similarity: 1018/1421 (71.6%), Gaps: 203/1421 (14.3%)

```
Asator-PD      1  KTMILLCEVSRKANAPGSGGCS--CPSSMDQVYAPNNCONLLATYPPPPKPPF  58
Submitted_Seq  1  HQF-LLFRDNCASSPFGNNHNTQPPCNQEQVLSNRDQKQLFLRPPPPKPPF  59

Asator-PD      59  AGATITELTDTSPSLADGAGAAWELLVYNNQKATLPAKNNKLVKSRVTFN  118
Submitted_Seq  60  AGATLQSKLLHQISGGAGK---KAAALEHRYFIALQKATLPAKNNKLVKSRVTFN  116

Asator-PD      119  PSSSLFAGQSYSGG--RNQVAAAKGGLAVNNAKESVSMTEGLQVSRVYQKQV  176
Submitted_Seq  117  PSSITFAVDPSGEPDQGNQVAAERDSTVLNPHRESCHMTSEDLQPSVYKRNQV  176

Asator-PD      177  KTKGGGSGEIVSQQLITREQVALKVESARQKQVLAHVAVLQKQKERVCFISG  236
Submitted_Seq  177  KYGGGGSGEIVYKQQLITREQVALKVESARQKQVLAHVAVLQKQKERVCFISG  236

Asator-PD      237  RNRDITVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  296
Submitted_Seq  237  RNRDITVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  296

Asator-PD      297  SNRFPVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  356
Submitted_Seq  297  SNRFPVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  356

Asator-PD      357  RNDLNSLFLVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  416
Submitted_Seq  357  RNDLNSLFLVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  416

Asator-PD      417  LTYADRPVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  476
Submitted_Seq  417  LTYADRPVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGLHRIE  476

Asator-PD      477  QNTTQCTVAASASGTE--TDDALHAYATATDPCALHVDONCHATSIAQKASGE  536
Submitted_Seq  477  QNTTQCTVAASASGTEYVSSGDIETATITATEPETHQVVDONCHATSIAQKASGE  536

Asator-PD      537  RNVQGRUANNQHTSKS--LQQSTLT--NSQVALANIQKASG-----IERKQV  594
Submitted_Seq  537  RNVQGRUANNQHTSKS--LQQSTLT--NSQVALANIQKASG-----IERKQV  594

Asator-PD      595  YTKLFGAPRTITMPPHRCES--VHIAAKTFK  618
Submitted_Seq  595  YTKLFGAPRTITMPPHRCES--VHIAAKTFK  618

Asator-PD      619  KVVLEAD--DLYRANSLQVLAATCS--LQNNQVYFANRQLQDITVYVNLQKQDAM  677
Submitted_Seq  619  KVVLEAD--DLYRANSLQVLAATCS--LQNNQVYFANRQLQDITVYVNLQKQDAM  677

Asator-PD      678  VSTTGLKLVLTAPPNVNQLPQGGGKQVNLQKQDAMLTAAATG--NLSSTKTK  737
Submitted_Seq  678  VSTTGLKLVLTAPPNVNQLPQGGGKQVNLQKQDAMLTAAATG--NLSSTKTK  737

Asator-PD      738  QVNDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  791
Submitted_Seq  738  QVNDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  791

Asator-PD      792  VAGDSVITQALDCEINVALQVQVMSGSLTASQKSGFQDSQEDITDHEWREPSGE  851
Submitted_Seq  792  VAGDSVITQALDCEINVALQVQVMSGSLTASQKSGFQDSQEDITDHEWREPSGE  851

Asator-PD      852  NLEQLTSLTSLPQKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV  911
Submitted_Seq  852  NLEQLTSLTSLPQKQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV  911

Asator-PD      912  STPKCTYFAMQVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGL  969
Submitted_Seq  912  STPKCTYFAMQVYVNLQKQDAMLRQAQKQKATSLSTTLRSLQKLPATESTRVGL  969

Asator-PD      970  TSSVLAHLSPFFQDQALQSTHSLKSRHSLPVSVDIFDQLQKQDAM----  1025
Submitted_Seq  970  TSSVLAHLSPFFQDQALQSTHSLKSRHSLPVSVDIFDQLQKQDAM----  1025

Asator-PD      1026  LSSATGNNCCISGLRIVIPK--TSPHDSVYVNLQKQDAMLTAAATG--NLSSTKTK  1084
Submitted_Seq  1026  LSSATGNNCCISGLRIVIPK--TSPHDSVYVNLQKQDAMLTAAATG--NLSSTKTK  1084

Asator-PD      1085  TQVATSAVIAFNNKIS--KSPYDQATYKATGSLTSLPQKQVQVQVQVQVQVQV  1143
Submitted_Seq  1085  TQVATSAVIAFNNKIS--KSPYDQATYKATGSLTSLPQKQVQVQVQVQVQV  1143

Asator-PD      1144  STAP--FVQVNDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1203
Submitted_Seq  1144  STAP--FVQVNDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1203

Asator-PD      1204  KPLKSAKVPQIPYVNDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1263
Submitted_Seq  1204  KPLKSAKVPQIPYVNDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1263

Asator-PD      1264  KQVVDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1323
Submitted_Seq  1264  KQVVDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1323

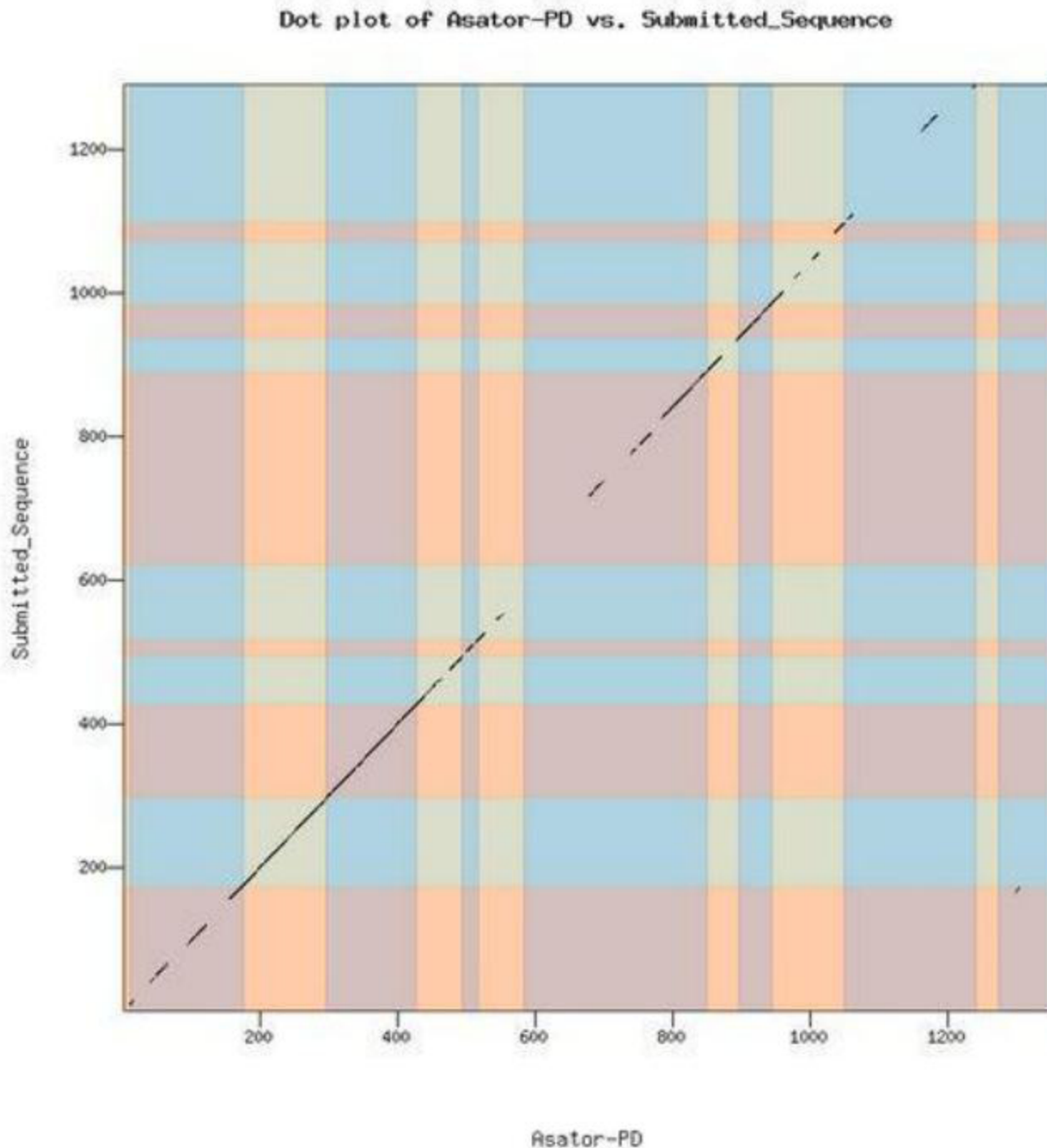
Asator-PD      1324  KQVVDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1383
Submitted_Seq  1324  KQVVDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1383

Asator-PD      1384  KQVVDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1443
Submitted_Seq  1384  KQVVDITGLAAMPVNRASATSLPSSSANG--RISDSTIGKAVGNNHTAS  1443
```

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.



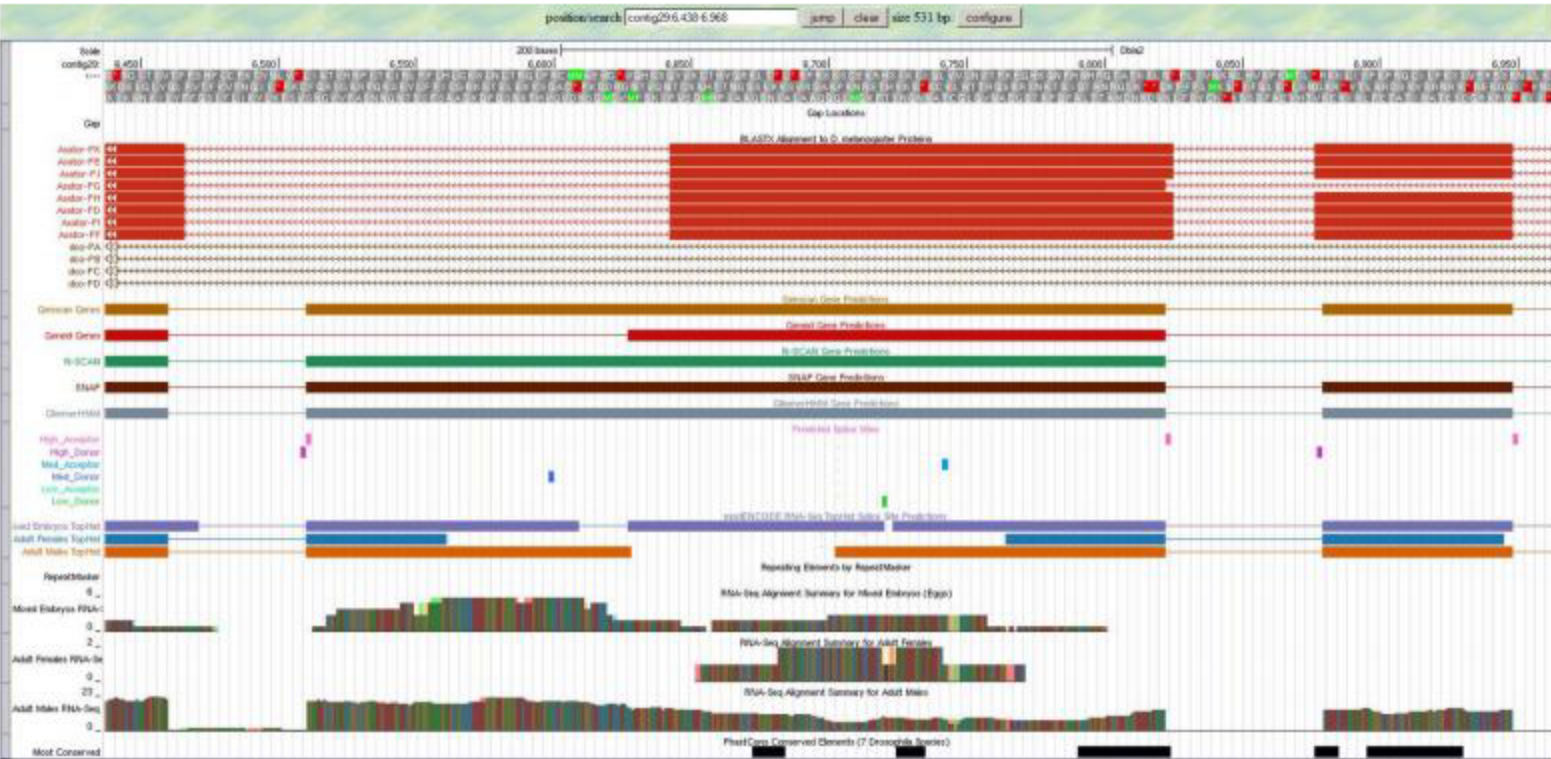
The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in *D. melanogaster*. An open reading frame the entire way through is present as well as computational evidence of tophat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

```

D.Wil  FVDSVLQNVNSGINQNPFPKT-----LAQ-----QQLAAANPNLPSALI
D.Moja SADCVVQNVNLGNPNLLKHQ-----QQLVIVSNTSY-----
D.Viri TADCAVQNVNSGI--QNT-----TLPKQHQ-----QQLVIVSNTSY-----
D.Grim TAESALQNSGTLNALPKQ-----Q-----QQLVIVSNTSY-----
D.Anan LVEAASQYG--INNQNNAKELLQ-----QKSCECL-----L
D.pseud KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Pers  KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Erec  SGDAIVPHGNTANNQNTITKRLQ-----STLT--TNSQV--AIPNIQV--
D.Yak   SGEATIQHGNTANNQNTITKGLQ-----STMT--TNSQV--AIPNIQSAPI
D.mel   SGEPMVQHGNNAANNQNTITKGLQ-----STL--TNSQV--AIPNIQSAP-
D.Sim   SGEALIQHGNTANNQNTITKGLQ-----STL--TNSQV--AIPNIQSAPI
D.      SGEALTQHGNTANNQNTITKGLQ-----STL--TNSQV--AIPNIQSAPI
:
D.Wil  KSNHENPAAI-----VGQIKSLHGHTAIHTKKCPHVGTGSLISNQNSAPVYQYPHA
D.Moja
D.Viri KSST-VGIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNSAPVY-YPES
D.Grim KPSTVIGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNSAPVYQYPS
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---RDHDSHQQHSNQELFISPSNVSSSSSKNNI
D.Pers -----MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNVSSSSSKNNI
D.Biar MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEYISLNRD---CQKNLF
D.Erec -----
D.Yak -----
D.mel -----MFWHLLCVF--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 MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEYISLNRDCQKNLFRLLHppppskpppLA 180
M + L NDENAS+P DGN +CQP Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct 1 MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPpppskpppLV 58

Query 181 GAILQSRLKQISSGanaenaealee---HRYPNALQRSATLPAKHNLGVRSRVTFKVP 351
GAILQ+RLL QIS A A+ L YPN LQRSATLPAKHNLGVRSRVTFKVP
Sbjct 59 GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNLGVRSRVTFKVP 118

Query 352 SSITPAVDPGS-EPDPGQNQVVAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK 522
SS PA D S +P +NQV A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct 119 SSNLPAQDSYSHQP---RNQVAVAAKDGIIVDVKAKESVKMTSEDLLQPGHVVKERWK 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:**

Configure Gene Model

Model Details

Fosmid Sequence File:

Ortholog in D. melanogaster:

Coding Exon Coordinates:

Annotated Untranslated Regions?
☐ Yes ☒ No

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

Completeness of Gene Model Translation:
☒ Complete ☐ Partial

Stop Codon Coordinates:

Project Details

Project Group:

Project Name:

Checklist

Dot Plot

Transcript Sequence

Peptide Sequence

Extra

Expand All

Collapse All

Vi...	Crit...	Status	Message
	Ch...	Pas	
	Ac...	Skip	Already checked for Start Codon
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Pas	
	Ac...	Pas	
	Do...	Skip	Already checked for Stop Codon
	Ch...	Pas	
	Ad...	Pas	
	Nu...	War	Gene model has 10 CDS's, ortholog has 9 CDS's

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:

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  MTSEDLLQPGHVVKERWVVRKIGGGGFGIYEGQDLITREQVALKVESARQPKQVLKME  60
Submitted_Seq  1  MTSEDLLQPGHVVKERWVVRKIGGGGFGIYEGQDLITREQVALKVESARQPKQVLKME  60

Asator-PF      61  VAVLKKLQGGKEHVCRFIGCGRNDRFNYVVMQLQGGKLAELRRAQPRGAFSLSTTLRLGLQ  120
Submitted_Seq  61  VAVLKKLQGGKEHVCRFIGCGRNDRFNYVVMQLQGGKLAELRRAQPRGAFSLSTTLRLGLQ  120

Asator-PF      121  ILKAIESIHSVGFLLHRDIKPSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA  180
Submitted_Seq  121  ILKAIESIHSVGFLLHRDIKPSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA  180

Asator-PF      181  GFRGTVRYASINAHNRNEMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKQVGLTKEKYDH  240
Submitted_Seq  181  GFRGTVRYASINAHNRNEMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKQVGLTKEKYDH  240

Asator-PF      241  RILLKHLPSDLKQFLEHIQSLTYGDRPDYAMLIGLFFERCMKRRGVKESDPYDWEKVDSTA  300
Submitted_Seq  241  RILLKHLPSDLKQFLEHIQSLTYADRPDYAMLIGLFFERCMKRRGVKESDPYDWEKVDSTA  300

Asator-PF      301  IGNISATGNPSIPIKSDYMHGNIQTMTVAASNASGTEYIRKRAEIIETAHITATDPLNIKE  360
Submitted_Seq  301  IGNISTSGNPPVPVKNDYIHGNITQMTVAASNASGTEYVRKRGDIETAHITATEPLHIKE  360

Asator-PF      361  KVDKNCNATSLAQPAKSGCEPMVQHGNAANNQNIITSKG-LQQQSTLT-NSQVAIANIQSA  418
Submitted_Seq  361  KVDKNCNATTILAFQPKTSGEANVQLGCTANNQNIITPKGMLQQAALAINSQAAIPHMQSV  420

Asator-PF      419  PSM-----IEREDVQ-----YTKLEEGAPTKFITM  443
Submitted_Seq  421  PIKSPVMVGSHDVQVHTKNSQPQKGAASFSSSTNQNNAPVYGNSTLQLEDKAPNMFLEP  480

Asator-PF      444  KPNGECDN-VDIAAKCIFEQKHVEAND-DIVGRASLSGVEQHYKSIQIKHNSPEIANKQI  501
Submitted_Seq  481  KPNAEESTVDVAPKSIFFEEKFVDSNDAECANKAFLSGEQQQKSQVKKLNLPESANQQV  540

Asator-PF      502  QRTGTIVNDKISEVNRSTEEQKSTFGRLRVLTAPFMSVHDLFSGGGHSHQVSDLSGKQDE  561
Submitted_Seq  541  QKLENVANEKSSSEDKRASQEPKSTFGRLRVLTAPFMSVHDLTTGG-HIQQGTDLISIKQDE  599

Asator-PF      562  YAATSNAAPIGINSSSTKFG-SQHGQIFGLAAMPPIINRRSATSTNLRPSSSSASQ-----R  615
Submitted_Seq  600  --SSSNAGPAAGNSSSSKLAINQHGGQIFGITLMPQVNRSATSTNLRPSSSGGNTNPIHR  657

Asator-PF      616  INSGSTIGGAVGNGSNTARSSVAGDHSVTQFALIDDENVSALQQVTKGGALTILASQWKSQ  675
Submitted_Seq  658  INIGSA-GGGGGTGSNTARSSVAGDHSVTQFALIDDENVSALQQVTKGGALTILASQWKSQ  716

Asator-PF      676  FDDSEDITDNEWNREHQLQPNLEQLIKLDIPLPLNEAKFPQHGVAAGTKLINPPGEAKG  735
Submitted_Seq  717  FDDSEDITDNEWNREPCSQPNLEQLIKLDIPLPLNEAKHFCQNVVITDGLITKPIEGNE  776

Asator-PF      736  RPKRYTLNITGIENYEALRISIPNCWSEPAMGNVLRKGLEPPAVQQAADFDDTVSILNKS  795
Submitted_Seq  777  KPKRYTLNITGIENYEALRISIPNCWSEPAMGNVLRKLDLEPPAVQQAADFDDTNSLNN---  833

Asator-PF      796  EVIIYIMRAVTFCE  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  MTSEDLLQPGHVVKERKVVVRKIGGGGFGIEYEGQDLITREQVALKVESARQPKQVLKME 60
Submitted_Seq  1  MTSEDLLQPGHVVKERKVVVRKIGGGGFGIEYEGQDLITREQVALKVESARQPKQVLKME 60

Asator-PF     61  VAVLKKLQGGKEHVCRFIGCGRNDRFNYVVMQLQGKNAELRRAQPRGAFSLSTTLRLGLQ 120
Submitted_Seq 61  VAVLKKLQGGKEHVCRFIGCGRNDRFNYVVMQLQGKNAELRRAQPRGAFSLSTTLRLGLQ 120

Asator-PF    121  ILKAIESIHSVGFLLHRDIKSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA 180
Submitted_Seq 121  ILKAIESIHSVGFLLHRDIKSNFSVGRLPYNCRRVYMLDFGLARQYTTGTGEVRCPRAAA 180

Asator-PF    181  GFRGTVRYASINAHNRNEMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKEQVGLTKEKYDH 240
Submitted_Seq 181  GFRGTVRYASINAHNRNEMGRHDDLWSLFYMLVEFVNGQLPWRKIKDKEQVGLTKEKYDH 240

Asator-PF    241  RILLKHLPSDLKQFLEHIQSLTYGDRPDYAMLIGLFCMKRRGVKESDPYDWEKVDSTA 300
Submitted_Seq 241  RILLKHLPSDLKQFLEHIQSLTYADRPDYAMLIGLFCMKRRGVKESDPYDWEKVDSAT 300

Asator-PF    301  IGNISATGNPSIPIKSDYMHGNIQTMTVAASNASGTEYIRKRAEIEAHITATDPLNIKE 360
Submitted_Seq 301  IGNISTSGNPPVPVKNDYIHGNIQTMTVAASNASGTEYVRKRGDIETAHITATEPLHIKE 360

Asator-PF    361  KVDKNCNATSLAQPAKSGSEPMVQHGNNAANNQNIISKG-LQQSTLT-NSQVAIANIQSA 418
Submitted_Seq 361  KVDKNCNATTIAPQPKTSGEANVQLGCTANNQNIITPKGMLQQQAALAINSQAAIPHMQSV 420

Asator-PF    419  PSM-----IEREDVQ-----YTKLEEGAPTKFITM 443
Submitted_Seq 421  PIKSPVMVGMSHDVQVHTKNSQPQKGAASFSTNQNNASAPVYGNSTYLQLEDKAPNMFLEP 480

Asator-PF    444  KPNGECDN-VDIAAKCIFEQKHVEAND-DIVGRASLSGVEQHYKSQIKKHNSPEIANKQI 501
Submitted_Seq 481  KPNAESESTVDVAPKSIFFEEKFVDSNDAECANKAFLSGEQQQKSQVKKLNLPEANQQV 540

Asator-PF    502  QRTGTVINDKTSEVNRSTEEQKSTFGRLRLVTAPPMSVHDLPSGGGSHSHQVSDLSGKQDF 561
Submitted_Seq 541  QKLENVANEEKSSDKRASQEPKSTFGRLRLVTAPPMSVHDLTTGG-HIQGGTDLISIKQDF 599

Asator-PF    562  YAATSNAAPIGINSSSTKFG-SQHGQIFGLAAMPFINRRSATSTNLRPSSSSASQ-----R 615
Submitted_Seq 600  --SSSNAGPAAGNSSSSKLAINQHGGQIFGITLMPQVNRSATSTNLRPSSSGGNTNPIHR 657

Asator-PF    616  INSGSTIGGAVGNGSNITARRSSVAGDHSTVTFALIDDENVSALQQVTKGGALTLASQWKSQ 675
Submitted_Seq 658  INIGSA-GGGGGTGSNITARRSSVAGDHSTVTFALIDDENVSALQQVTKGGALTLASQWKSQ 716

Asator-PF    676  FDDSEDITDNEWNREHQLQPNLEQLIKLDIPLPLNEAKFPQHGVAAGTKGLINPPGEAKG 735
Submitted_Seq 717  FDDSEDITDNEWNREPQSQPNLEQLIKLDIPLPLNEAKHFCQNVVTDITGILTKPIEGNE 776

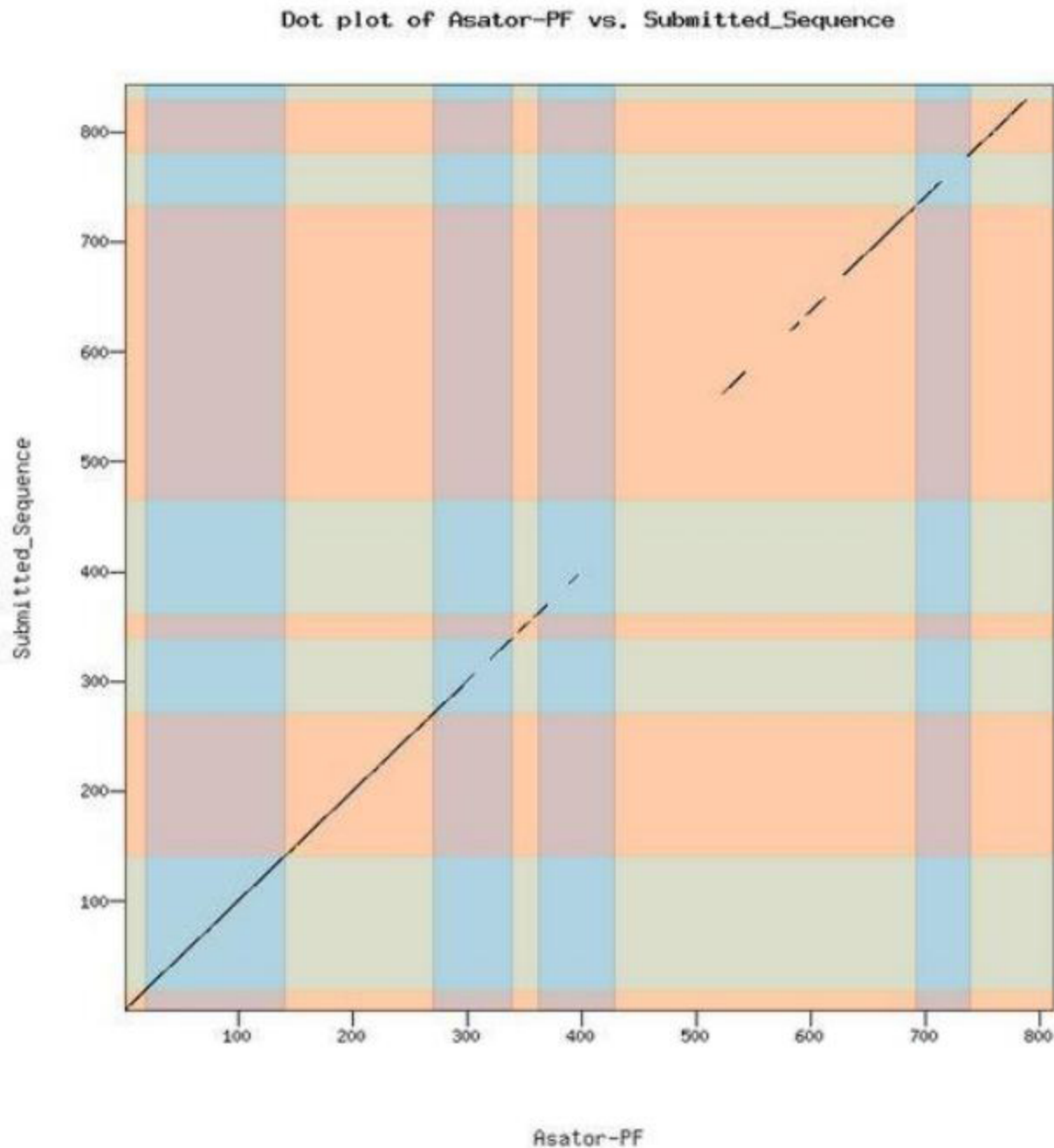
Asator-PF    736  RPKRYTLNITGIENYEALRISIPNCWSEPAMGNVLRKGLEPPAVQQAADFDDTVSILNKS 795
Submitted_Seq 777  KPKRYTLNITGIENYEALRISIPNCWSEPAMGNVLRKLDLEPPAVQQAADFDDTVSLNN--- 833

Asator-PF    796  FVLIYYMHAVTFPCE 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.



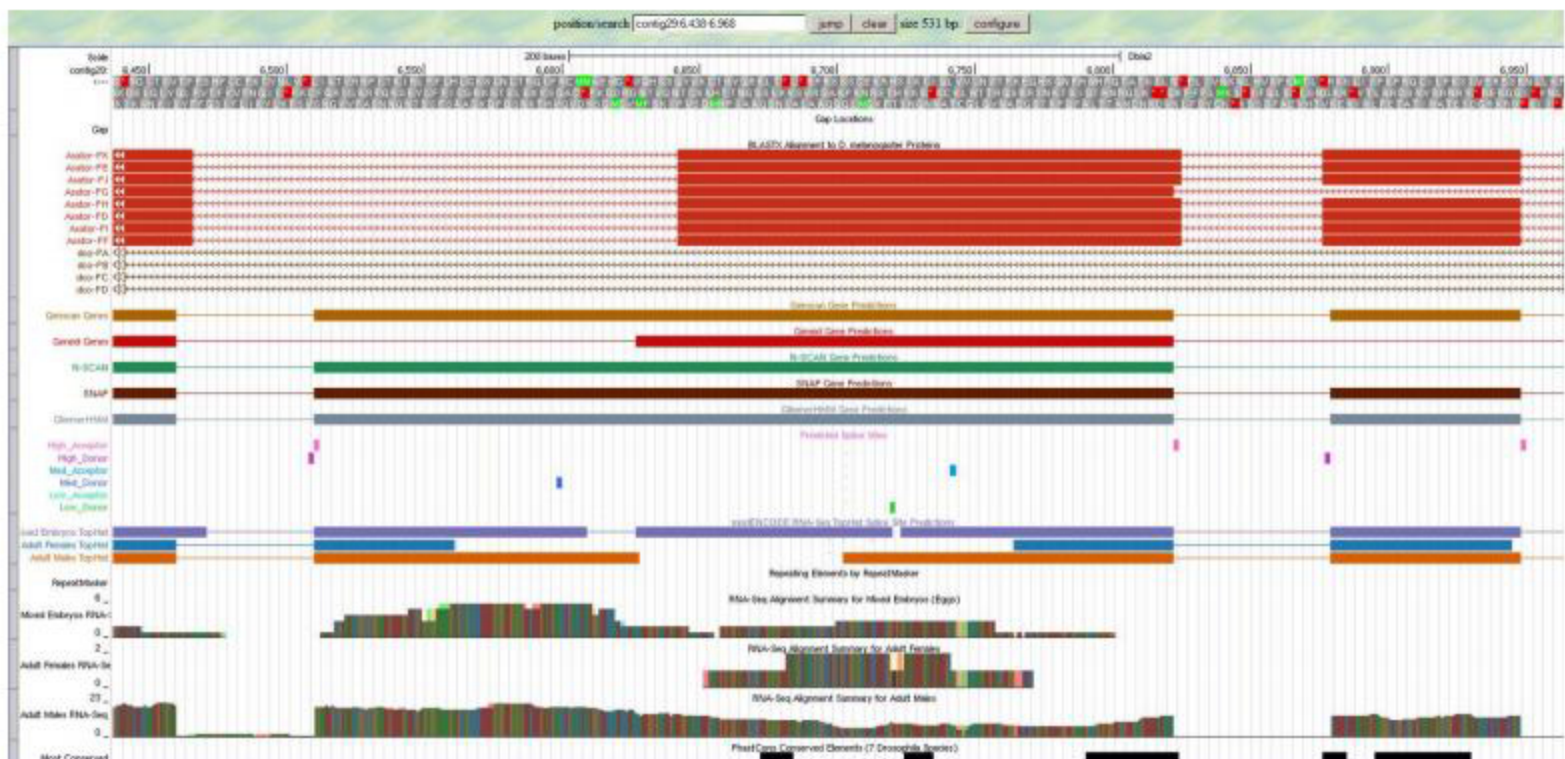
The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in *D. melanogaster*. An open reading frame the entire way through is present as well as computational evidence of tophat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

```

D.Wil  FVDSVLQNVNSGINQNPFPKT-----LAQ-----QQLAAANPNLPSALI
D.Moja SADCVVQNVNLGNPNLLKHQXXXXXXXXXXXXXXXXXXXXQQLVIVSNTSY-
D.Viri TADCAVQNVNSGI--QNT-----TLPKQHQHQHQHQVULVANTSIAIPNLPSALI
D.Grim TAESALQNSGTLNALPKQ-----QHQHQHQHQHQHQVULVNTAIAVANLPSALI
D.Anan  LVEAASQYG--INNQNNLAKELLQ-----QKSCECL-----L
D.pseud KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Pers  KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Erec  SGDAIVPHGNTANNQNTITKRLQ-----STLT--TNSQV--AIPNIQV--
D.Yak   SGEATIQHNTANNQNIKTGLQ-----STMT--TNSQV--AIPNIQSAPI
D.mel   SGEPMVQHNAANNQNTITKGLQ-----STL--TNSQV--AIANIQSAP-
D.Sim   SGEALIQHNTANNQNTITKGLQ-----STL--TNSQV--AIPNIQSAPI
D.      SGEALTQHNTANNQNTITKGLQ-----STL--TNSQV--AIPNIQSAPI
:
D.Wil  KSNHENPAAI-----VGQIKSLHGHTAIHTKKCPHVGTGSLISISNQNNSAPVYQYPHA
D.Moja
D.Viri KSST-VGIIVNHEVSGSTHNAQLKST----HGKRCQPQTVSGSYTASNQNNSAPVY-YPES
D.Grim KPSTVIGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQYPS
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---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Pers -----MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Biar MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEYISLNRD---CQKNLF
D.Erec -----
D.Yak -----
D.mel -----MFWHLLCVF--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} .

```

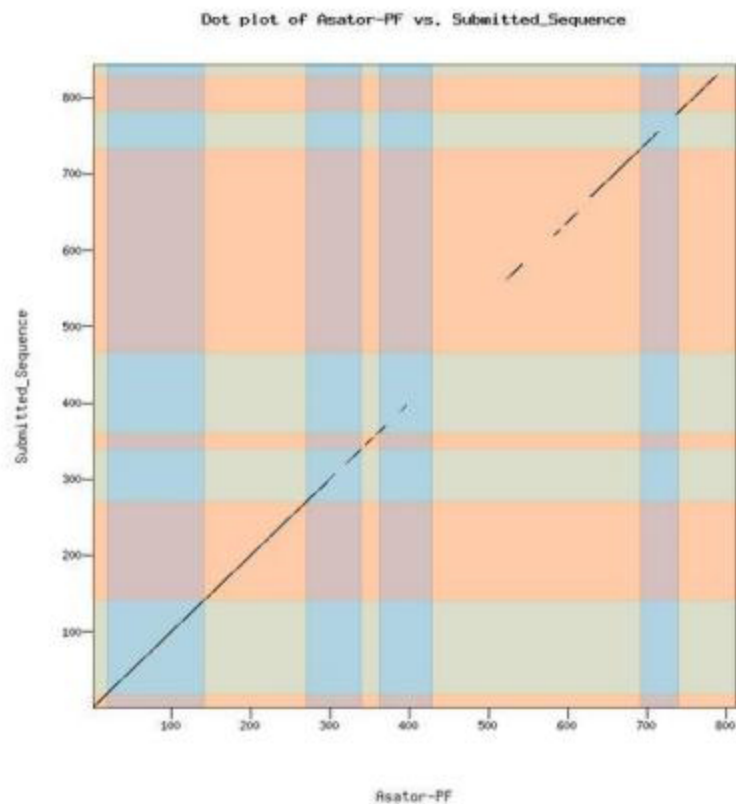
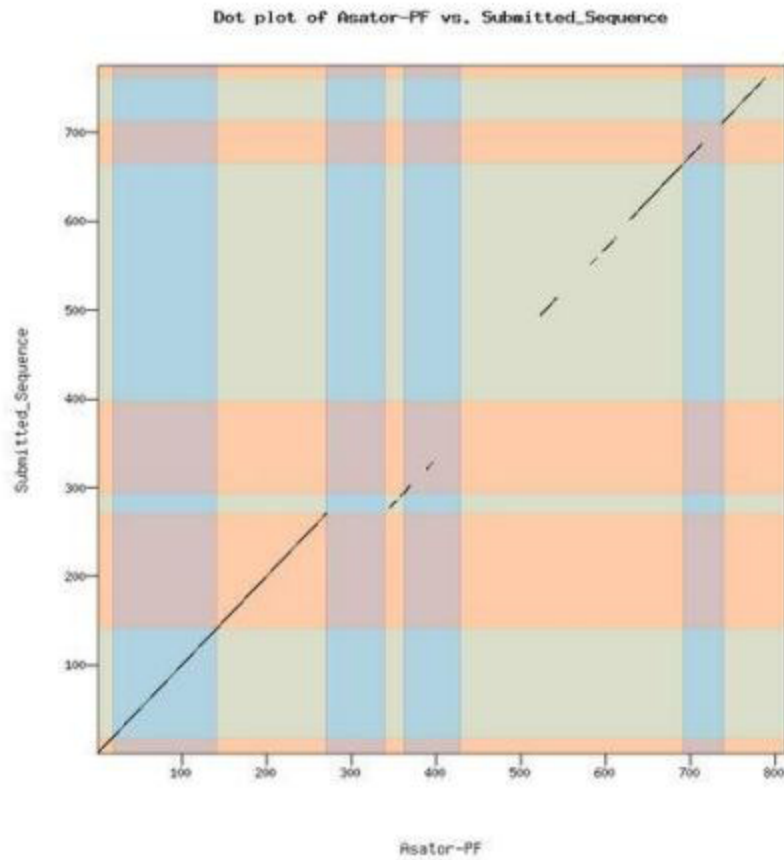
Query 1 MQFLLFFRNDENASSPGDGNHNHTCQPPCNQEYISLNRDCQKNLFRLLHppppskpppLA 180
M + L NDENAS+P DGN +CQP Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct 1 MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPPPPSKPPPLV 58

Query 181 GAILQSRLKQISSGanaenaealee---HRYPNALQRSATLPAKHNLGVRSRVTFKVP 351
GAILQ+RLL QIS A A+ L YPN LQRSATLPAKHNLGVRSRVTFKVP
Sbjct 59 GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNLGVRSRVTFKVP 118

Query 352 SSITPAVDPGS-EPDPGQNQVVAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK 522
SS PA D S +P +NQV A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct 119 SSNLPAQDSYSHQP---RNQVAVAAKDGIIVDVKAKESVKMTSEDLLQPGHVVKERWK 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:**

The screenshot displays the 'Configure Gene Model' window, which is divided into two main sections: 'Model Details' and 'Project Details'.

Model Details:

- Fosmid Sequence File:** C:\fakepath\contig29.fasta (with a 'Browse...' button)
- Ortholog in D. melanogaster:** Asator-PG
- Coding Exon Coordinates:** 11050-10529, 8535-8170, 8057-7668, 7605-7402, 6822-6511, 6460-5657, 1680-1424, 1366-1287, 894-312
- Annotated Untranslated Regions?** ☐ Yes ☒ No
- Orientation of Gene Relative to Query Sequence:** ☐ Plus ☒ Minus
- Completeness of Gene Model Translation:** ☐ Complete ☒ Partial
- Region Missing:** Missing 3' end of translated region (selected from a dropdown menu)

Project Details:

- Project Group:** D. biarmipes Dot (selected from a dropdown menu)
- Project Name:** contig29

Checklist:

The 'Checklist' tab is active, showing a table of criteria and their status. The table has columns for 'View', 'Criteria', 'Status', and 'Message'.

View	Criteria	Status	Message
	Check for Start Co...	Pass	
	Acceptor for CDS 1	Skip	Already checke...
	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	
	Check for Stop Co...	Skip	Partial gene wit...
	Additional Checks	Pass	

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:

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%)

```
Asator-PG      1 MFKHLLCVNDENASAPINGNQS--CQSSKQVQLSPWRCQGNLLGLYPPPTSKVPPV 58
Submitted_Seq  1 MQF-LLFRRNDEHASSPGDGNHHTCQSPCNQEQYISLHRDCQNLFRHLFFPFSKFPFL 59

Asator-PG      59 VEA LLNTRALLQISPSA LADAHNUNAVGELLYPVLQSSATLPADNRGLGVRKVTYKY 118
Submitted_Seq  60 AGAILQSRLLKQISSGANAL---NAEALFEHRYPNALQRSATLPKHNRLGVRKVTYKY 116

Asator-PG      119 PSSNLPADSTESMQ--RQVVAANDGTVGVNANQSVNMTSLDLQSSVWQEMRV 176
Submitted_Seq  117 PSSITPAVDPGSEDPGQNVVAERDGVLPKRESCQMTSEDLQPGHVKERNV 176

Asator-PG      177 RKIGGGGFGIYEGQQLITREQVALKVESARQPKVLMQHVAVLQKQKENVCFIQC 236
Submitted_Seq  177 RKIGGGGFGIYEGQQLITREQVALKVESARQPKVLMQHVAVLQKQKENVCFIQC 236

Asator-PG      237 RNDRFYVVMQLQGKHLAELRAQPRGAFSLSTTLRLGLQILKATESIRSVGLHNDIK 296
Submitted_Seq  237 RNDRFYVVMQLQGKHLAELRAQPRGAFSLSTTLRLGLQILKATESIRSVGLHNDIK 296

Asator-PG      297 SNFSVGRLPYKCPVVMLOFGLARYTTIGTEVRCPRAAAGFRTVRYASINAHREMG 356
Submitted_Seq  297 SNFSVGRLPYKCPVVMLOFGLARYTTIGTEVRCPRAAAGFRTVRYASINAHREMG 356

Asator-PG      357 RNDLNSLFLYKVEFVNGQLPWRKNDKQVGLTKKDYDRILLKHLPSDLKQFLHQS 416
Submitted_Seq  357 RNDLNSLFLYKVEFVNGQLPWRKNDKQVGLTKKDYDRILLKHLPSDLKQFLHQS 416

Asator-PG      417 LTYGSRPOVA LIGLFEHCKRRGVKESDFYDWEKVDSTAGNISATGNPSIPKSDVM 476
Submitted_Seq  417 LTYGSRPOVA LIGLFEHCKRRGVKESDFYDWEKVDSTAGNISATGNPSIPKSDVM 476

Asator-PG      477 GNITQMTVAASNASGTEYVDMKCNATSLAQANSGSLPMVQNGNAANPNITSE-1000 535
Submitted_Seq  477 GNITQMTVAASNASGTEYVDMKCNATSLAQANSGSLPMVQNGNAANPNITSE-1000 536

Asator-PG      536 STLT KSVATANIQAPSM-----TEKENV----- 561
Submitted_Seq  537 AALAINSQALPMMQSVFIKSPMVGMGSHGVVNTKISQPKGAASFSSIHQNISAFVYG 596

Asator-PG      562 --VTKLEGAFTKFTMKPMGECN-VDIAACIFEQKQVAND-DIVERASLSGVEQHY 617
Submitted_Seq  597 NSTYLQLEKAPNMLPTKPNASESTVINAPKSIPEKTVDSNDACANKATLSGQQQQ 656

Asator-PG      618 KQIKKHNSPEIANMQIQRTIGVINDKTSVIRSTEEQKSTFGRLRVLTAPPMSVHQLPS 677
Submitted_Seq  657 KQVQKGLPESANQQVQLENVANEKSSDQKGAQPKSTFGLRVLTAPPMSVHQLPS 716

Asator-PG      678 GGHSHQVSDLSGMDQFYAATSHAAFIGINSSSTKFG-SQHQIFGLAAMPFINRRSATS 736
Submitted_Seq  717 GG-HHQGTQLSIKQDF--SSNAGPAAGNSSSKLALNMQQIFGLTAMPQVRRSATS 773

Asator-PG      737 TNLAPSSASQ-----RINSSTIGSAGVSHNTARSVAGDHSVTQFALIDENVSALQ 791
Submitted_Seq  774 TNLAPSSSGGNTPIHDLNIGSA-GGGGAGSNTARSVAGDHSVTQFALIDENVSALQ 832

Asator-PG      792 QVINGGALTILASQWKSQFDSEDTIDNEWGHEAVYRMDIANNVCVRETYSEITRL-AR 849
Submitted_Seq  833 QVINGGALTILASQWKSQFDSEDTIDNEWGHEAVYRMDIANNVCVRETYSDITPLDKAK 892

Asator-PG      850 PSTSSVLRMLPSPFKQDSALQASTHSLDSCHRONSLPVSVDNIFDLQMKALD 907
Submitted_Seq  893 PATESLVRVVLSPFKELATFQLNLSINISQAKLGRASLPVSVADLFDQPIHNSISDAI 952

Asator-PG      908 LGSAIQANCCI SGALLIRVIPKTSHPDQSVYYDAMGAVENTITANEGHNRDQKV 964
Submitted_Seq  953 NVALSSAVQENGCISGALLIRVIPKTSHELDQSVYYDALAPIKNTITAMPDQGISOKAN 1012

Asator-PG      965 INCDEMEATSAVIAFFNKSISKINSPPGRDATEERTGASLCSLSYAGVHKLKNG----- 1019
Submitted_Seq  1013 IFCDEIEENAAVIALPSNTANKKQTYTDKIEACIANPCSIHATGVNKLKNGSETC 1072

Asator-PG      1020 ---NTAP-----RTQFKMGSTIDGFGENSESDFLLNPSKIFVRQSKCASNAGADF-I 1068
Submitted_Seq  1073 HALNDQPHYKSGDYKTFPTGGSTDSYRETDGCDLFLNPSKIFVRQSKCASNAGADITV 1132

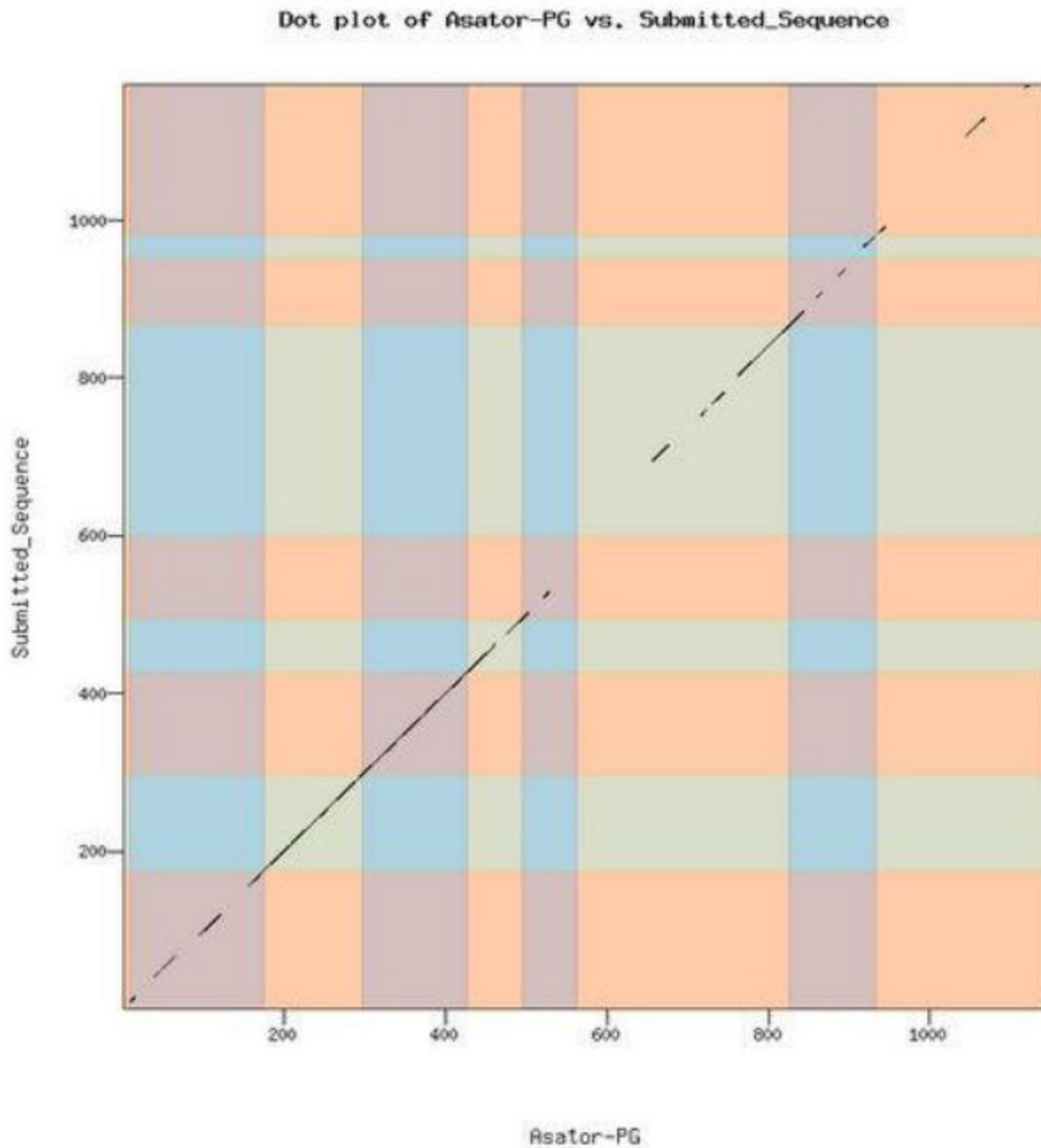
Asator-PG      1069 SASKFLSAAVPCIPYHPQSDITYVIDSIFVAKTITSIALECPNHSIDLTPLGLSYFYC 1128
Submitted_Seq  1133 YSSKTLPRDALPEIPNPQTNT-----YSTALEYPPNITDLTPG----- 1172

Asator-PG      1129 NIVVPLRLFSILLTES 1144
Submitted_Seq  1173 ----- 1172
```

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.



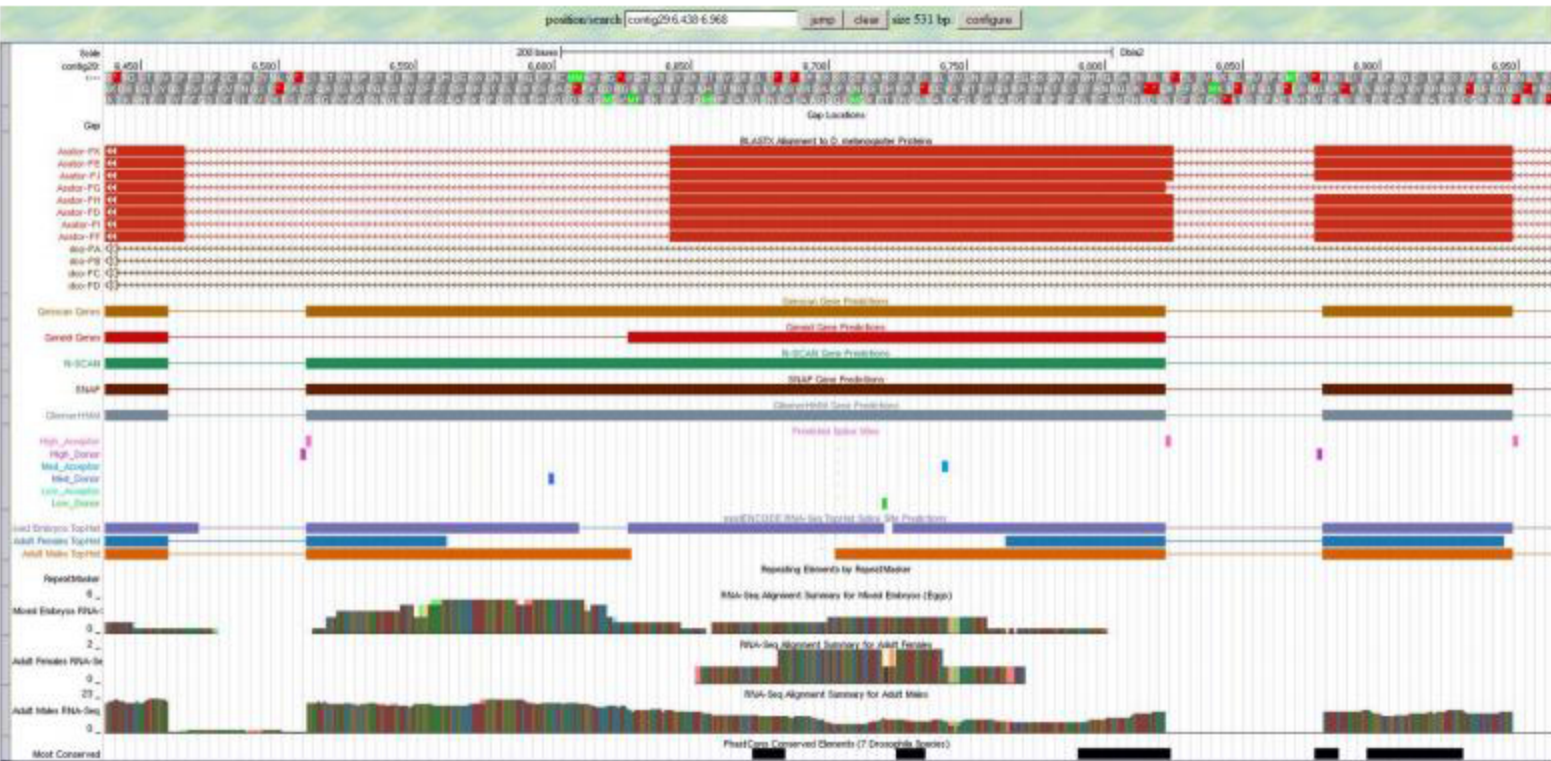
The conservation at the 5' end of the exon (6822-6511) is very poor. Four different gene predictors predict the exon's actual length to be much 45 amino acids longer than the ortholog in *D. melanogaster*. An open reading frame the entire way through is present as well as computational evidence of tophat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

```

D.Wil  FVDSVLQNVNSGINQNP LPKT-----LAQ-----QQLAAANPNLPSALI
D.Moja SADCVVQNVNLGNPNLLKHQXXXXXXXXXXXXXXXXXXXXQQLVIVSNTSY-----
D.Viri  TADCAVQNVNSGI--QNT-----TLPKQHQHQHQHQV L VANTSI AIPNLPSALI
D.Grim  TAESALQNSGTLNALPKQ-----QHQHQHQHQHQV L VNTAIAVANLPSALI
D.Anan  LVEAASQYG--INNQNNAKELLQ-----QKSCECL-----L
D.pseud KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Pers  KCETSLQHANSANNQNTLPKALLQ-----QHPPQSANSNVAMATPNLPNALI
D.Erec  SGDAIVPHGNTANNQNITTKRLQ-----STLT--TNSQV--AIPNIQV---
D.Yak   SGEATIQHNTANNQNIKTGKLQ-----STMT--TNSQV--AIPNIQSAPI
D.mel   SGEPMVQHNAANNQNITSKGLQ-----STL--TNSQV--AIANIQSAP-
D.Sim   SGEALIQHNTANNQNITSKGLQ-----STL--TNSQV--AIPNIQSAPI
D.      SGEALTQHNTANNQNITSKGLQ-----STL--TNSQV--AIPNIQSAPI
:
D.Wil  KSNHENPAAI-----VGQIKSLHGHTAIHTKKCPHVGTGSLISISNQNNSAPVYQYPHA
D.Moja
D.Viri  KSST-VGI VNEHVS GSTHNAQLKST----HGKRCQPQT VSGSYTASNQNNSAPVY-YPES
D.Grim  KPSTVIGIANHEVSTTNSNAPLRSTHGVIHGKRCPAQTVSGSYTTTNQNNSAPVYQY P ES
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---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Pers -----MHLQDNPRNDENVAAS---RDHDSHQQHSNQELFISPSNVSSSSKNNI
D.Biar MLLRFVRVSLFWDLFCEN--DENASSPGDGNHNHTCQPPCNQEYISLNRD---CQKNLF
D.Erec -----
D.Yak -----
D.mel -----MFWHLLCVF--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 MQFLLFFRNNDENASSPGDGNHNHTCQPPCNQEYISLNRDCQKNLFRHppppskpppLA 180
M + L NDENAS+P DGN +CQP Q+QY+S NR+CQKNL RL+PPPPSKPPPL
Sbjct 1 MFWHLLCVNDENASAPDDGNQ--SCQPSSKQDQYLSPNRNCQKNLLRLYPpppskpppLV 58

Query 181 GAILQSRLKQISSGanaenaealee---HRYPNALQRSATLPAKHNLGVRSRVTFKVP 351
GAILQ+RLL QIS A A+ L YPN LQRSATLPAKHNLGVRSRVTFKVP
Sbjct 59 GAILQTRLLHQISPSAIADADADLNAVGELLYPNVLQRSATLPAKHNLGVRSRVTFKVP 118

Query 352 SSITPAVDPGS-EPDPGQNQVVAERDGIVLNPKERESCKMTSEDLLQPGHVVKERWK 522
SS PA D S +P +NQV A +DGI+++ K +ES KMTSEDLLQPGHVVKERWK
Sbjct 119 SSNLPAQDSYSHQP---RNQVAVAAKDGIIVDVKAKESVKMTSEDLLQPGHVVKERWK 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.

