F Element Project: Annotation Report

Faculty instructor(s): Anya Goodman
College/university: California Polytechnic State University – San Luis Obispo
Course number: CHEM 441
Course name: Bioinformatics Applications

Authorship Information for GEP Scientific Publications

Initials:	By entering my/our initials, I/we grant permission for the Genomics Education Partnership (GEP) to use the annotation data produced in this report in future scientific publications.
RS/VY	

Note: Please skip the rest of this section if **more than three students** contribute to this annotation report. When more than three students contribute to an annotation project, the class as a whole will be acknowledged in future GEP scientific publications.

Co-authors Responsibilities

In order to be a co-author on a GEP publication, you must review, critique, and approve the final gene models and manuscript, responding promptly to requests to read and approve. As part of the preparations for the microPublication article, co-authors are required to validate specific data within the manuscript, supplemental materials, and GenBank submission (the specific details will depend on each annotation project). In most cases, the manuscript preparation process will take approximately 3–5 hours of your time.

The above requirements mean that we must be able to contact you when the GEP microPublication, and later, the scientific paper with meta-analysis, is ready for your review and approval. **If we cannot reach you at that time, you will not be a co-author on our GEP scientific publications,** as scientific journals require all co-authors to have read and approved the manuscript.

Please provide your contact information below. Note that your name and contact information will be publicly available through the scientific publication and the GenBank record (this is standard for all scientific publications.). Please list the authors in ascending alphabetical order by last name. (The actual order of the student co-authors in the scientific publication will be determined by a random number generator.)

Contact information for Author #1

First name	Vianna
Middle initials	C
Last name	Yang
Author name (name that will appear on the publication):	Vianna Chen Yang
Permanent Email address (one you will use five years from now):	viannachenyang@gmail.com
Alternative Email address (optional):	_____
Enter your initials to indicate that you have read and accept the co-authors responsibilities	VY

Contact information for Author #2

First name	Rachel
Middle initials	L
Last name	Stockwell
Author name (name that will appear on the publication):	Rachel Stockwell
Permanent Email address (one you will use five years from now):	rachelstockwell02@gmail.com
Alternative Email address (optional):	_____
Enter your initials to indicate that you have read and accept the co-authors responsibilities	RS

Project Details

Project name: Contig60
Project species: D. Bipectinata
Date of submission: June 2, 2022
Size of project in base pairs: 49,290
Number of genes in project: 5

Does this report cover all of the genes or is it a partial report? Full 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.

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 within this report. Be sure to create enough Isoform Report Forms within your Gene Report Form for all isoforms. If you cannot find evidence for any protein-coding genes in the project, jump to the " REF _Ref91436230 \h Check for additional features in your project" section.

Gene Report Form

Gene name (*e.g.*, *D. bipectinata eyeless*): *D. bipectinate CG13293*
Gene symbol (*e.g.*, *dbip_ey*): *dbip_ CG13293*

Approximate location in project (from 5' end to 3' end): 4185-6917
Number of isoforms in *D. melanogaster:* 4
Number of isoforms in this project: 4

Complete the following table, including all of the isoforms in this project:

Name(s) of unique isoform(s) based on coding sequence	List of isoforms with identical coding sequences
CG13293 – PC	
CG13293 – PB	CG13293 - PE
CG13293 - PD	

Names of the isoforms with unique coding sequences in *D. melanogaster* that are absent in this species: N/A (Error Report Form removed from doc as there is no suggested change)

Isoform Report Form

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

Gene-isoform symbol: dbip_13293-PB

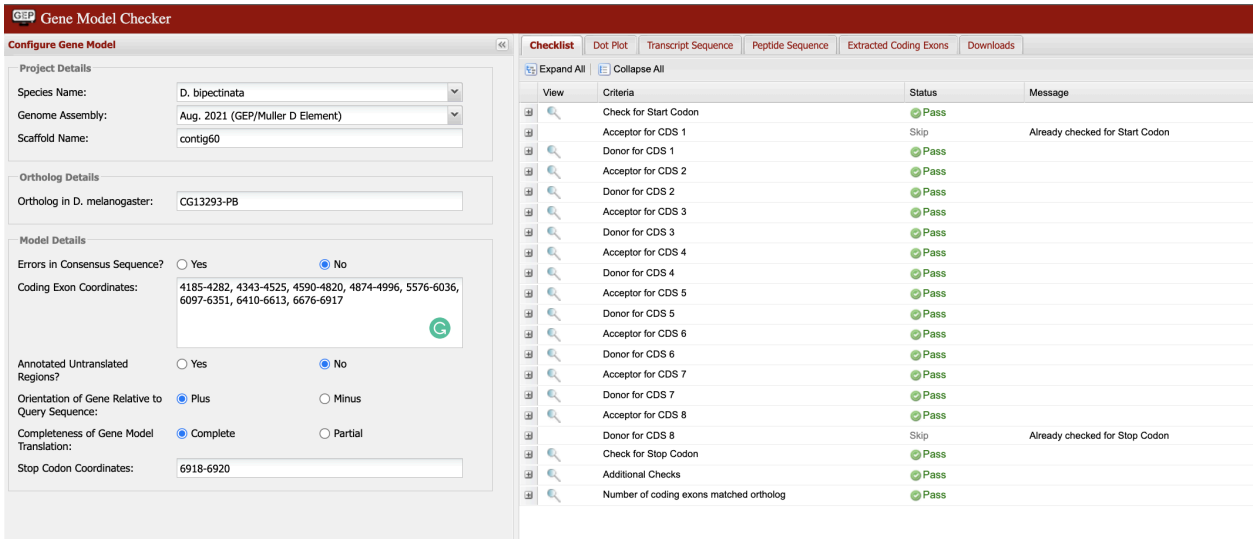
Names of any additional isoforms with identical coding sequences:
dbip_13293-PE

Is the 5' end of this isoform missing from the end of the project? NO
Is the 3' end of this isoform missing from the end of the project? NO

1. *Gene Model Checker* checklist

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

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

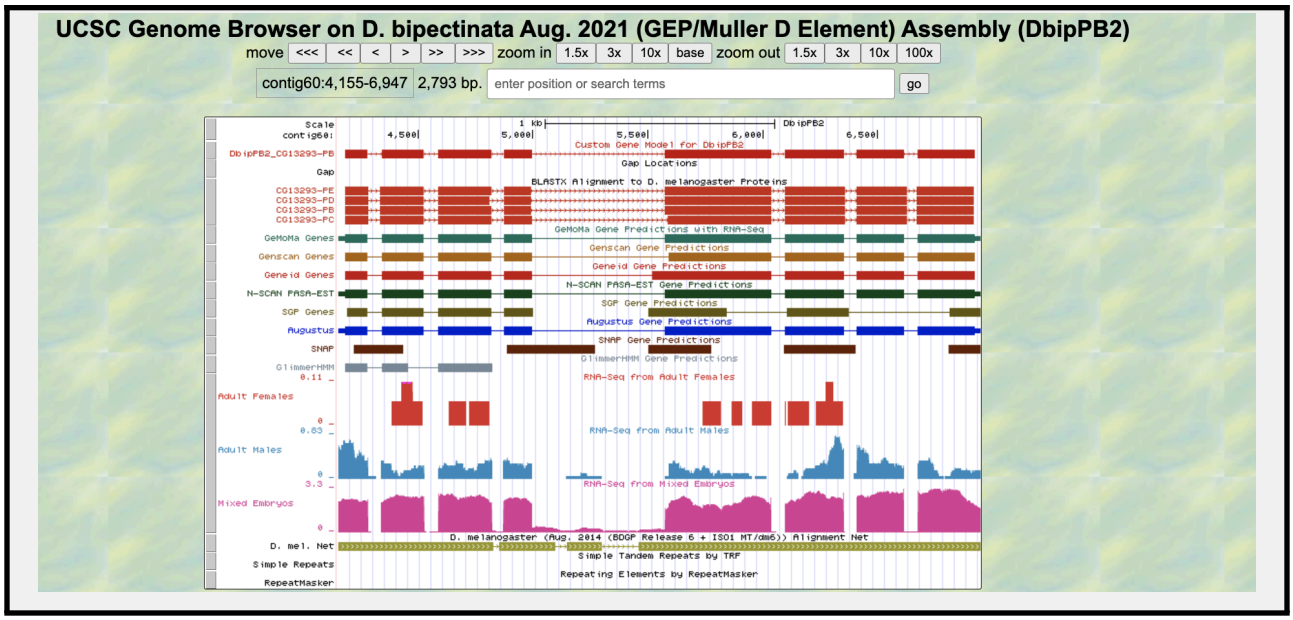


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

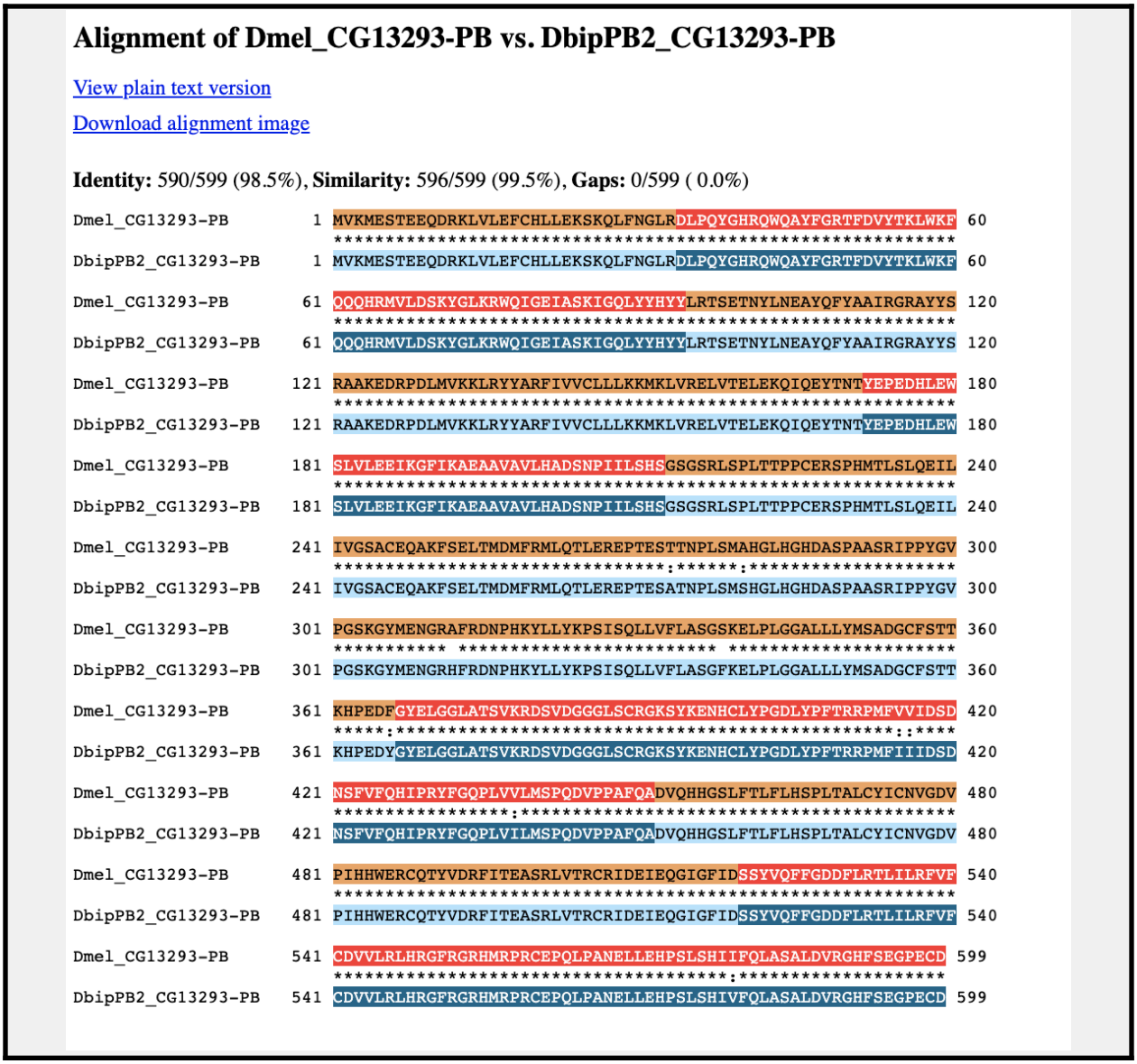
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_CG13293-PB vs. DbipPB2_CG13293-PB

View plain text version

Download alignment image

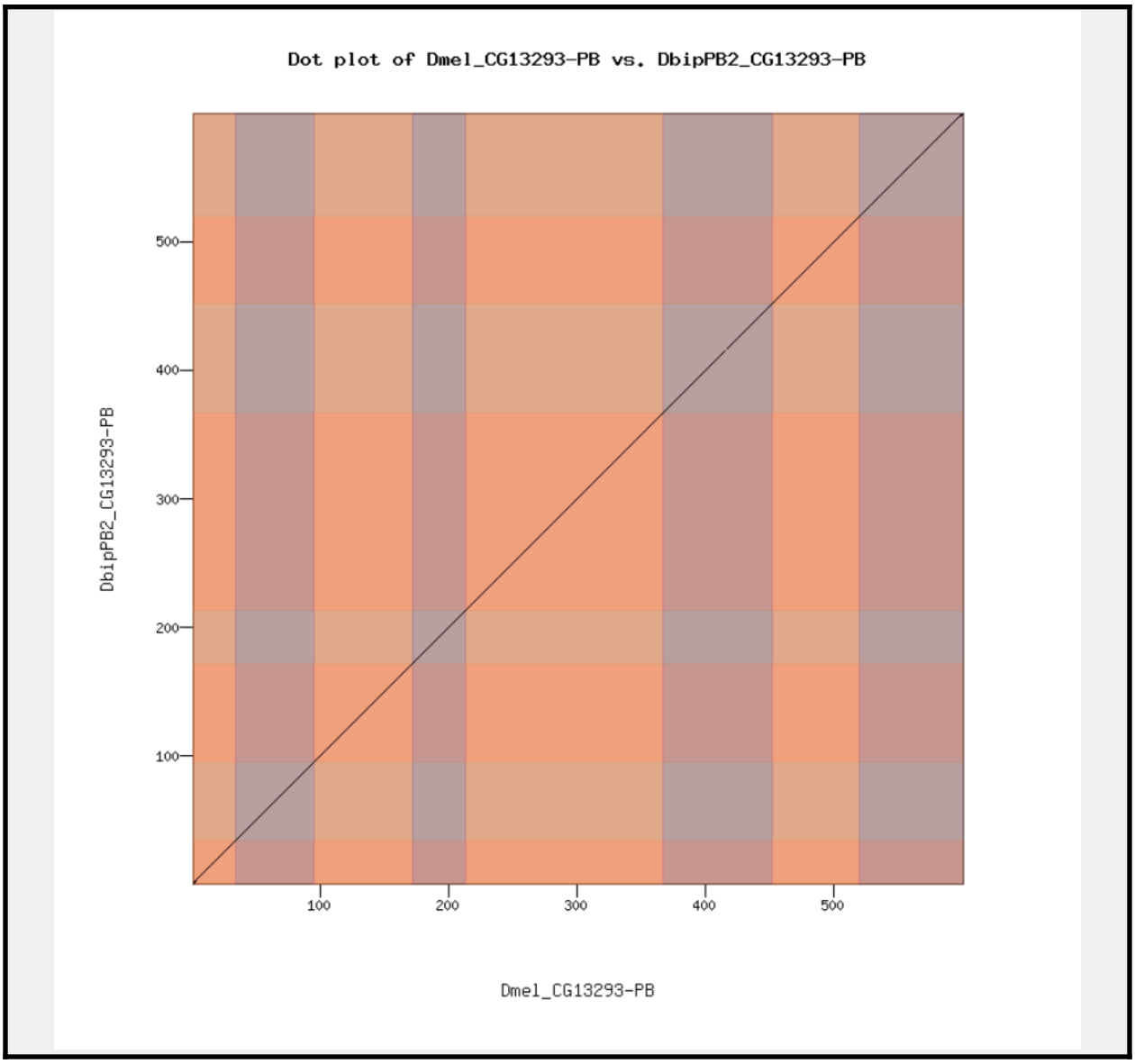
Identity: 590/599 (98.5%), **Similarity:** 596/599 (99.5%), **Gaps:** 0/599 (0.0%)

```
Dmel_CG13293-PB         1 MVKMESTEEQDRKLVLEFCHLLEKSKQLFNGLRDLPQYGHRQWQAYFGRTFDVYTKLWKF 60
                          ************************************************************
DbipPB2_CG13293-PB      1 MVKMESTEEQDRKLVLEFCHLLEKSKQLFNGLRDLPQYGHRQWQAYFGRTFDVYTKLWKF 60

Dmel_CG13293-PB        61 QQQHRMVLDSKYGLKRWQIGEIASKIGQLYYHYYLRTSETNYLNEAYQFYAAIRGRAYYS 120
                          ************************************************************
DbipPB2_CG13293-PB     61 QQQHRMVLDSKYGLKRWQIGEIASKIGQLYYHYYLRTSETNYLNEAYQFYAAIRGRAYYS 120

Dmel_CG13293-PB       121 RAAKEDRPDLMVKKLRYYARFIVVCLLLKKMKLVRELVTELEKQIQEYTNTYEPEDHLEW 180
                          ************************************************************
DbipPB2_CG13293-PB    121 RAAKEDRPDLMVKKLRYYARFIVVCLLLKKMKLVRELVTELEKQIQEYTNTYEPEDHLEW 180

Dmel_CG13293-PB       181 SLVLEEIKGFIKAEAAVAVLHADSNPIILSHSGSGSRLSPLTTPPCERSPHMTLSLQEIL 240
                          ************************************************************
DbipPB2_CG13293-PB    181 SLVLEEIKGFIKAEAAVAVLHADSNPIILSHSGSGSRLSPLTTPPCERSPHMTLSLQEIL 240

Dmel_CG13293-PB       241 IVGSACEQAKFSELTMDMFRMLQTLEREPTESTTNPLSMAHGLHGHDASPAASRIPPYGV 300
                          ********************************:******:********************
DbipPB2_CG13293-PB    241 IVGSACEQAKFSELTMDMFRMLQTLEREPTESATNPLSMSHGLHGHDASPAASRIPPYGV 300

Dmel_CG13293-PB       301 PGSKGYMENGRAFRDNPHKYLLYKPSISQLLVFLASGSKELPLGGALLLYMSADGCFSTT 360
                          ********** ********************** ********************
DbipPB2_CG13293-PB    301 PGSKGYMENGRHFRDNPHKYLLYKPSISQLLVFLASGFKELPLGGALLLYMSADGCFSTT 360

Dmel_CG13293-PB       361 KHPEDFGYELGGLATSVKRDSVDGGGLSCRGKSYKENHCLYPGDLYPFTRRPMFVVIDSD 420
                          *****:**************************************************::****
DbipPB2_CG13293-PB    361 KHPEDYGYELGGLATSVKRDSVDGGGLSCRGKSYKENHCLYPGDLYPFTRRPMFIIIDSD 420

Dmel_CG13293-PB       421 NSFVFQHIPRYFGQPLVVLMSPQDVPPAFQADVQHHGSLFTLFLHSPLTALCYICNVGDV 480
                          ****************:*****************************************
DbipPB2_CG13293-PB    421 NSFVFQHIPRYFGQPLVILMSPQDVPPAFQADVQHHGSLFTLFLHSPLTALCYICNVGDV 480

Dmel_CG13293-PB       481 PIHHWERCQTYVDRFITEASRLVTRCRIDEIEQGIGFIDSSYVQFFGDDFLRTLILRFVF 540
                          ************************************************************
DbipPB2_CG13293-PB    481 PIHHWERCQTYVDRFITEASRLVTRCRIDEIEQGIGFIDSSYVQFFGDDFLRTLILRFVF 540

Dmel_CG13293-PB       541 CDVVLRLHRGFRGRHMRPRCEPQLPANELLEHPSLSHIIFQLASALDVRGHFSEGPECD 599
                          ************************************:***********************
DbipPB2_CG13293-PB    541 CDVVLRLHRGFRGRHMRPRCEPQLPANELLEHPSLSHIVFQLASALDVRGHFSEGPECD 599
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol: dbip_13293-PC

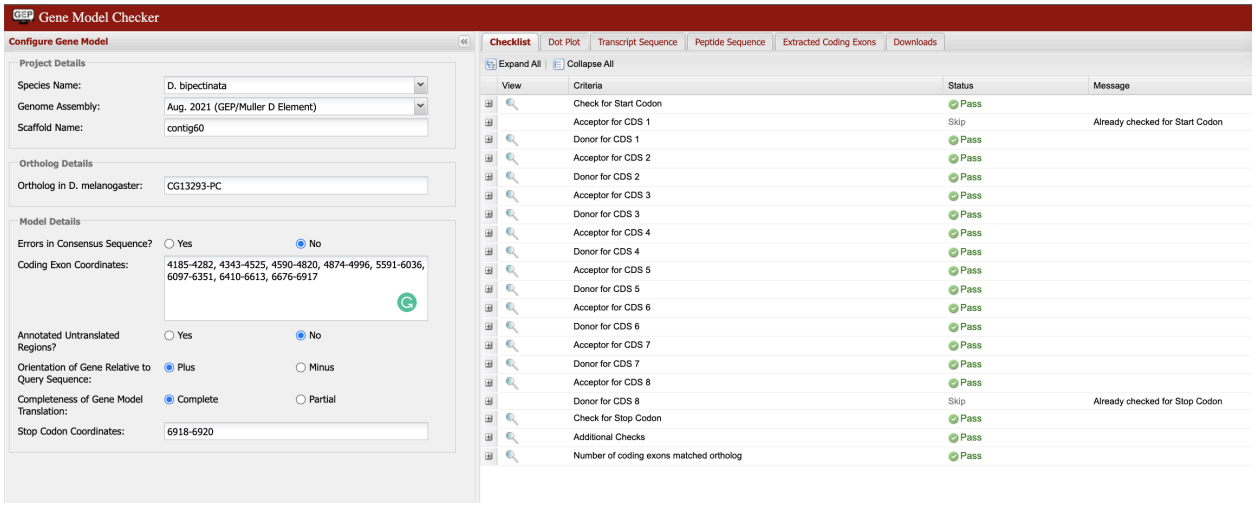
Names of any additional isoforms with identical coding sequences:

Is the 5' end of this isoform missing from the end of the project? NO
Is the 3' end of this isoform missing from the end of the project? NO

1. *Gene Model Checker* checklist

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

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.



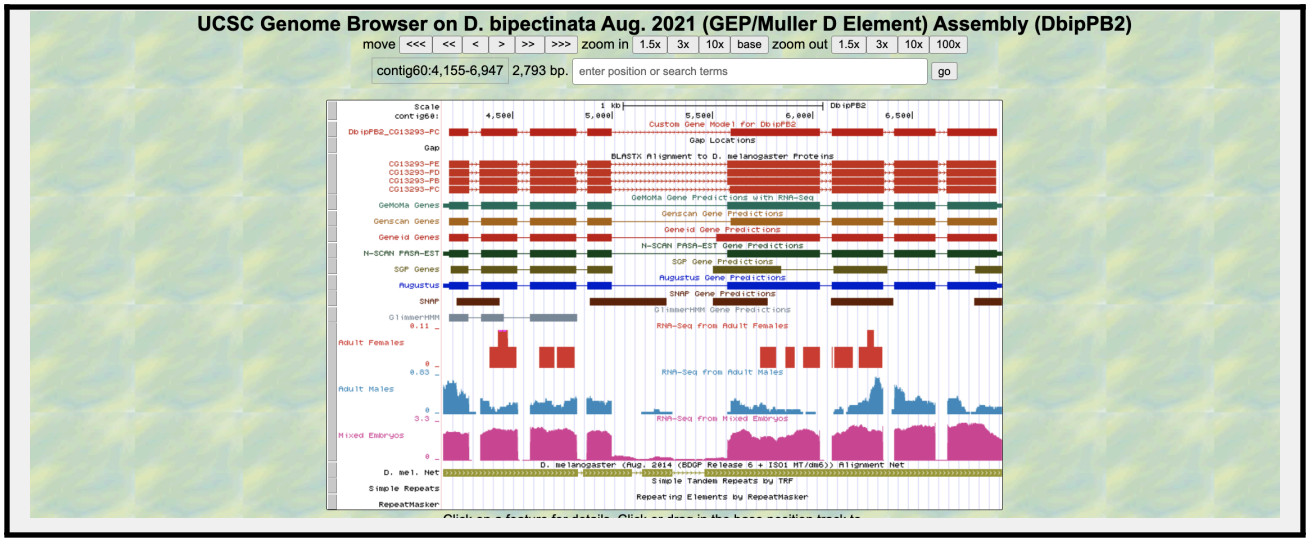
2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in**

the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:

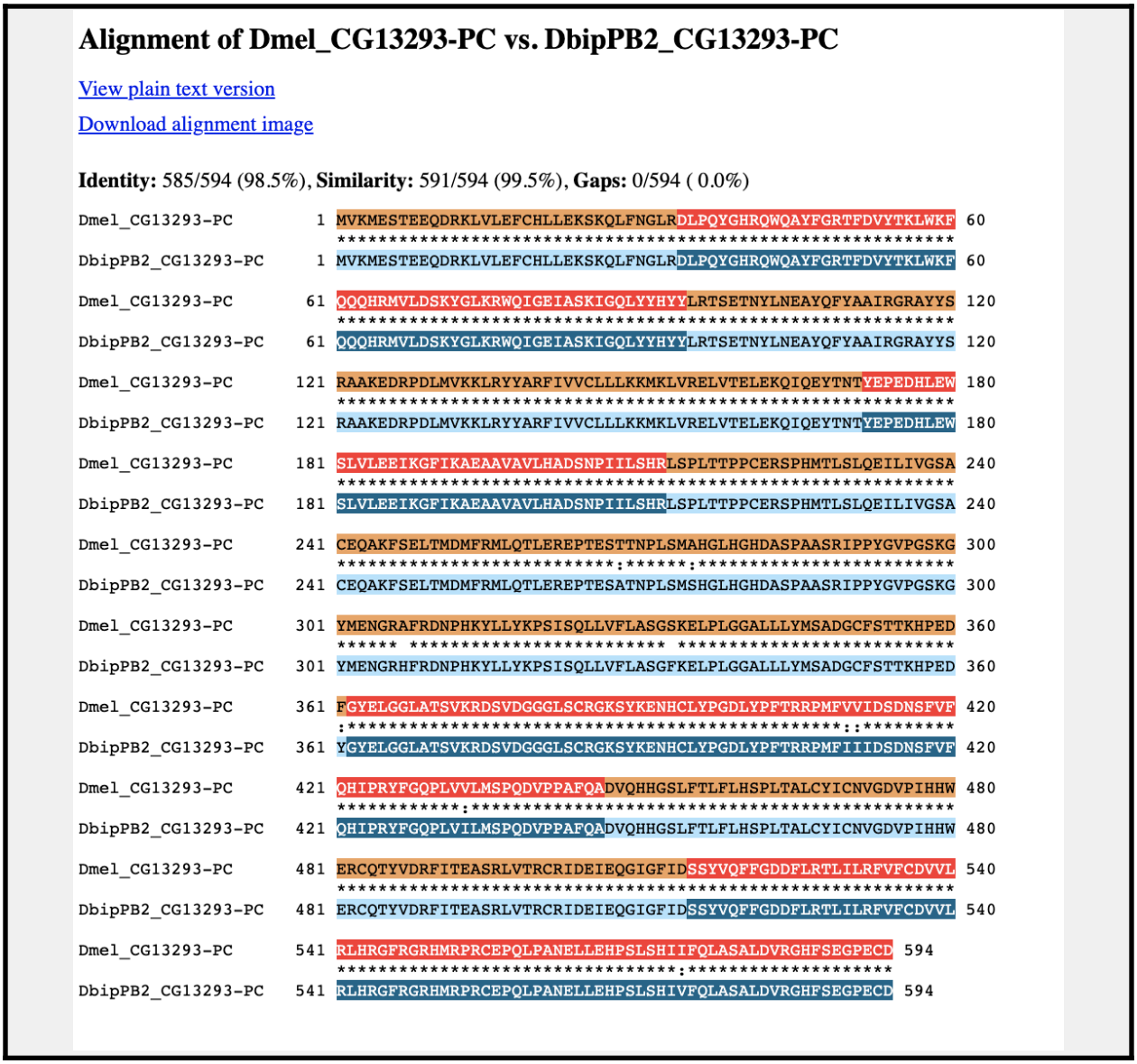
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_CG13293-PC vs. DbipPB2_CG13293-PC

View plain text version

Download alignment image

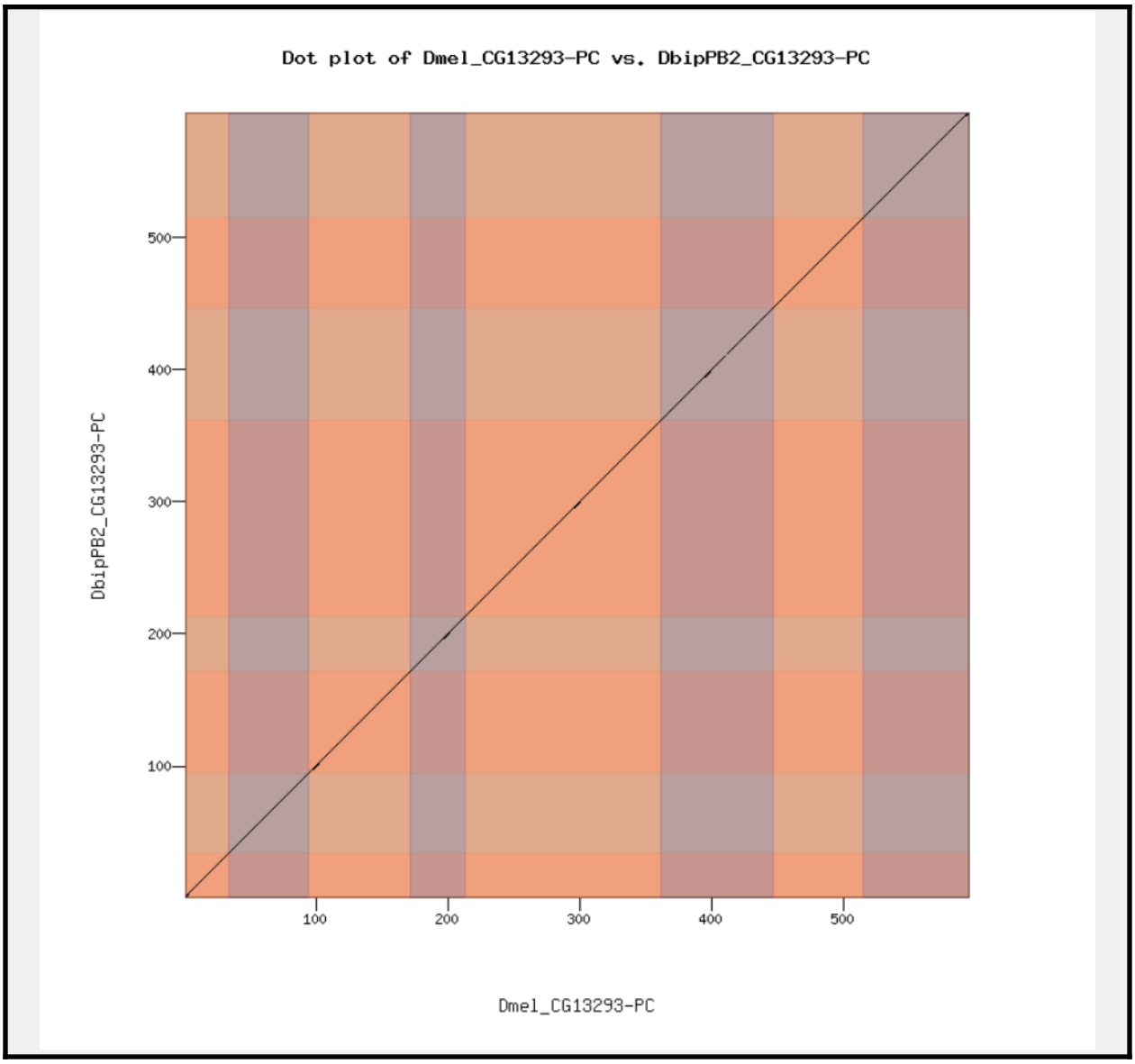
Identity: 585/594 (98.5%), **Similarity:** 591/594 (99.5%), **Gaps:** 0/594 (0.0%)

```
Dmel_CG13293-PC         1 MVKMESTEEQDRKLVLEFCHLLEKSKQLFNGLRDLPQYGHRQWQAYFGRTFDVYTKLWKF 60
                          ************************************************************
DbipPB2_CG13293-PC      1 MVKMESTEEQDRKLVLEFCHLLEKSKQLFNGLRDLPQYGHRQWQAYFGRTFDVYTKLWKF 60

Dmel_CG13293-PC        61 QQQHRMVLDSKYGLKRWQIGEIASKIGQLYYHYYLRTSETNYLNEAYQFYAAIRGRAYYS 120
                          ************************************************************
DbipPB2_CG13293-PC     61 QQQHRMVLDSKYGLKRWQIGEIASKIGQLYYHYYLRTSETNYLNEAYQFYAAIRGRAYYS 120

Dmel_CG13293-PC       121 RAAKEDRPDLMVKKLRYYARFIVVCLLLKKMKLVRELVTELEKQIQEYTNTYEPEDHLEW 180
                          ************************************************************
DbipPB2_CG13293-PC    121 RAAKEDRPDLMVKKLRYYARFIVVCLLLKKMKLVRELVTELEKQIQEYTNTYEPEDHLEW 180

Dmel_CG13293-PC       181 SLVLEEIKGFIKAEAAVAVLHADSNPIILSHRLSPLTTPPCERSPHMTLSLQEILIVGSA 240
                          ************************************************************
DbipPB2_CG13293-PC    181 SLVLEEIKGFIKAEAAVAVLHADSNPIILSHRLSPLTTPPCERSPHMTLSLQEILIVGSA 240

Dmel_CG13293-PC       241 CEQAKFSELTMDMFRMLQTLEREPTESTTNPLSMAHGLHGHDASPAASRIPPYGVPGSKG 300
                          ****************************:******:************************
DbipPB2_CG13293-PC    241 CEQAKFSELTMDMFRMLQTLEREPTESATNPLSMSHGLHGHDASPAASRIPPYGVPGSKG 300

Dmel_CG13293-PC       301 YMENGRAFRDNPHKYLLYKPSISQLLVFLASGSKELPLGGALLLYMSADGCFSTTKHPED 360
                          ****** ********************** ************************** ******
DbipPB2_CG13293-PC    301 YMENGRHFRDNPHKYLLYKPSISQLLVFLASGFKELPLGGALLLYMSADGCFSTTKHPED 360

Dmel_CG13293-PC       361 FGYELGGLATSVKRDSVDGGGLSCRGKSYKENHCLYPGDLYPFTRRPMFVVIDSDNSFVF 420
                          :*************************************************::**********
DbipPB2_CG13293-PC    361 YGYELGGLATSVKRDSVDGGGLSCRGKSYKENHCLYPGDLYPFTRRPMFIIIDSDNSFVF 420

Dmel_CG13293-PC       421 QHIPRYFGQPLVVLMSPQDVPPAFQADVQHHGSLFTLFLHSPLTALCYICNVGDVPIHHW 480
                          ************:***********************************************
DbipPB2_CG13293-PC    421 QHIPRYFGQPLVILMSPQDVPPAFQADVQHHGSLFTLFLHSPLTALCYICNVGDVPIHHW 480

Dmel_CG13293-PC       481 ERCQTYVDRFITEASRLVTRCRIDEIEQGIGFIDSSYVQFFGDDFLRTLILRFVFCDVVL 540
                          ************************************************************
DbipPB2_CG13293-PC    481 ERCQTYVDRFITEASRLVTRCRIDEIEQGIGFIDSSYVQFFGDDFLRTLILRFVFCDVVL 540

Dmel_CG13293-PC       541 RLHRGFRGRHMRPRCEPQLPANELLEHPSLSHIIFQLASALDVRGHFSEGPECD 594
                          ********************************:*****************
DbipPB2_CG13293-PC    541 RLHRGFRGRHMRPRCEPQLPANELLEHPSLSHIVFQLASALDVRGHFSEGPECD 594
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol: dbip_13293-PD

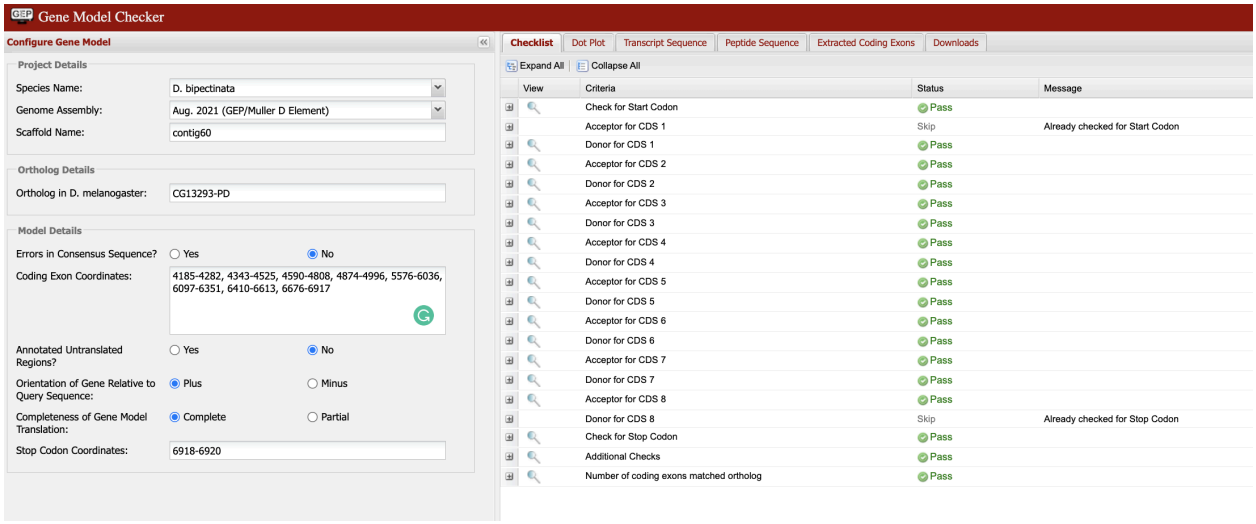
Names of any additional isoforms with identical coding sequences:

Is the 5' end of this isoform missing from the end of the project? NO
Is the 3' end of this isoform missing from the end of the project? NO

1. *Gene Model Checker* checklist

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

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.



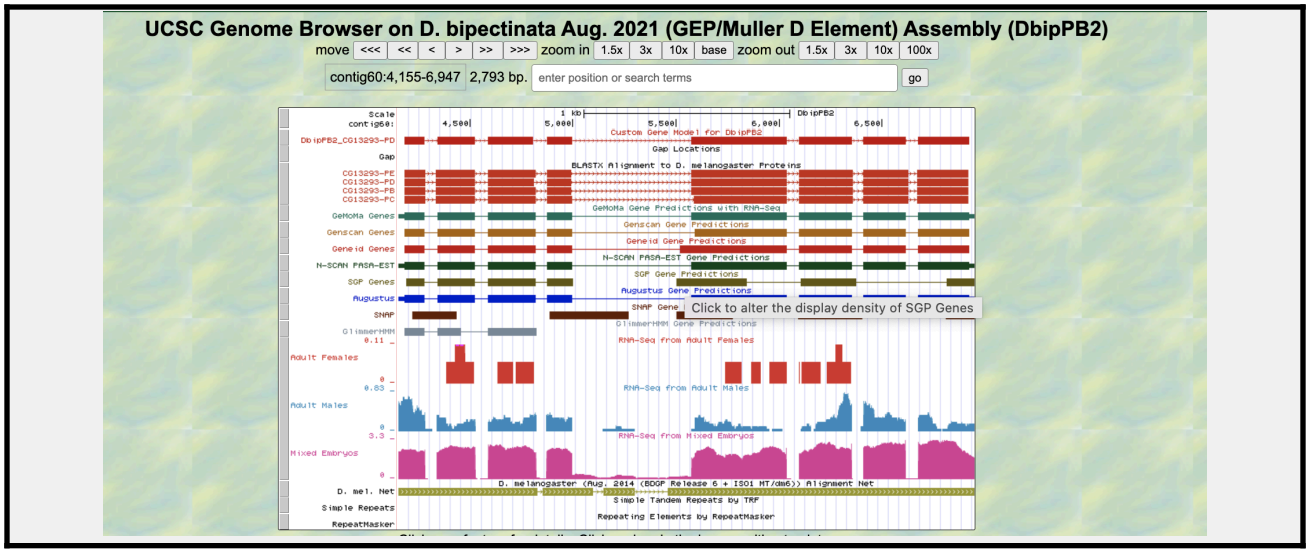
2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in**

the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:

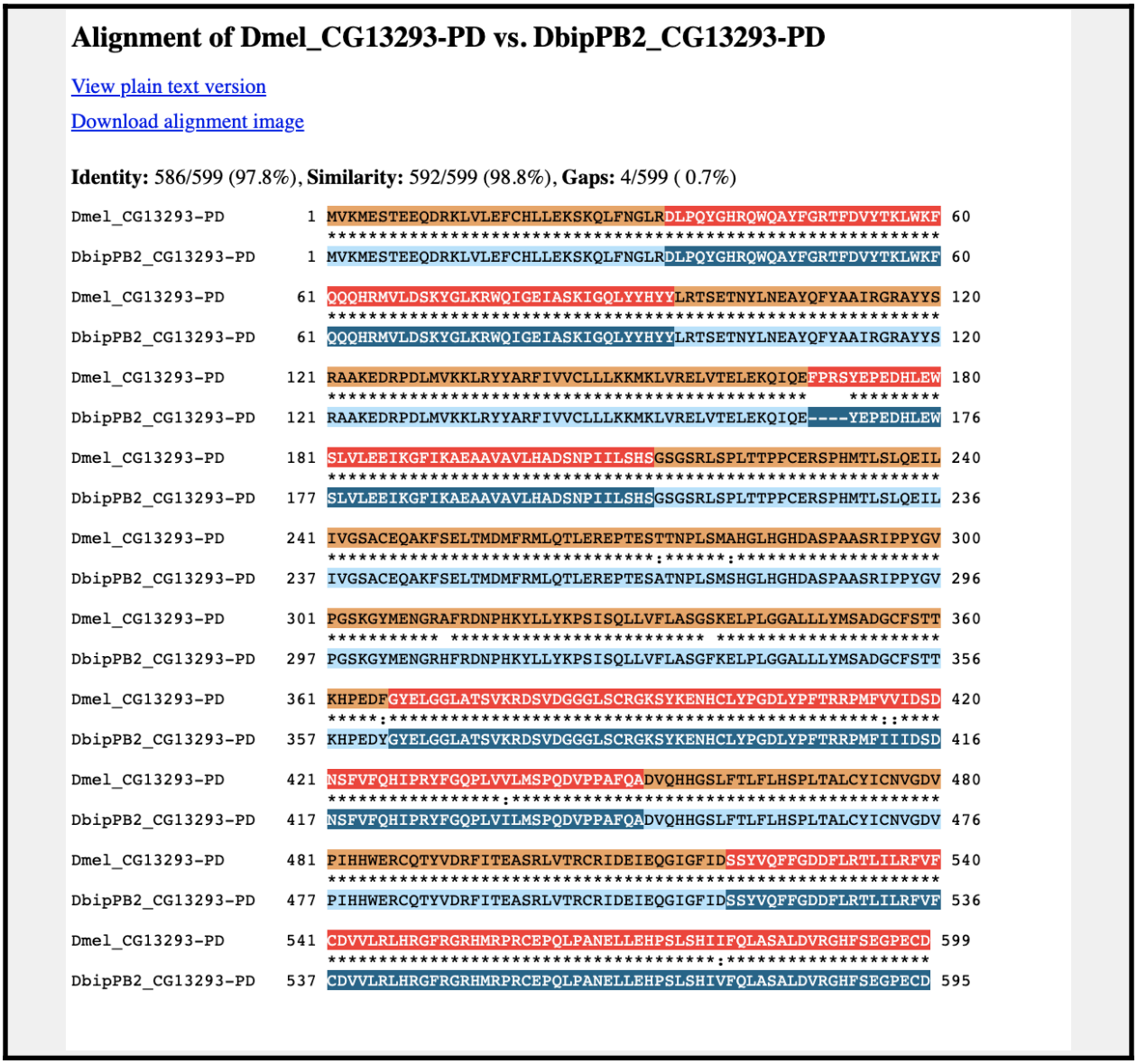
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_CG13293-PD vs. DbipPB2_CG13293-PD

View plain text version

Download alignment image

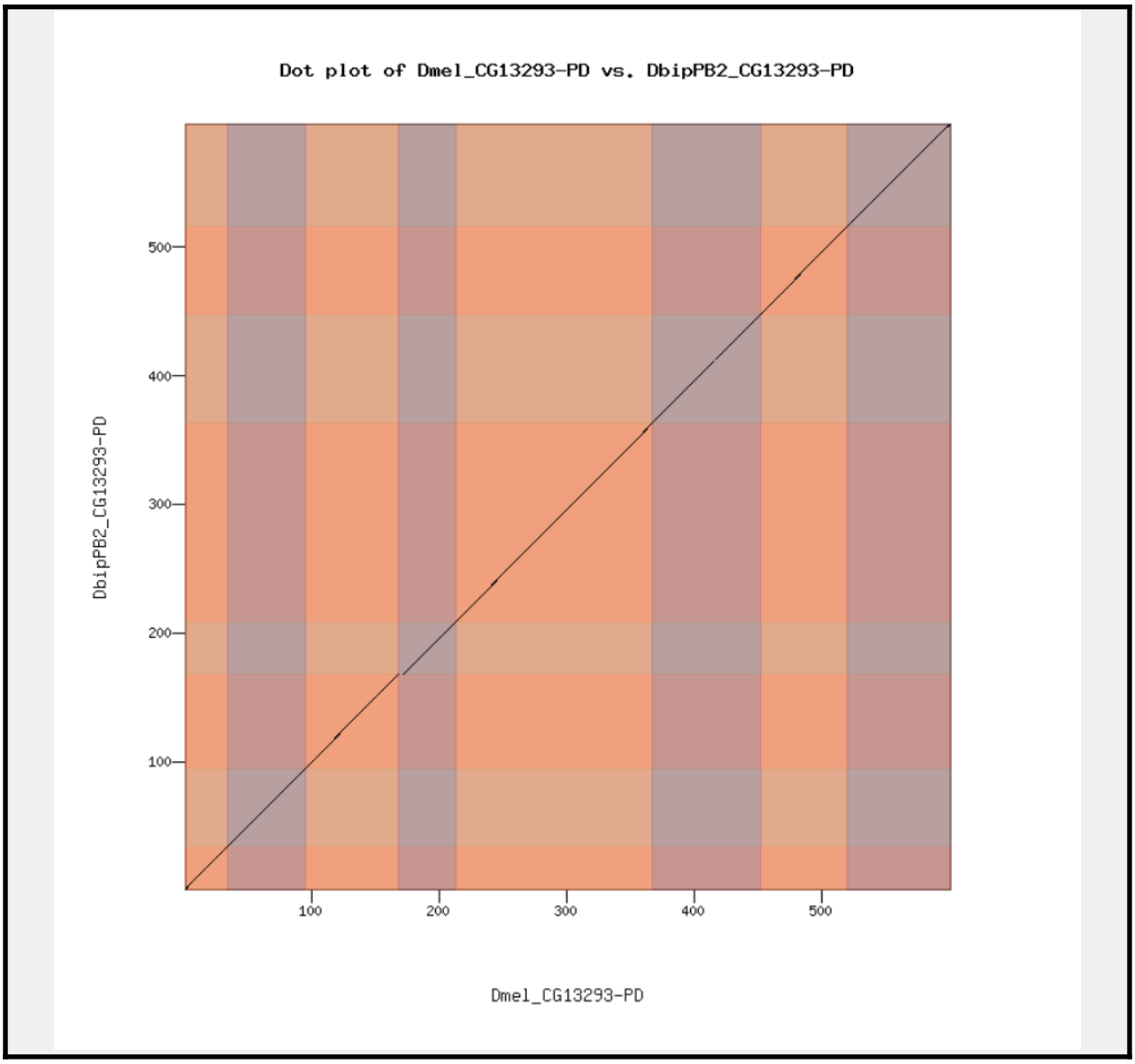
Identity: 586/599 (97.8%), **Similarity:** 592/599 (98.8%), **Gaps:** 4/599 (0.7%)

```
Dmel_CG13293-PD       1 MVKMESTEEQDRKLVLEFCHLLEKSKQLFNGLRDLPQYGHRQWQAYFGRTFDVYTKLWKF 60
                        ************************************************************
DbipPB2_CG13293-PD    1 MVKMESTEEQDRKLVLEFCHLLEKSKQLFNGLRDLPQYGHRQWQAYFGRTFDVYTKLWKF 60

Dmel_CG13293-PD      61 QQQHRMVLDSKYGLKRWQIGEIASKIGQLYYHYYLRTSETNYLNEAYQFYAAIRGRAYYS 120
                        ************************************************************
DbipPB2_CG13293-PD   61 QQQHRMVLDSKYGLKRWQIGEIASKIGQLYYHYYLRTSETNYLNEAYQFYAAIRGRAYYS 120

Dmel_CG13293-PD     121 RAAKEDRPDLMVKKLRYYARFIVVCLLLKKMKLVRELVTELEKQIQEFPRSYEPEDHLEW 180
                        ********************************************     **********
DbipPB2_CG13293-PD  121 RAAKEDRPDLMVKKLRYYARFIVVCLLLKKMKLVRELVTELEKQIQE----YEPEDHLEW 176

Dmel_CG13293-PD     181 SLVLEEIKGFIKAEAAVAVLHADSNPIILSHSGSGSRLSPLTTPPCERSPHMTLSLQEIL 240
                        ************************************************************
DbipPB2_CG13293-PD  177 SLVLEEIKGFIKAEAAVAVLHADSNPIILSHSGSGSRLSPLTTPPCERSPHMTLSLQEIL 236

Dmel_CG13293-PD     241 IVGSACEQAKFSELTMDMFRMLQTLEREPTESTTNPLSMAHGLHGHDASPAASRIPPYGV 300
                        ********************************:******:********************
DbipPB2_CG13293-PD  237 IVGSACEQAKFSELTMDMFRMLQTLEREPTESATNPLSMSHGLHGHDASPAASRIPPYGV 296

Dmel_CG13293-PD     301 PGSKGYMENGRAFRDNPHKYLLYKPSISQLLVFLASGSKELPLGGALLLYMSADGCFSTT 360
                        *********** ************************ **********************
DbipPB2_CG13293-PD  297 PGSKGYMENGRHFRDNPHKYLLYKPSISQLLVFLASGFKELPLGGALLLYMSADGCFSTT 356

Dmel_CG13293-PD     361 KHPEDFGYELGGLATSVKRDSVDGGGLSCRGKSYKENHCLYPGDLYPFTRRPMFVVIDSD 420
                        *****:**************************************************::****
DbipPB2_CG13293-PD  357 KHPEDYGYELGGLATSVKRDSVDGGGLSCRGKSYKENHCLYPGDLYPFTRRPMFIIIDSD 416

Dmel_CG13293-PD     421 NSFVFQHIPRYFGQPLVVLMSPQDVPPAFQADVQHHGSLFTLFLHSPLTALCYICNVGDV 480
                        ****************:*******************************************
DbipPB2_CG13293-PD  417 NSFVFQHIPRYFGQPLVILMSPQDVPPAFQADVQHHGSLFTLFLHSPLTALCYICNVGDV 476

Dmel_CG13293-PD     481 PIHHWERCQTYVDRFITEASRLVTRCRIDEIEQGIGFIDSSYVQFFGDDFLRTLILRFVF 540
                        ************************************************************
DbipPB2_CG13293-PD  477 PIHHWERCQTYVDRFITEASRLVTRCRIDEIEQGIGFIDSSYVQFFGDDFLRTLILRFVF 536

Dmel_CG13293-PD     541 CDVVLRLHRGFRGRHMRPRCEPQLPANELLEHPSLSHIIFQLASALDVRGHFSEGPECD 599
                        ************************************:***********************
DbipPB2_CG13293-PD  537 CDVVLRLHRGFRGRHMRPRCEPQLPANELLEHPSLSHIVFQLASALDVRGHFSEGPECD 595
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Transcription Start Sites (TSS) Report Form (Optional)

Note: Complete this section if you have annotated the TSS for the gene above. This section is **optional** and you do not need to complete this section to submit the project.

Gene name: *D. bipectinata CG13293*
Gene symbol: *dbip CG13293*

Name(s) of isoform(s) with unique TSS	List of isoforms with identical TSS
CG13293 - PB	CG13293 - PE
	CG13293 - PD
	CG13293 - PC

Names of the isoforms with unique TSS in *D. melanogaster* that are absent in this species:
N/A

Isoform TSS Report

Complete an Isoform TSS report (through page PAGEREF _Ref91436230 \h 17) for each unique TSS listed in the table above. Copy and paste this form to create as many copies as needed.

Gene-isoform name : *dbip* CG13293 - PB

Names of the isoforms with the same TSS as this isoform:
dbip CG13293 – PE, *dbip* CG13293 – PD, *dbip* CG13293 - PC

Type of core promoter in *D. melanogaster* (see table below):
(Peaked / Intermediate / Broad / Insufficient Evidence)
Peaked

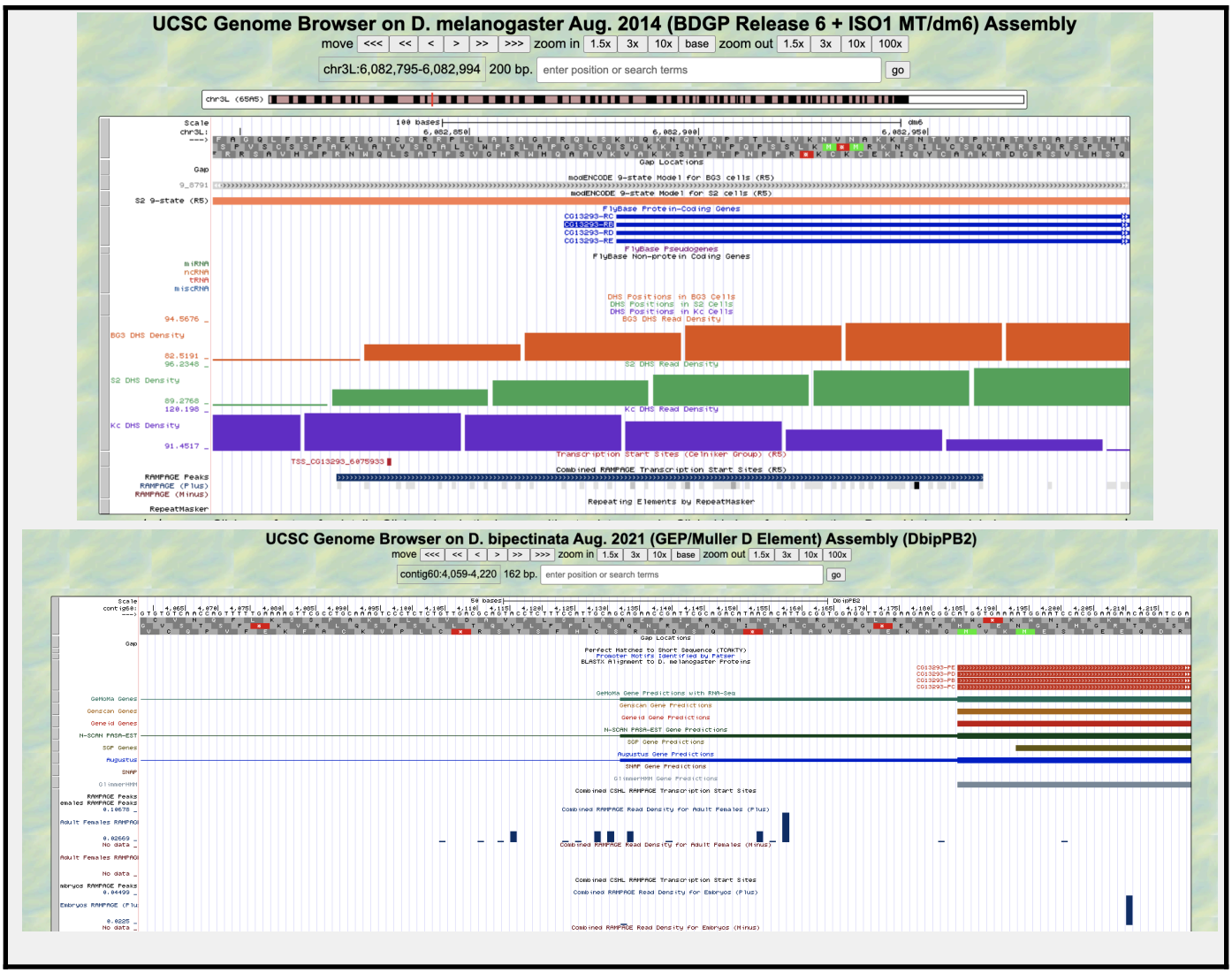
The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

1. Turn on RAMPAGE evidence tracks (Only applies to projects with these tracks)

Coordinates of the TSS position based on position with the highest RAMPAGE read density
Bipectinata: 4105-4182

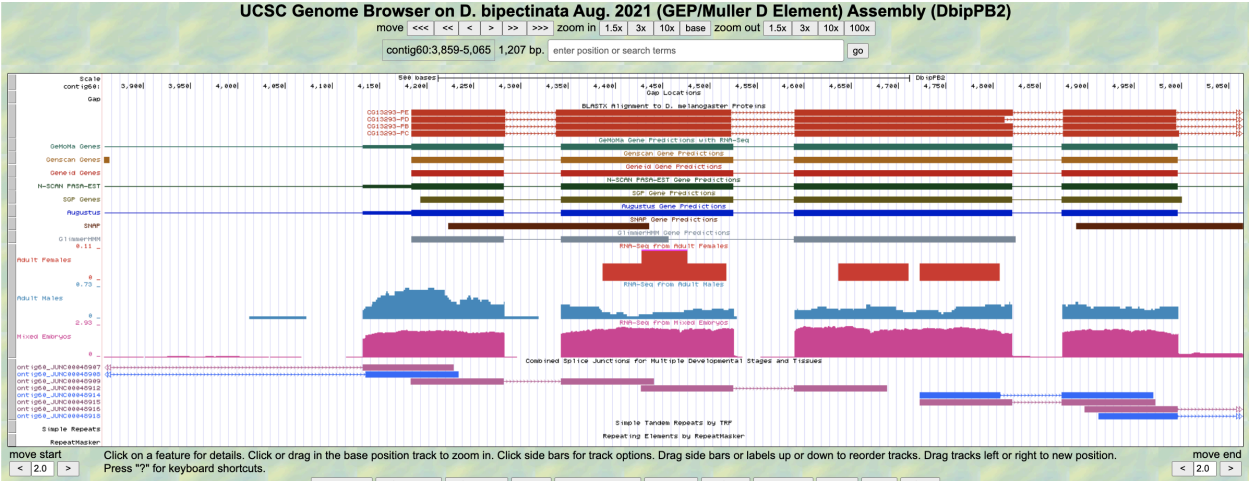
Coordinates of the narrow TSS search region based on RAMPAGE peaks
Bipectinata: 3,767

If the TSS position and narrow TSS search region are supported by RAMPAGE data, **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the Combined RAMPAGE TSS evidence track:**



2. Turn on RNA-Seq evidence tracks (N/A for this Gene)

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (*e.g.*, *regtools*), **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the following evidence tracks:**



1. RNA-Seq Coverage or RNA-Seq Alignment Summary
2. Combined Splice Junctions or RNA-Seq *TopHat*

If the RNA-Seq evidence tracks indicate a TSS search region, list it here: **4133-4184**

3. Annotate the first transcribed exon

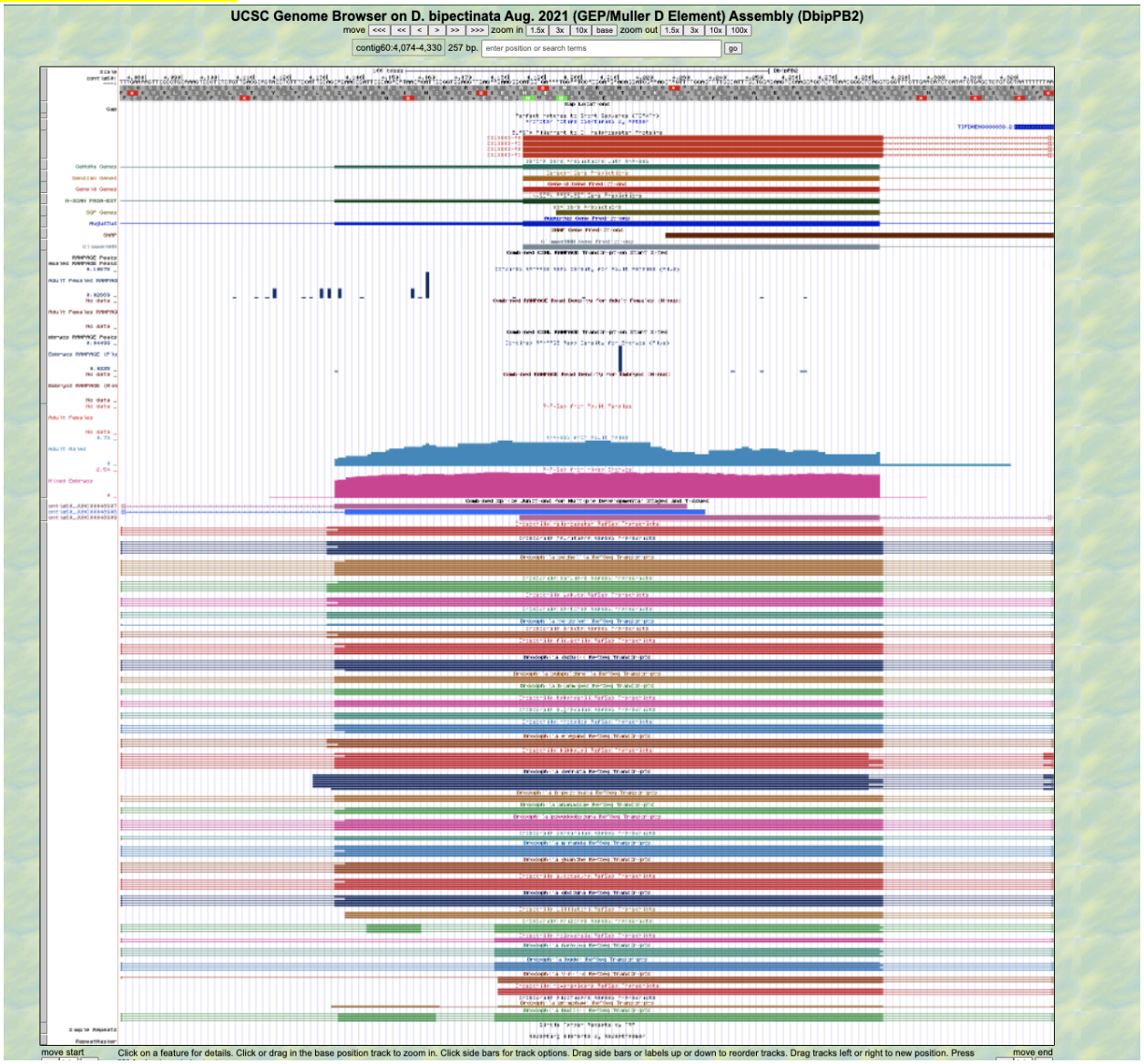
Coordinates of the first transcribed exon based on *BLASTN* alignment:
279-406 (suggests TSS at 279 which makes no sense = refute)
Does the *BLASTN* alignment cover the entire *D. melanogaster* first transcribed exon?
NO

If not, specify the parts of the *D. melanogaster* exon that are missing from the *BLASTN* alignment.
Query length of the entire exon is 471. BLATN covers 1-128

If the TSS annotation is supported by *BLASTN* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *BLASTN* alignment into the box below:**

4. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, **paste a screenshot of the multiple sequence alignment (*e.g.*, from *Clustal Omega*, *ROAST*) into the box below:**



5. Summarize the evidence that supports the TSS annotation postulated above

Coordinate(s) of the TSS position(s):

Based on RAMPAGE data (if applicable): 4105-4182 (also corresponds to the same region in respect to first exon in Melanogaster RAMPAGE)

Based on RNA-Seq data: **4133-4184**

Based on *BLASTN* alignment: 279

Note: If the *BLASTN* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter "Insufficient evidence" if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS search region based on the available evidence? **YES, 4105-4184 with most evidence pointing to 4133 as the TSS**
If so, indicate in the table below the evidence that supports this TSS position

For each evidence type, enter an "X" in the cell to indicate whether the line of evidence supports, refutes, or neither supports nor refutes the TSS annotation:

Evidence type	Support	Refute	Neither
RAMPAGE peaks and read density	X		
RNA-Seq coverage and splice junctions	X		
BLASTN alignment of the initial exon from *D. melanogaster*		X	
Sequence conservation with other *Drosophila* species (*e.g.,* "Conservation" track on the Genome Browser)	X		

Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against the annotated TSS; check "Refute." In contrast, the lack of RNA-Seq read coverage is a negative result that neither supports nor refutes the TSS annotation; check "Neither."

Provide an explanation if the TSS annotation is inconsistent with at least one of the evidence types specified above:

Gene Report Form

Gene name: *D. bipectinate CG10469*
Gene symbol: *dbip_CG10469*

Approximate location in project (from 5' end to 3' end): 9014-9871
Number of isoforms in *D. melanogaster:* 1
Number of isoforms in this project: 1

Complete the following table, including all of the isoforms in this project:

Name(s) of unique isoform(s) based on coding sequence	List of isoforms with identical coding sequences
CGG10469 - PA	

Names of the isoforms with unique coding sequences in *D. melanogaster* that are absent in this species: N/A (report form deleted from this document because of N/A)

Isoform Report Form

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

Gene-isoform symbol: dbip_CGG10469 - PA

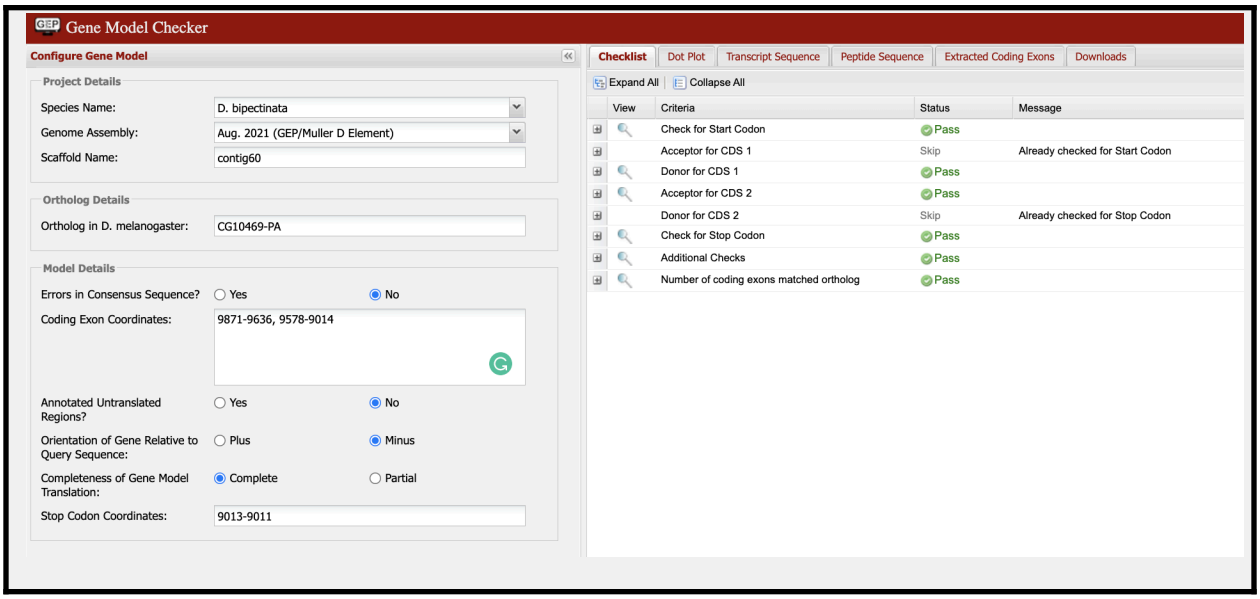
Names of any additional isoforms with identical coding sequences:
N/A

Is the 5' end of this isoform missing from the end of the project? NO
Is the 3' end of this isoform missing from the end of the project? NO

1. *Gene Model Checker* checklist

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

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

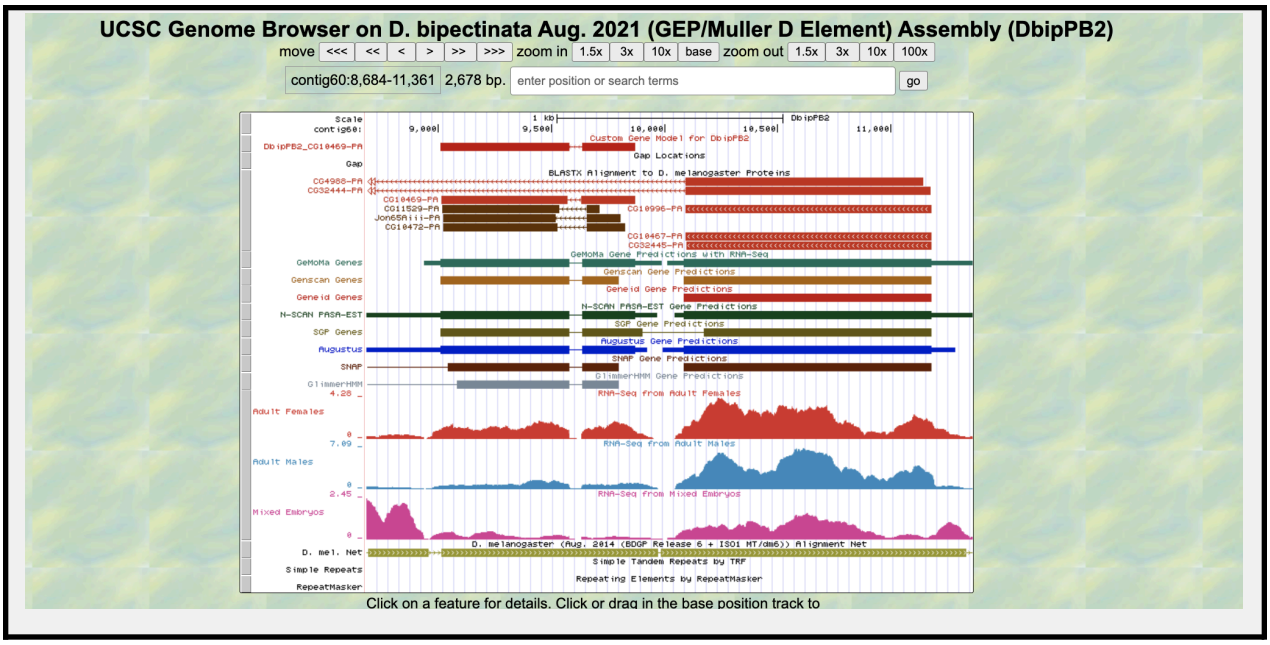


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

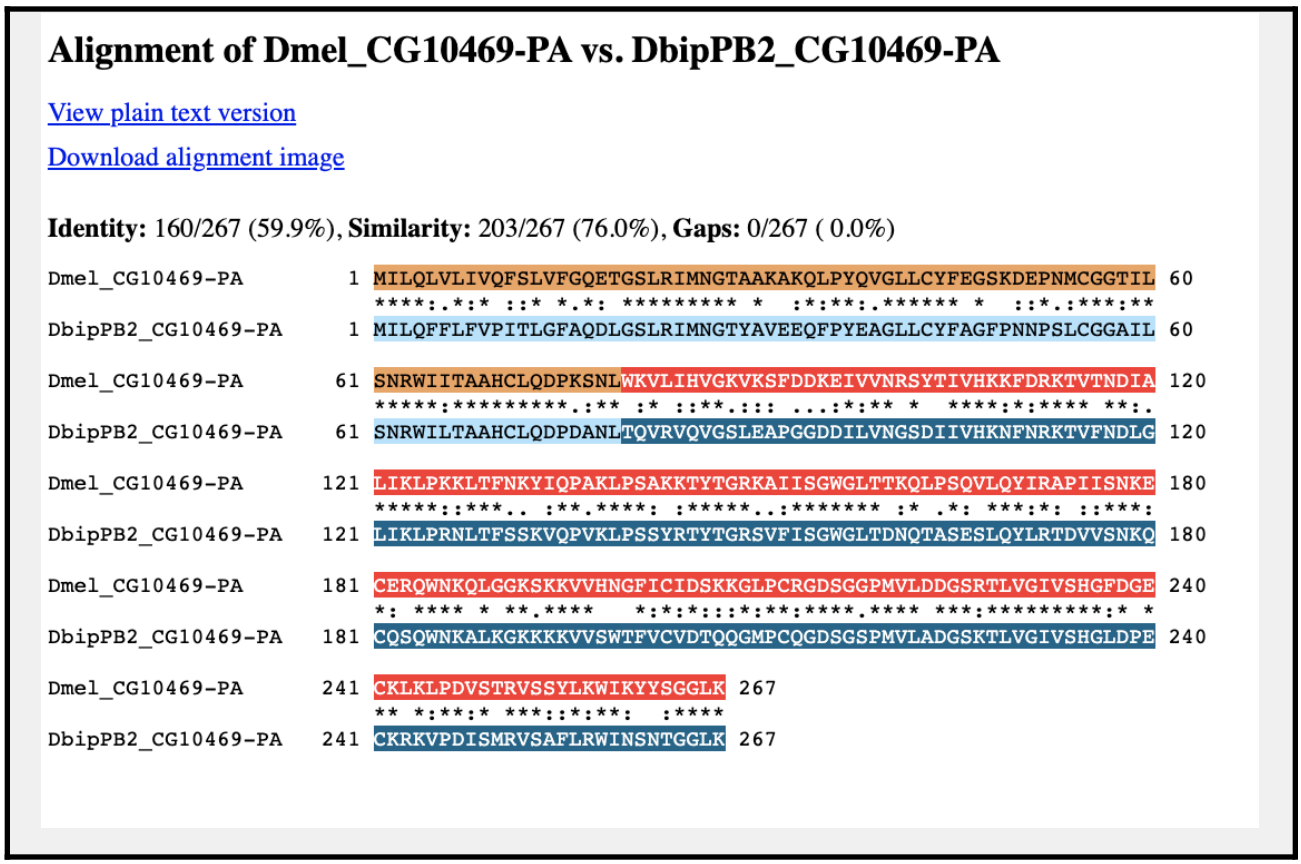
5. A sequence alignment track (*e.g.*, D. mel Proteins)
6. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
7. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
8. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box below:


UCSC Genome Browser on D. bipectinata Aug. 2021 (GEP/Muller D Element) Assembly (DbipPB2)
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
contig60:8,684-11,361 2,678 bp. enter position or search terms go
Scale
contig60:
DbipPB2_CG10469-PA
Gap
CG4988-PA
CG32444-PA
CG10469-PA
CG11529-PA
Jon65Aiii-PA
CG10472-PA
GeMoMa Genes
Genscan Genes
Geneid Genes
N-SCAN PASA-EST
SGP Genes
Augustus
SNAP
GlimmerHMM
Adult Females
Adult Males
Mixed Embryos
D. mel. Net
Simple Repeats
RepeatMasker
Custom Gene Model for DbipPB2
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
CG10996-PA
CG10467-PA
CG32445-PA
GeMoMa Gene Predictions with RNA-Seq
Genscan Gene Predictions
Geneid Gene Predictions
N-SCAN PASA-EST Gene Predictions
SGP Gene Predictions
Augustus Gene Predictions
SNAP Gene Predictions
GlimmerHMM Gene Predictions
RNA-Seq from Adult Females
RNA-Seq from Adult Males
RNA-Seq from Mixed Embryos
D. melanogaster (Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6)) Alignment Net
Simple Tandem Repeats by TRF
Repeating Elements by RepeatMaster
Click on a feature for details. Click or drag in the base position track to

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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_CG10469-PA vs. DbipPB2_CG10469-PA

View plain text version

Download alignment image

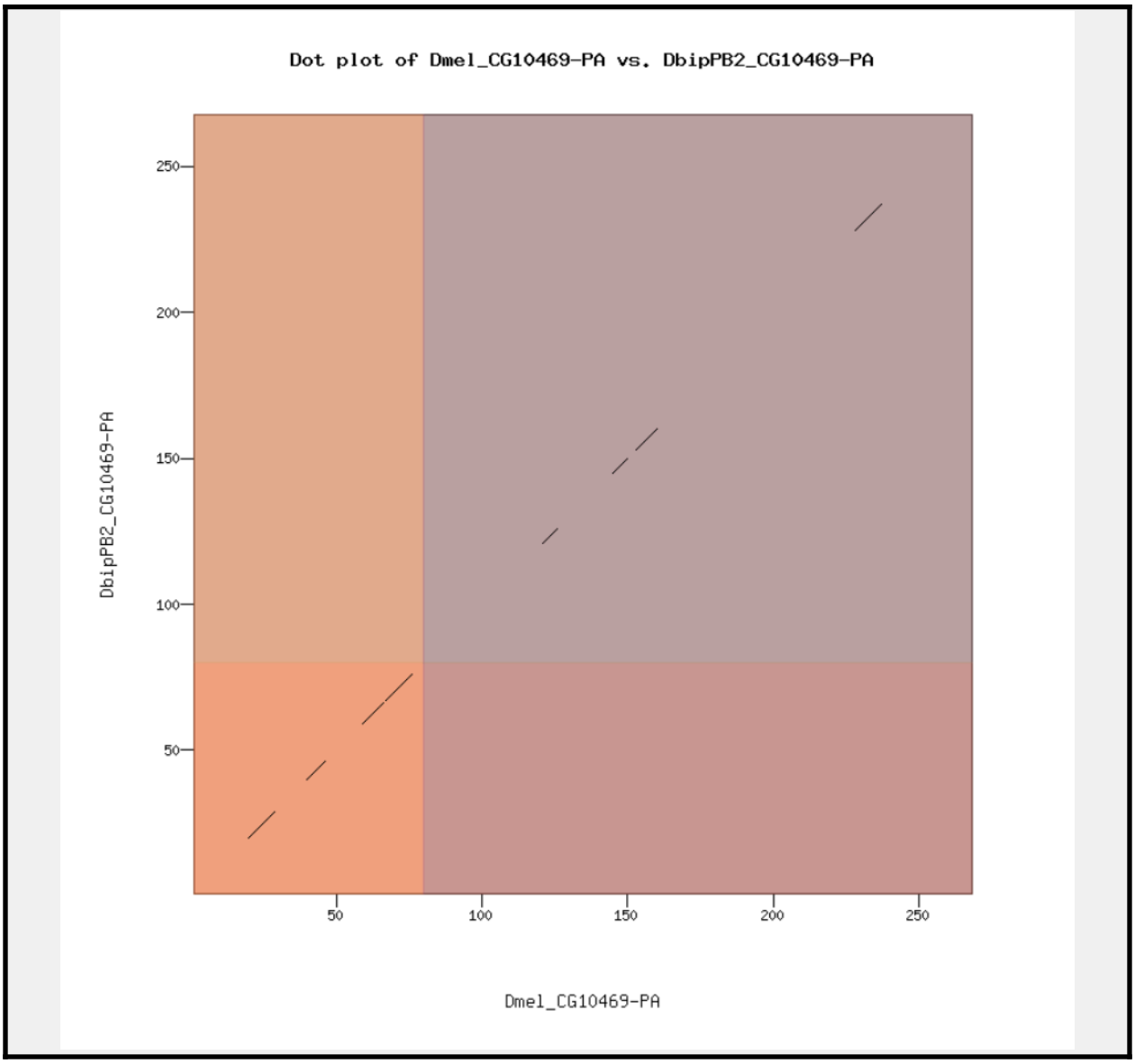
Identity: 160/267 (59.9%), **Similarity:** 203/267 (76.0%), **Gaps:** 0/267 (0.0%)

```
Dmel_CG10469-PA         1 MILQLVLIVQFSLVFGQETGSLRIMNGTAAKAKQLPYQVGLLCYFEGSKDEPNMCGGTIL 60
                          ****:.*:* ::* *.*: ********* *  :*:**:.****** *  ::*.:***:**
DbipPB2_CG10469-PA      1 MILQFFLFVPITLGFAQDLGSLRIMNGTYAVEEQFPYEAGLLCYFAGFPNNPSLCGGAIL 60

Dmel_CG10469-PA        61 SNRWIITAAHCLQDPKSNLWKVLIHVGKVKSFDDKEIVVNRSYTIVHKKFDRKTVTNDIA 120
                          *****:*********.:** :* ::**.::: ...:*:** *  ****:*:**** **:.
DbipPB2_CG10469-PA     61 SNRWILTAAHCLQDPDANLTQVRVQVGSLEAPGGDDILVNGSDIIVHKNFNRKTVFNDLG 120

Dmel_CG10469-PA       121 LIKLPKKLTFNKYIQPAKLPSAKKTYTGRKAIISGWGLTTKQLPSQVLQYIRAPIISNKE 180
                          *****::***.. :**.****: :*****..:******* :* .*: ***:*: ::***:
DbipPB2_CG10469-PA    121 LIKLPRNLTFSSKVQPVKLPSSYRTYTGRSVFISGWGLTDNQTASESLQYLRTDVVSNKQ 180

Dmel_CG10469-PA       181 CERQWNKQLGGKSKKVVHNGFICIDSKKGLPCRGDSGGPMVLDDGSRTLVGIVSHGFDGE 240
                          *: **** * **.****   *:*:*:::*:**:****.**** ***:*********:* *
DbipPB2_CG10469-PA    181 CQSQWNKALKGKKKKVVSWTFVCVDTQQGMPCQGDSGSPMVLADGSKTLVGIVSHGLDPE 240

Dmel_CG10469-PA       241 CKLKLPDVSTRVSSYLKWIKYYSGGLK 267
                          ** *:**:* ***::*:**:  :****
DbipPB2_CG10469-PA    241 CKRKVPDISMRVSAFLRWINSNTGGLK 267
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Transcription Start Sites (TSS) Report Form (Optional)

Note: Complete this section if you have annotated the TSS for the gene above. This section is **optional** and you do not need to complete this section to submit the project.

Gene name: *D. bipectinate CG10469*
Gene symbol: *dbip_CG10469*

Name(s) of isoform(s) with unique TSS	List of isoforms with identical TSS
CG10469-PA	

Names of the isoforms with unique TSS in *D. melanogaster* that are absent in this species:
N/A

Provide the evidence (text and figures) which support the hypothesis that these isoforms are absent in this species (*e.g.*, changes in canonical splice sites, gene structure, etc.):

Isoform TSS Report

Complete an Isoform TSS report (through page PAGEREF _Ref91436230 \h 17) for each unique TSS listed in the table above. Copy and paste this form to create as many copies as needed.

Gene-isoform name : dbip_ CGG10469 – PA

Names of the isoforms with the same TSS as this isoform:
N/A

Type of core promoter in *D. melanogaster* (see table below):
(Peaked / Intermediate / Broad / Insufficient Evidence)
Peaked

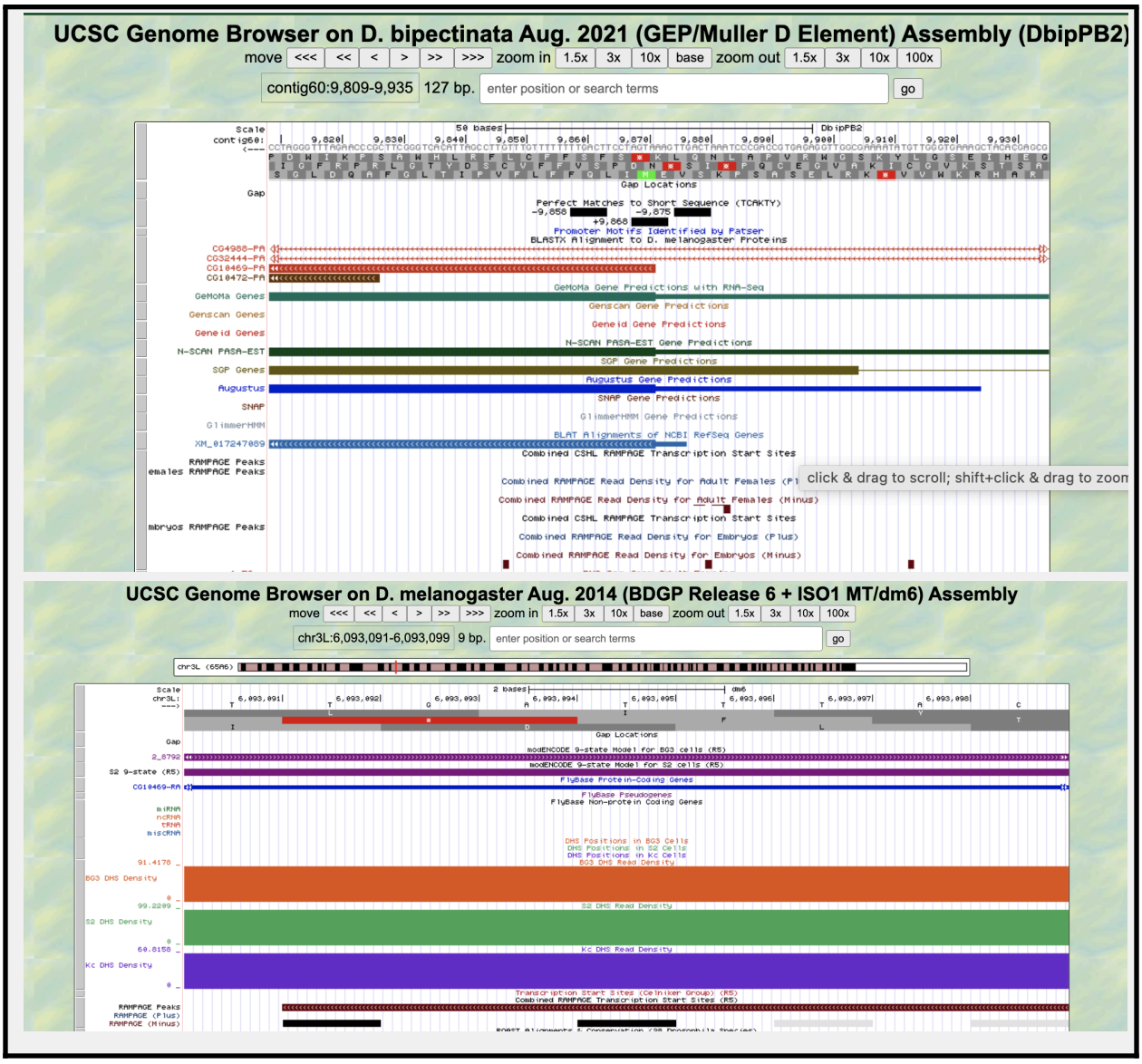
The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

1. Turn on RAMPAGE evidence tracks (Only applies to projects with these tracks)

Coordinates of the TSS position based on position with the highest RAMPAGE read density
Bipectinata: 10023-9877
Melanogaster: 6,093,098-6,093,091

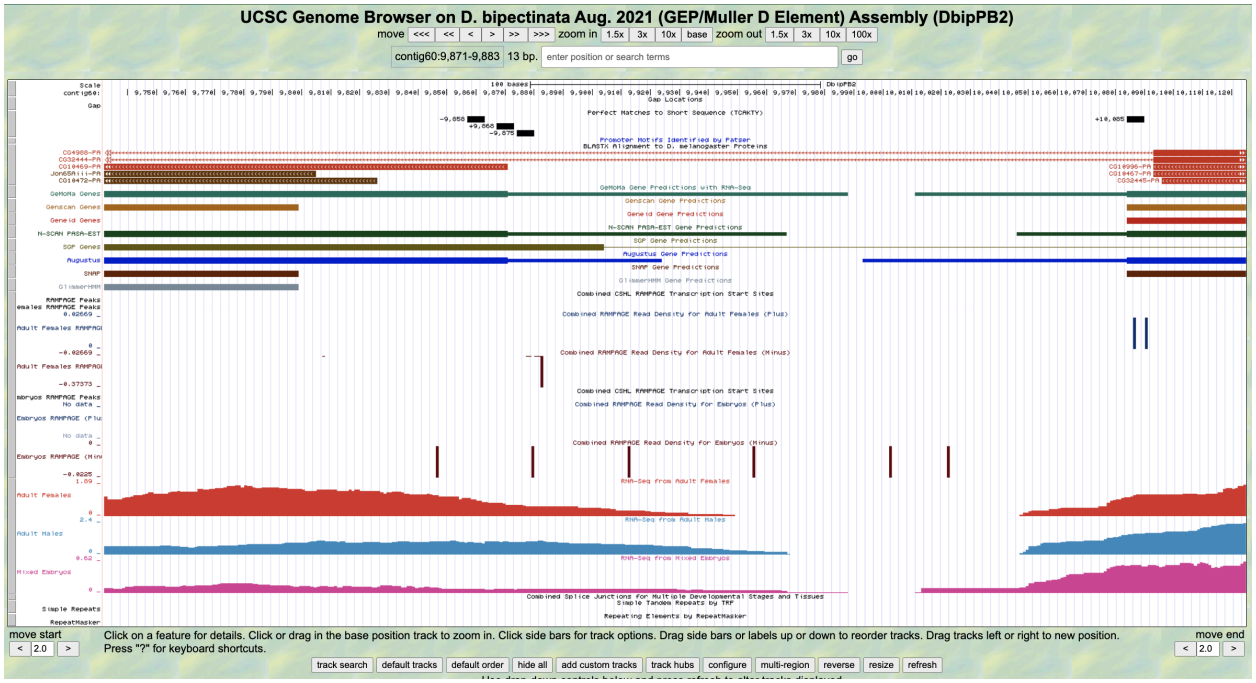
Coordinates of the narrow TSS search region based on RAMPAGE peaks
Bipectinata: 9880
Melanogaster: 6,093,095 and 6,093,092

If the TSS position and narrow TSS search region are supported by RAMPAGE data, **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the Combined RAMPAGE TSS evidence track:**



2. Turn on RNA-Seq evidence tracks

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (*e.g.*, *regtools*), **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the following evidence tracks:**



1. RNA-Seq Coverage or RNA-Seq Alignment Summary
2. Combined Splice Junctions or RNA-Seq *TopHat*

If the RNA-Seq evidence tracks indicate a TSS search region, list it here: downstream of 9950

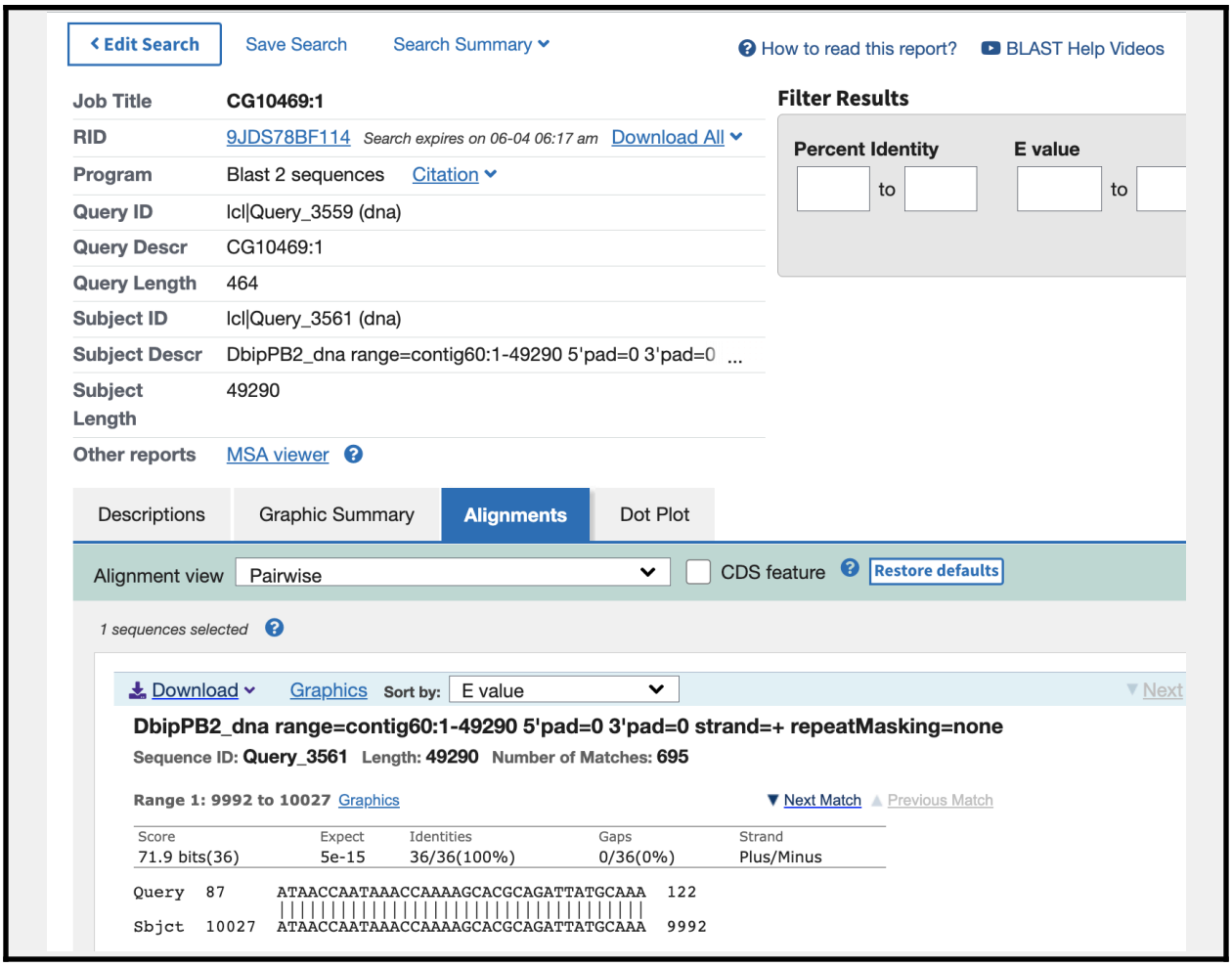
3. Annotate the first transcribed exon

Coordinates of the first transcribed exon based on *BLASTN* alignment:
10027-9992

Does the *BLASTN* alignment cover the entire *D. melanogaster* first transcribed exon?
No

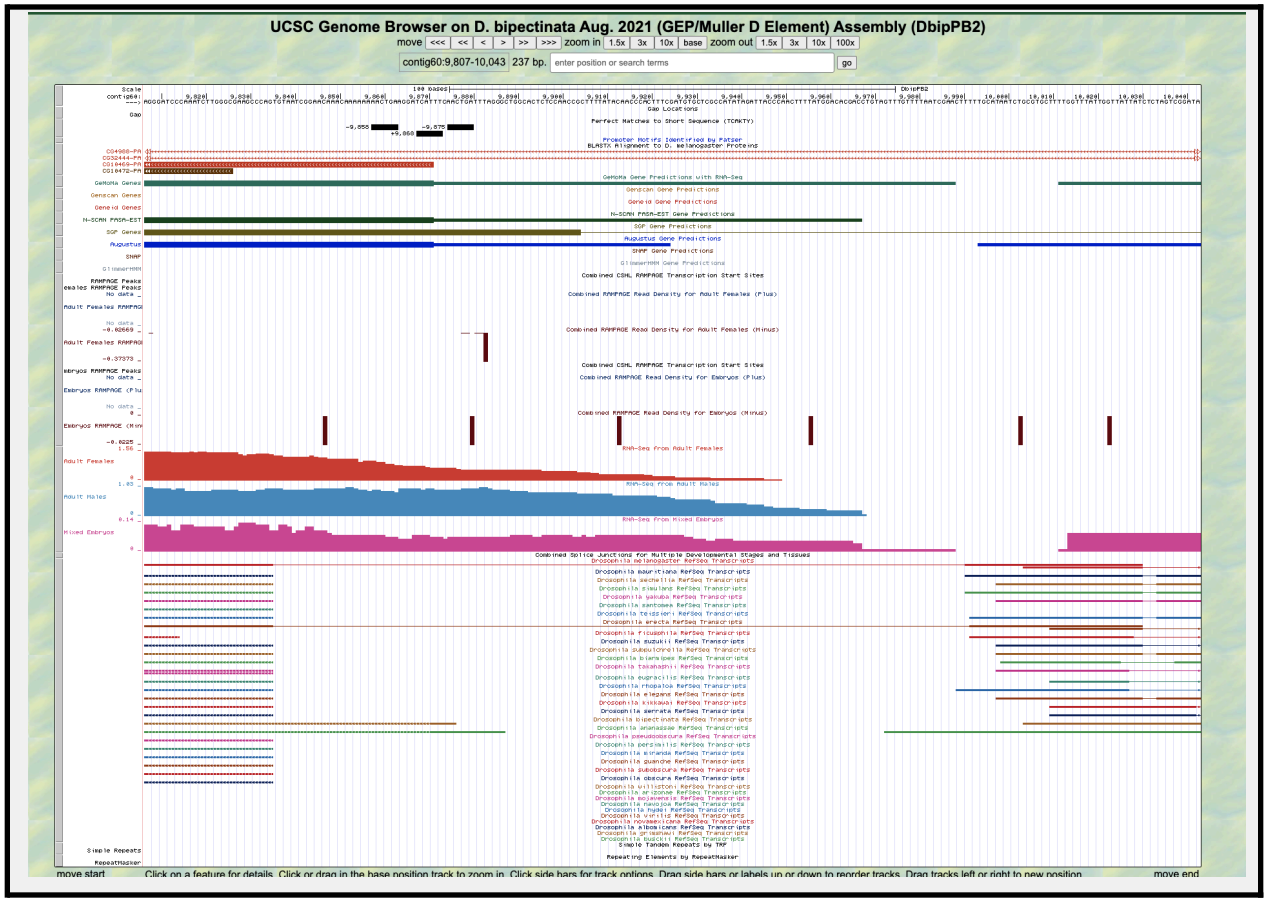
If not, specify the parts of the *D. melanogaster* exon that are missing from the *BLASTN* alignment.
Covers 87-122. Missing 1-187 and 88-464

If the TSS annotation is supported by *BLASTN* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *BLASTN* alignment into the box below:**


Edit Search
Save Search
Search Summary
How to read this report?
BLAST Help Videos
Job Title CG10469:1
RID 9JDS78BF114 Search expires on 06-04 06:17 am Download All
Program Blast 2 sequences Citation
Query ID lcl|Query_3559 (dna)
Query Descr CG10469:1
Query Length 464
Subject ID lcl|Query_3561 (dna)
Subject Descr DbipPB2_dna range=contig60:1-49290 5'pad=0 3'pad=0 ...
Subject Length 49290
Other reports MSA viewer
Filter Results
Percent Identity to
E value to
Descriptions
Graphic Summary
Alignments
Dot Plot
Alignment view Pairwise
CDS feature
Restore defaults
1 sequences selected
Download
Graphics
Sort by: E value
Next
DbipPB2_dna range=contig60:1-49290 5'pad=0 3'pad=0 strand=+ repeatMasking=none
Sequence ID: Query_3561 Length: 49290 Number of Matches: 695
Range 1: 9992 to 10027 Graphics
Next Match
Previous Match
Score 71.9 bits(36)
Expect 5e-15
Identities 36/36(100%)
Gaps 0/36(0%)
Strand Plus/Minus
Query 87 ATAACCAATAAACCAAAAGCACGCAGATTATGCAAA 122
Sbjct 10027 ATAACCAATAAACCAAAAGCACGCAGATTATGCAAA 9992

4. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, **paste a screenshot of the multiple sequence alignment (*e.g.*, from *Clustal Omega*, *ROAST*) into the box below:**



5. Summarize the evidence that supports the TSS annotation postulated above

Coordinate(s) of the TSS position(s):

Based on RAMPAGE data (if applicable): 10023-9877
Based on RNA-Seq data: N/A (unable to get RNA Seq alignment)
Based on *BLASTN* alignment: 10027-9992
Based on other evidence (ShortMatch (TCAKTY): 9873

Note: If the *BLASTN* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter "Insufficient evidence" if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS position based on the available evidence? NO

If not, were you able to define a TSS search region? YES, 10023-9873
If so, indicate in the table below the evidence that supports the TSS search region(s)

For each evidence type, enter an "X" in the cell to indicate whether the line of evidence supports, refutes, or neither supports nor refutes the TSS annotation:

Evidence type	Support	Refute	Neither
RAMPAGE peaks and read density	X		
RNA-Seq coverage and splice junctions			X
BLASTN alignment of the initial exon from *D. melanogaster*	X		
Sequence conservation with other *Drosophila* species (*e.g.,* "Conservation" track on the Genome Browser)	X		
Other (ShortMatch (TCAKTY))	X		

Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against the annotated TSS; check "Refute." In contrast, the lack of RNA-Seq read coverage is a negative result that neither supports nor refutes the TSS annotation; check "Neither."

Provide an explanation if the TSS annotation is inconsistent with at least one of the evidence types specified above:

Gene Report Form

Gene name: *D. bipectinata CG10467*
Gene symbol: *dbip_CG10467*

Approximate location in project (from 5' end to 3' end): 10094-11179
Number of isoforms in *D. melanogaster:* 1
Number of isoforms in this project: 1

Complete the following table, including all of the isoforms in this project:

Name(s) of unique isoform(s) based on coding sequence	List of isoforms with identical coding sequences
CG10467-PA	

Names of the isoforms with unique coding sequences in *D. melanogaster* that are absent in this species: N/A (removed report form from document due to N/A)

Isoform Report Form

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

Gene-isoform symbol: dbip_CG10467 - PA

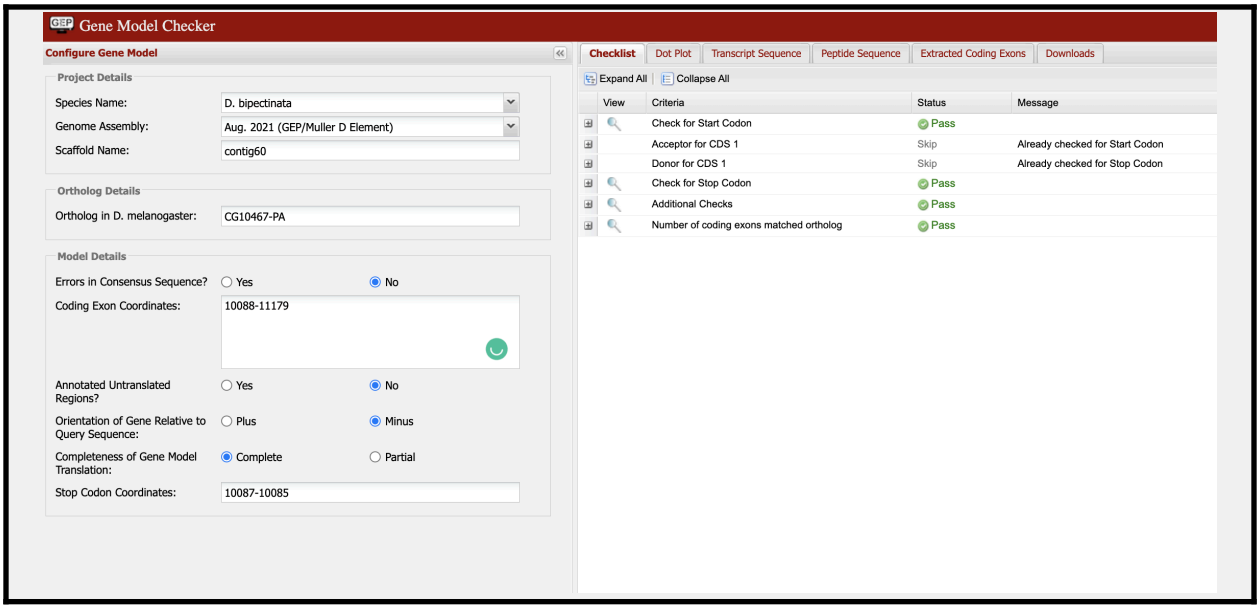
Names of any additional isoforms with identical coding sequences:
N/A

Is the 5' end of this isoform missing from the end of the project? NO
Is the 3' end of this isoform missing from the end of the project? NO

1. *Gene Model Checker* checklist

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

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

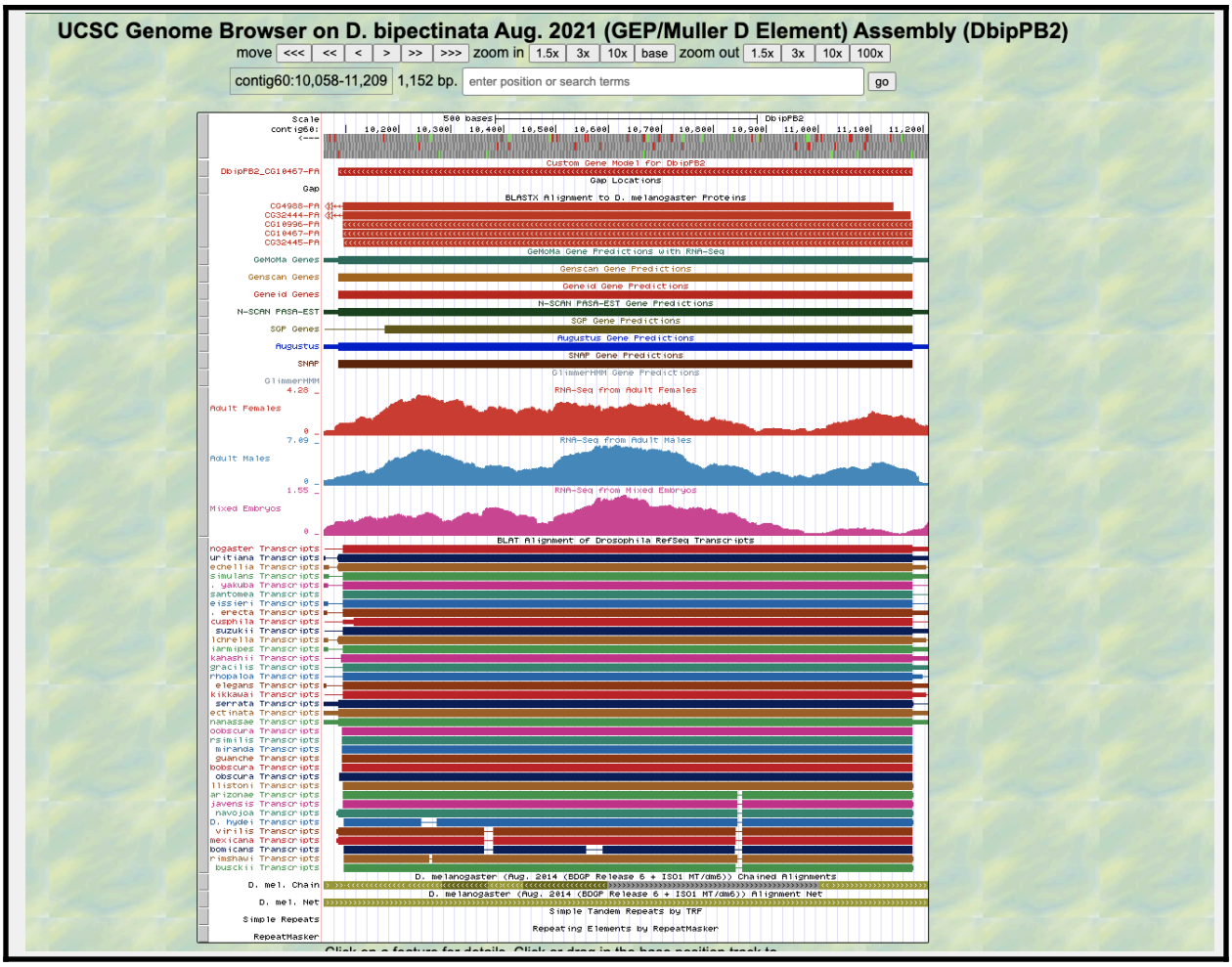


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

3. A sequence alignment track (*e.g.*, D. mel Proteins)
4. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
5. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
6. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box below:


UCSC Genome Browser on D. bipectinata Aug. 2021 (GEP/Muller D Element) Assembly (DbipPB2)
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
contig60:10,058-11,209 1,152 bp. enter position or search terms go
Scale
contig60:
500 bases
DbipPB2
10,200| 10,300| 10,400| 10,500| 10,600| 10,700| 10,800| 10,900| 11,000| 11,100| 11,200|
Custom Gene Model for DbipPB2
DbipPB2_CG10467-PA
Gap
Gap Locations
BLASTX Alignment to D. melanogaster Proteins
CG4988-PA
CG32444-PA
CG10996-PA
CG10467-PA
CG32445-PA
GeMoMa Gene Predictions with RNA-Seq
GeMoMa Genes
Genscan Gene Predictions
Genscan Genes
Geneid Gene Predictions
Geneid Genes
N-SCAN PASA-EST Gene Predictions
N-SCAN PASA-EST
SGP Gene Predictions
SGP Genes
Augustus Gene Predictions
Augustus
SNAP Gene Predictions
SNAP
GlimmerHMM Gene Predictions
GlimmerHMM
4.28
RNA-Seq from Adult Females
Adult Females
0
7.09
RNA-Seq from Adult Males
Adult Males
0
1.55
RNA-Seq from Mixed Embryos
Mixed Embryos
0
BLAT Alignment of Drosophila RefSeq Transcripts
nogaster Transcripts
uritiana Transcripts
echellia Transcripts
simulans Transcripts
. yakuba Transcripts
santomea Transcripts
eissieri Transcripts
. erecta Transcripts
ugphila Transcripts
suzukii Transcripts
ichrella Transcripts
ianaipes Transcripts
kanashii Transcripts
gracilis Transcripts
rhopaloa Transcripts
elegans Transcripts
kikkawai Transcripts
serrata Transcripts
ectinata Transcripts
nanassae Transcripts
obscura Transcripts
rsimilis Transcripts
miranda Transcripts
guanche Transcripts
bobscura Transcripts
obscura Transcripts
lliston Transcripts
arizonae Transcripts
javensis Transcripts
navojoa Transcripts
D. hydei Transcripts
virilis Transcripts
mexicana Transcripts
bomicans Transcripts
rimshawi Transcripts
busckii Transcripts
D. melanogaster (Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6)) Chained Alignments
D. mel. Chain
D. melanogaster (Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6)) Alignment Net
D. mel. Net
Simple Tandem Repeats by TRF
Simple Repeats
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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**

Alignment of Dmel_CG10467-PA vs. DbipPB2_CG10467-PA

View plain text version

Download alignment image

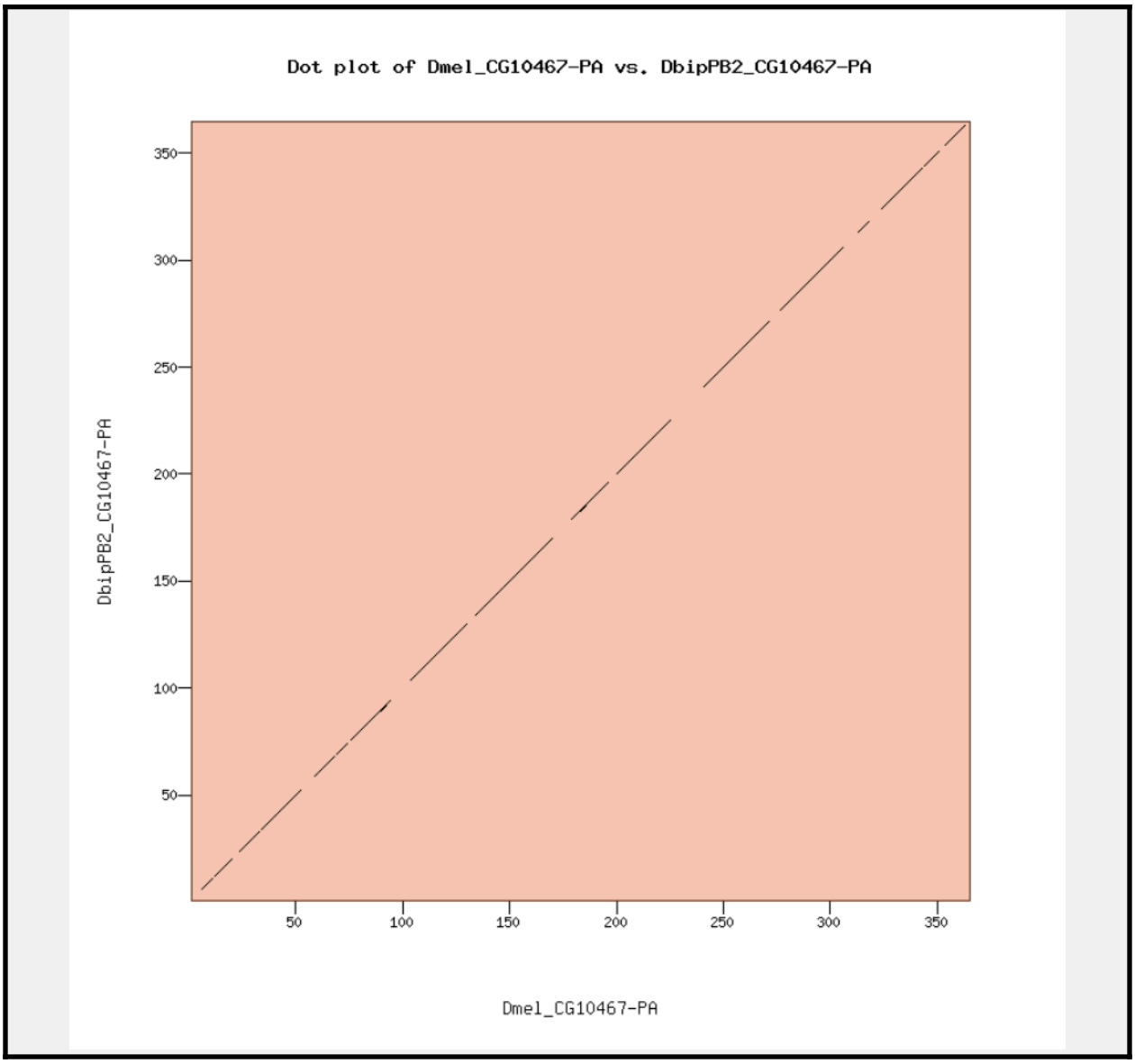
Identity: 313/364 (86.0%), **Similarity:** 336/364 (92.3%), **Gaps:** 0/364 (0.0%)

```
Dmel_CG10467-PA         1 MVNVTEDVFANGAVNPFTKSKDTIKRFTLTNGAGMSVQLITRGATITSIKTPDASGQIDD 60
                          ****.*****.********:.: ********* ****************** *:**::**
DbipPB2_CG10467-PA      1 MVNVKEDVFATGAVNPFTKATEDIKRFTLTNGYGMSVQLITRGATITSIKTMDSSGKVDD 60

Dmel_CG10467-PA        61 VTLGFDDLAGYQSERNPYFGATIGRVCNRIANGSFYLDGKLVQVSKNRDNKFQLHGGFVG 120
                          *******:*****::****************** * *:** ::*****:***********
DbipPB2_CG10467-PA     61 VTLGFDDVAGYQSDKNPYFGATIGRVCNRIANGRFQLNGKWIEVSKNRNNKFQLHGGFVG 120

Dmel_CG10467-PA       121 FDKAHWEVVEVRVDGVTLSHTNPDGHEGYPGKVTATASFTLSEDNCLHVQMSALADKTTP 180
                          ********* ** ********************************** *:* :**.**
DbipPB2_CG10467-PA    121 FDKAHWEVVAVRRDGVTLSHTNPDGHEGYPGKVTATASFTLSEDNCLHVLMTAETDKVTP 180

Dmel_CG10467-PA       181 VNLTNHSYFNLAGHKSGANGLYEHTIEINAYGITETDQSSIPTGRITPVEGTGFDLRVSS 240
                          ***************:**.*****:*************.*****:****:** :****:.
DbipPB2_CG10467-PA    181 VNLTNHSYFNLAGHKTGASGLYEHAIEINAYGITETDQDSIPTGKITPVNGTPYDLRVAG 240

Dmel_CG10467-PA       241 NLGERLKALQPARGYDDNFCVTFSPPQPLAKVARATHPPSGRWLEVVSNQPGVQFYTSNF 300
                          ****************************** ****:************************
DbipPB2_CG10467-PA    241 NLGERLKALQPARGYDDNFCVTFSPPQPLAMVARASHPPSGRWLEVVSNQPGVQFYTSNF 300

Dmel_CG10467-PA       301 MPDVERGESPIPGKDGAAYAKHCAFCLETQKFPDSVNHSNFPSTILRPGESYQHEVIYKF 360
                          *****.**:**.*****:*.** ********************:******* *.*******
DbipPB2_CG10467-PA    301 MPDVENGEAPISGKDGASYGKHGAFCLETQKFPDSVNHSNFPTTILRPGERYSHEVIYKF 360

Dmel_CG10467-PA       361 GVSH 364
                          ** :
DbipPB2_CG10467-PA    361 GVFN 364
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Transcription Start Sites (TSS) Report Form (Optional)

Note: Complete this section if you have annotated the TSS for the gene above. This section is **optional** and you do not need to complete this section to submit the project.

Gene name : *D. Bipectinata CG10467*
Gene symbol: dbip_CG10467

Name(s) of isoform(s) with unique TSS	List of isoforms with identical TSS
CG10467-PA	

Names of the isoforms with unique TSS in *D. melanogaster* that are absent in this species:
N/A

Provide the evidence (text and figures) which support the hypothesis that these isoforms are absent in this species (*e.g.*, changes in canonical splice sites, gene structure, etc.):

Isoform TSS Report

Complete an Isoform TSS report (through page PAGEREF _Ref91436230 \h 17) for each unique TSS listed in the table above. Copy and paste this form to create as many copies as needed.

Gene-isoform name : *dbip_CG10467-pA*

Names of the isoforms with the same TSS as this isoform:
N/A

Type of core promoter in *D. melanogaster* (see table below):
(Peaked / Intermediate / Broad / Insufficient Evidence)
Peaked

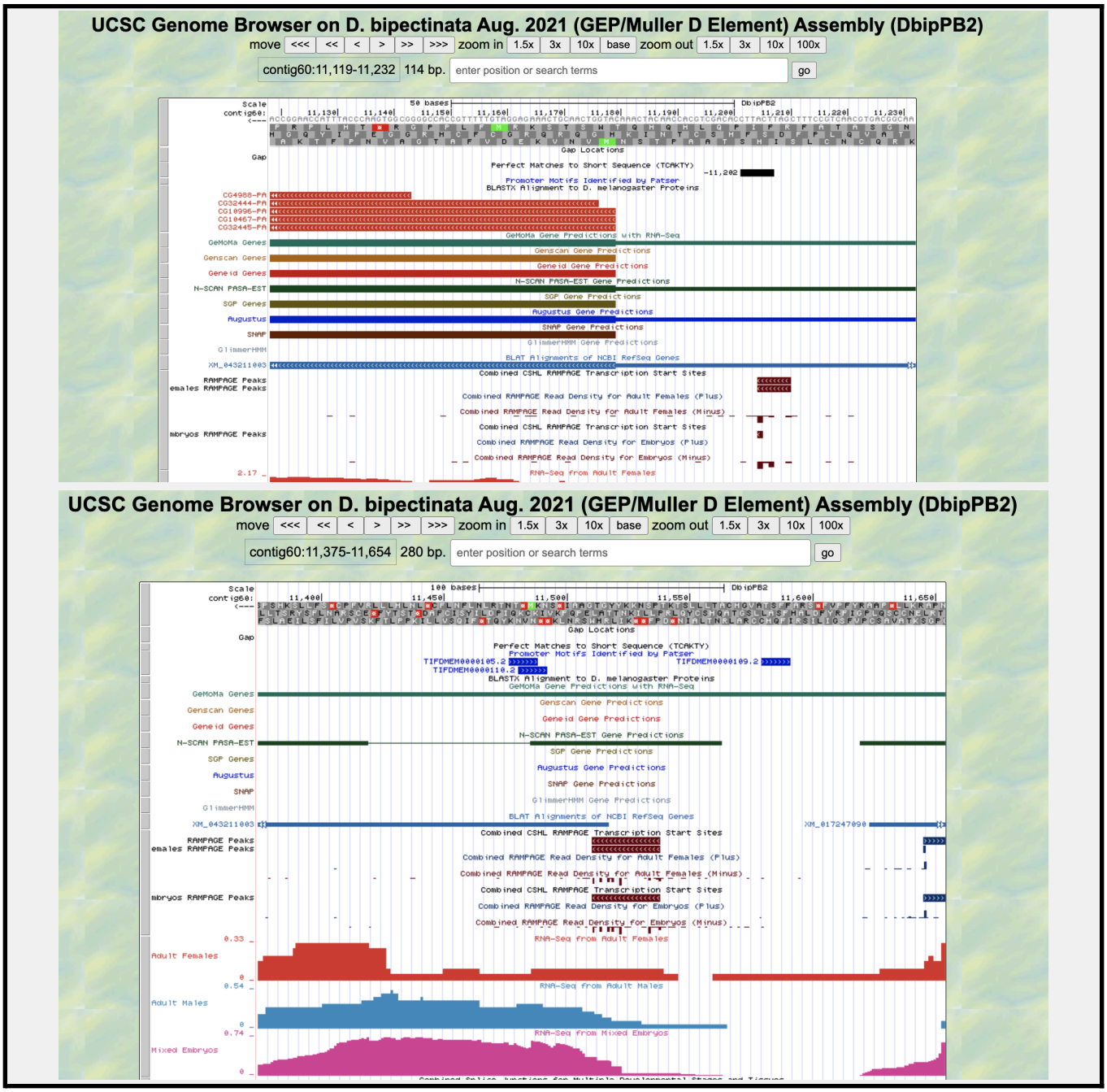
The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

1. Turn on RAMPAGE evidence tracks (Only applies to projects with these tracks)

Coordinates of the TSS position based on position with the highest RAMPAGE read density
11598-11161

Coordinates of the narrow TSS search region based on RAMPAGE peaks
11538-11510 and 11210-11204

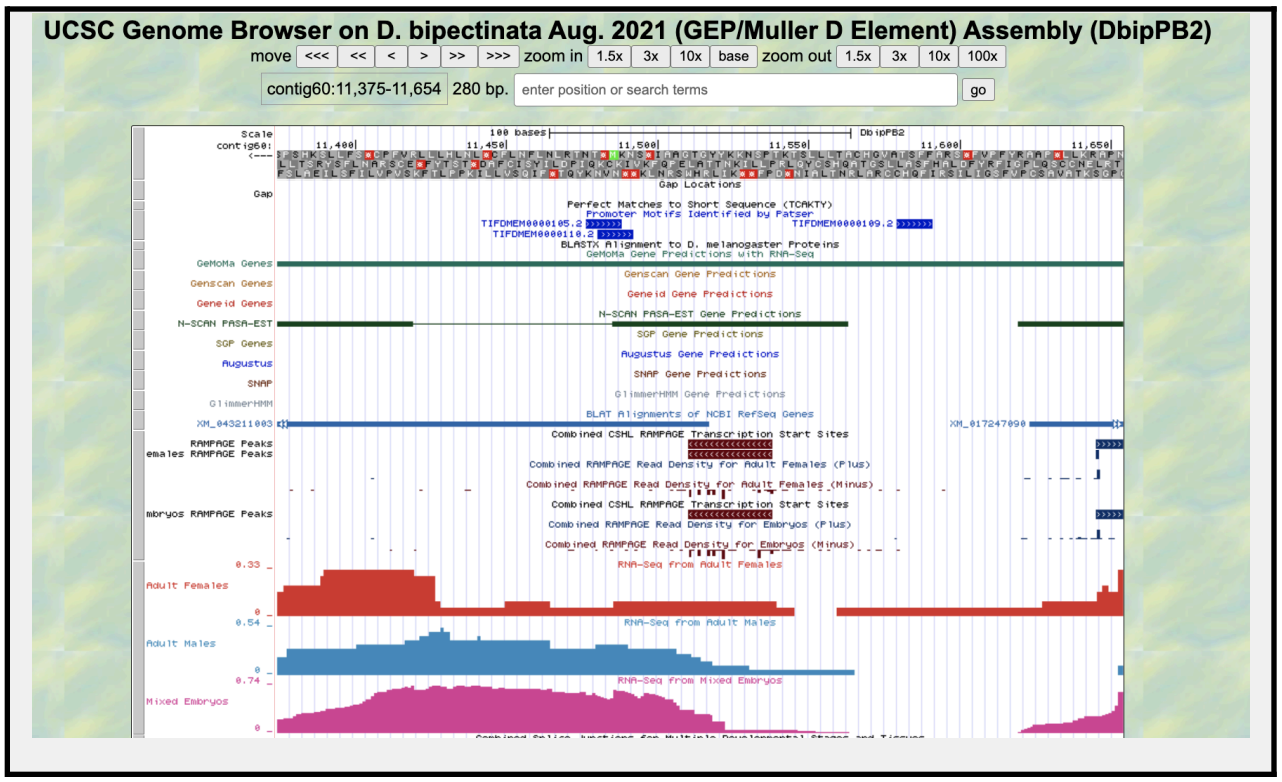
If the TSS position and narrow TSS search region are supported by RAMPAGE data, **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the Combined RAMPAGE TSS evidence track:**



2. Turn on RNA-Seq evidence tracks

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (*e.g.*, *regtools*), **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the following evidence tracks:**

7. RNA-Seq Coverage or RNA-Seq Alignment Summary
8. Combined Splice Junctions or RNA-Seq *TopHat*



If the RNA-Seq evidence tracks indicate a TSS search region, list it here: Adult Females/Males and mixed embryos because they are starting near the RAMPAGE site showing that TSS could be starting in this region.

3. Annotate the first transcribed exon

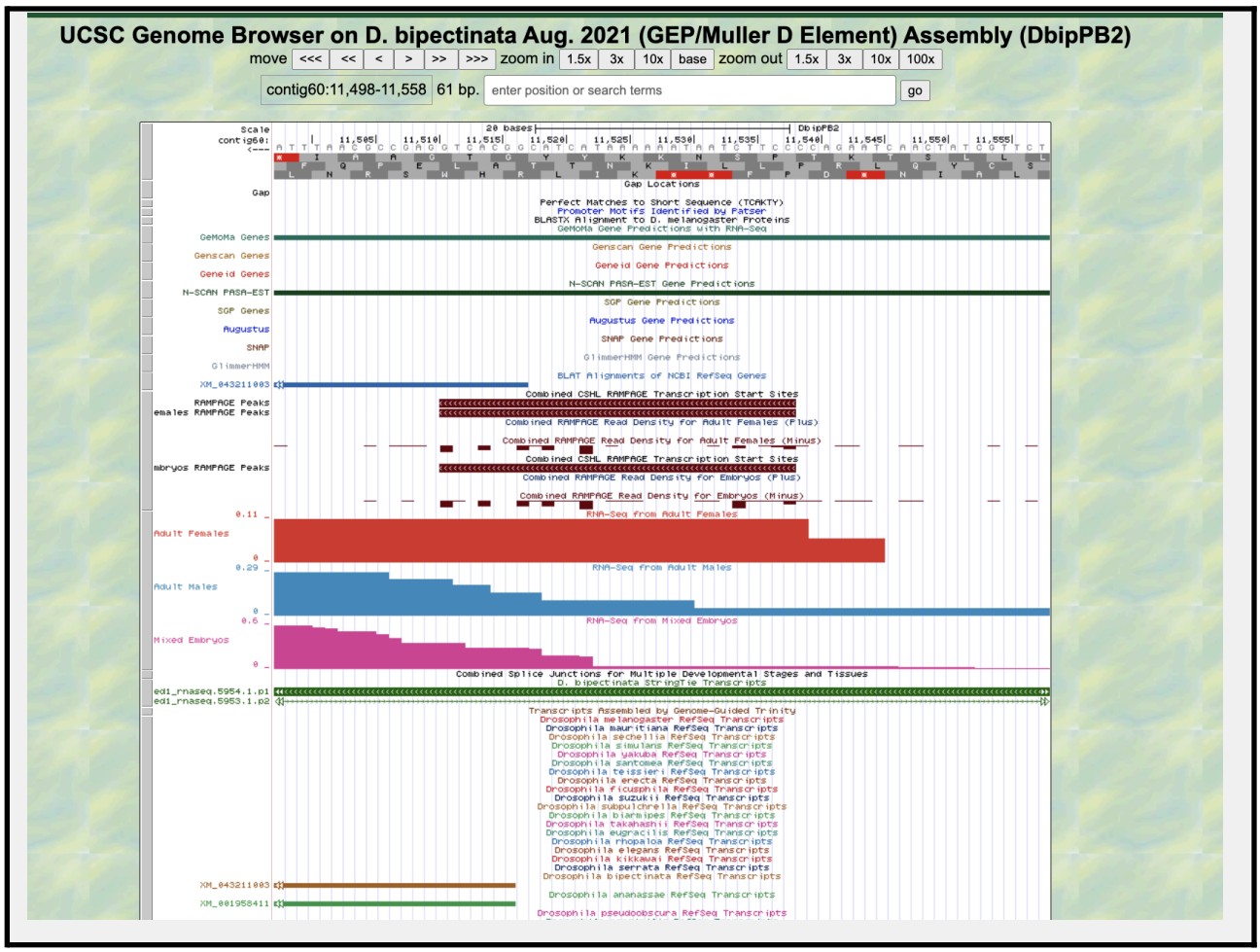
Coordinates of the first transcribed exon based on *BLASTN* alignment:
10101-11279

Does the *BLASTN* alignment cover the entire *D. melanogaster* first transcribed exon?
NO
If not, specify the parts of the *D. melanogaster* exon that are missing from the *BLASTN* alignment.
Covers 253-1431, missing 1-253 and 1431-1544

If the TSS annotation is supported by *BLASTN* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *BLASTN* alignment into the box below:**

4. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, **paste a screenshot of the multiple sequence alignment (*e.g.*, from *Clustal Omega*, *ROAST*) into the box below:**



5. Summarize the evidence that supports the TSS annotation postulated above

Coordinate(s) of the TSS position(s):

Based on RAMPAGE data (if applicable): range: 11598-11161 peak: 11538-11510
Based on RNA-Seq data: near 11545
Based on *BLASTN* alignment: 10101-11279
Based on other evidence (Shortmatch(TCAKTY))): 11,200

Note: If the *BLASTN* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter "Insufficient evidence" if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS position based on the available evidence? NO
If so, indicate in the table below the evidence that supports this TSS position

If not, were you able to define a TSS search region?11545- 11538
If so, indicate in the table below the evidence that supports the TSS search region(s)

For each evidence type, enter an "X" in the cell to indicate whether the line of evidence supports, refutes, or neither supports nor refutes the TSS annotation:

Evidence type	Support	Refute	Neither
RAMPAGE peaks and read density	X		
RNA-Seq coverage and splice junctions	X		
BLASTN alignment of the initial exon from *D. melanogaster*		X	
Sequence conservation with other *Drosophila* species (*e.g.*, "Conservation" track on the Genome Browser)	X		
Other (ShortMatch)		X	

Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against the annotated TSS; check "Refute." In contrast, the lack of RNA-Seq read coverage is a negative result that neither supports nor refutes the TSS annotation; check "Neither."

Provide an explanation if the TSS annotation is inconsistent with at least one of the evidence types specified above:

Gene Report Form

Gene name (*e.g.*, *D. bipectinata eyeless*): *D. bipectinata logjam*
Gene symbol (*e.g.*, *dbip_ey*): *dbip_loj*

Approximate location in project (from 5' end to 3' end): 11,848 – 13,107
Number of isoforms in *D. melanogaster:* 3
Number of isoforms in this project: 3

Complete the following table, including all of the isoforms in this project:

Name(s) of unique isoform(s) based on coding sequence	**List of isoforms with identical coding sequences**
loj-PD	loj-PE, loj-PF

Names of the isoforms with unique coding sequences in *D. melanogaster* that are absent in this species: None

Provide the evidence (text and figures) which support the hypothesis that these isoforms are absent in this species (*e.g.*, changes in canonical splice sites, gene structure, etc.):

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 their coding sequence is identical to that of another isoform.

Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): loj-PD

Names of any additional isoforms with identical coding sequences:
loj-PE and loj-PF

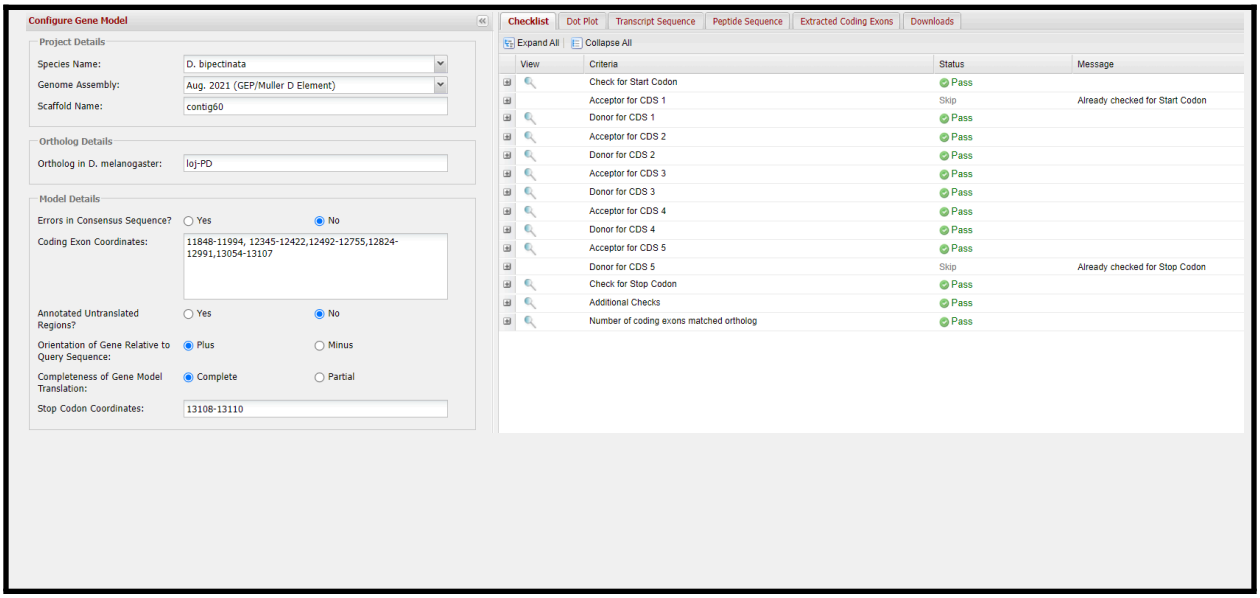
Is the 5' end of this isoform missing from the end of the project? __No_____
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? __No_____
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

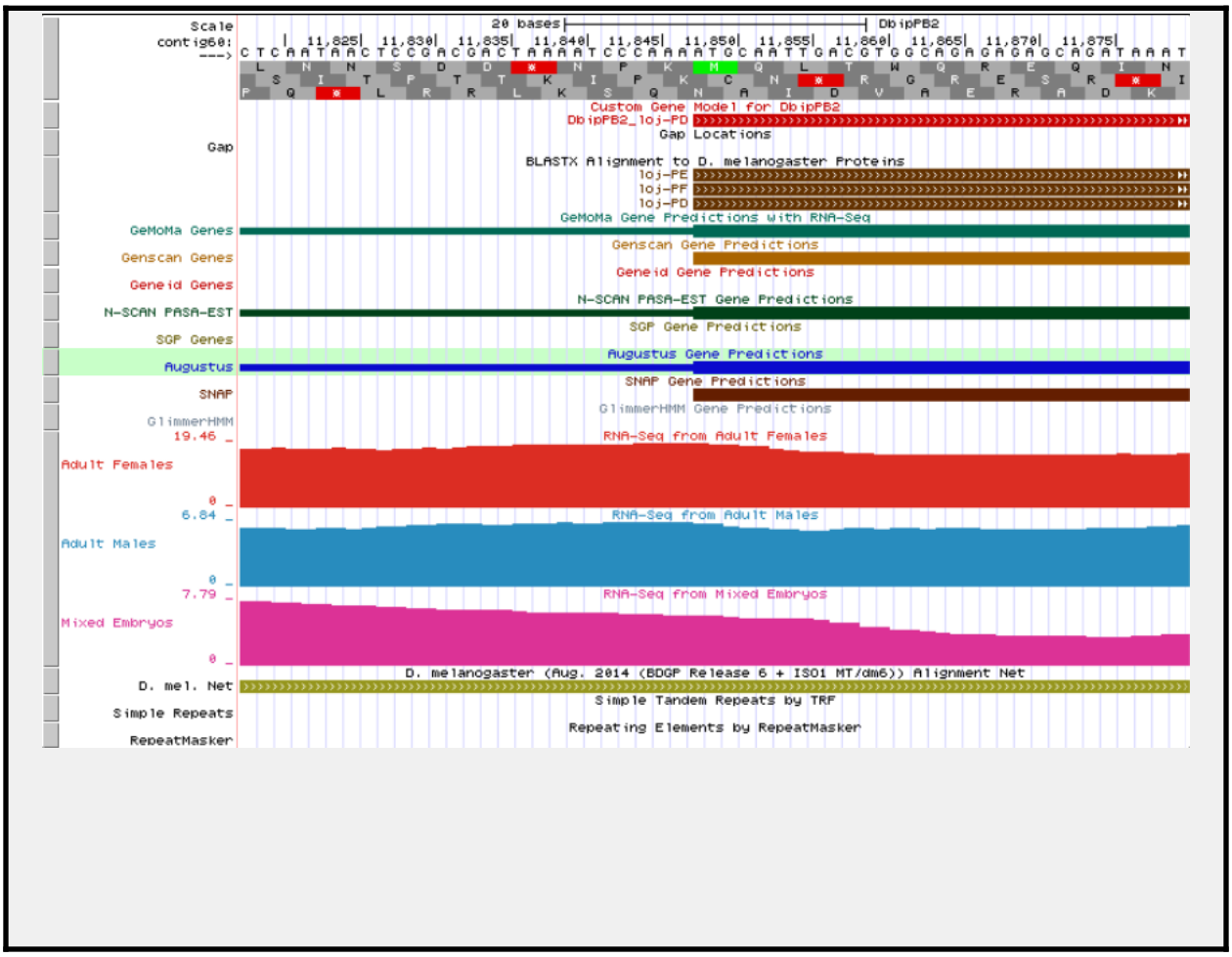


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

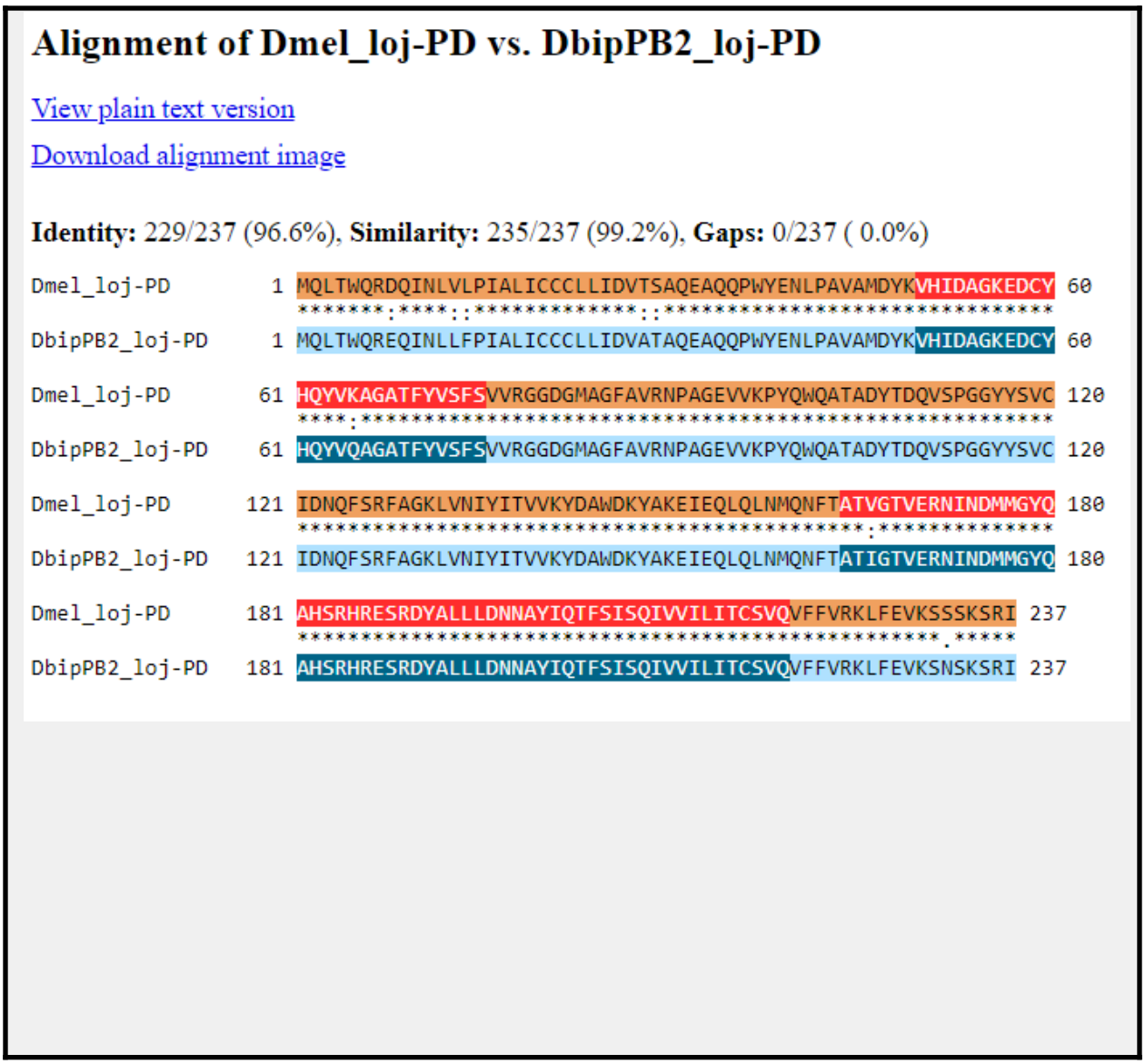
5. A sequence alignment track (*e.g.*, D. mel Proteins)
6. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
7. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
8. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_loj-PD vs. DbipPB2_loj-PD

View plain text version

Download alignment image

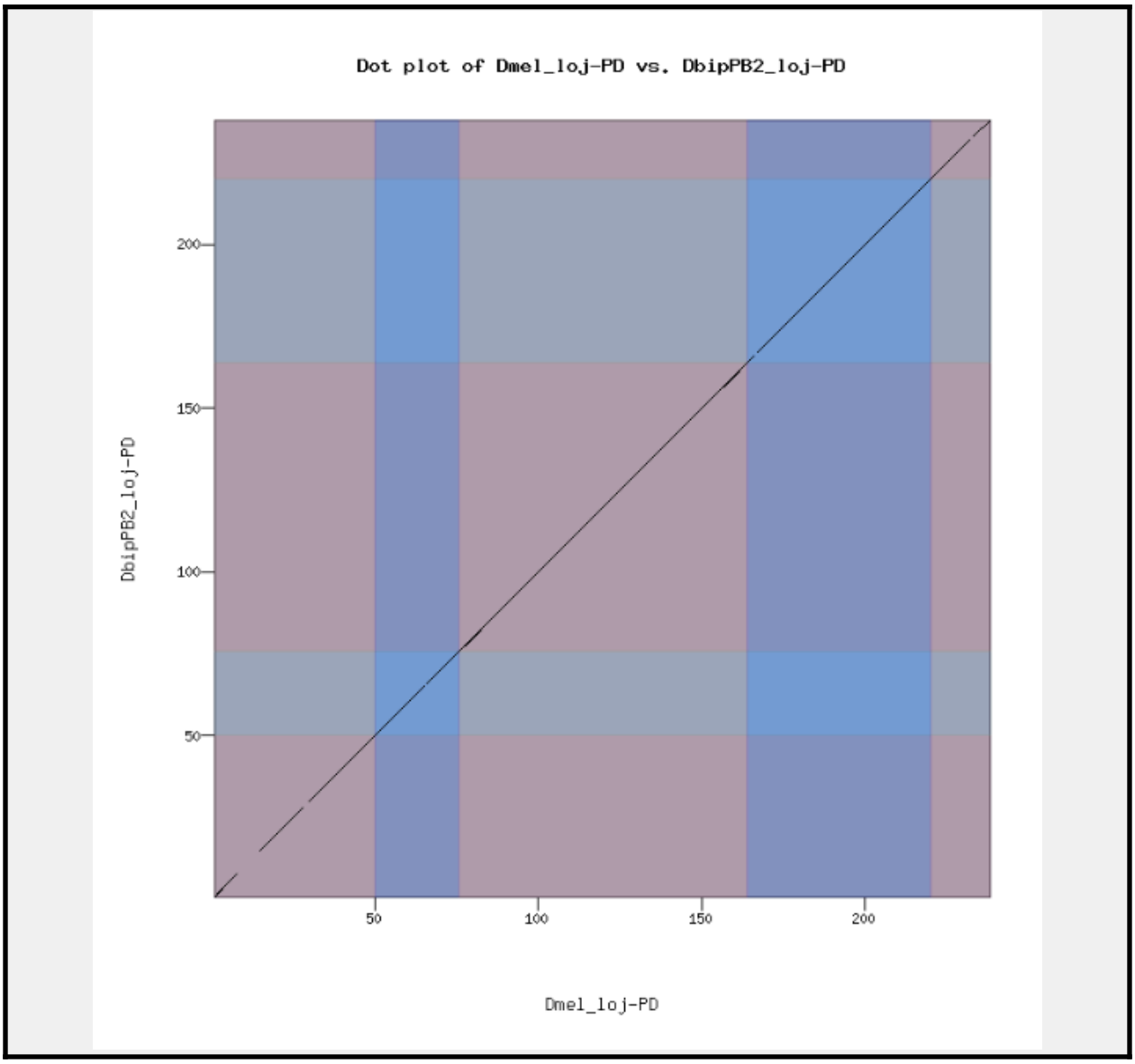
Identity: 229/237 (96.6%), **Similarity:** 235/237 (99.2%), **Gaps:** 0/237 (0.0%)

```
Dmel_loj-PD       1 MQLTWQRDQINLVLPIALICCCLLIDVTSAQEAQQPWYENLPAVAMDYKVHIDAGKEDCY 60
                    *******:****::*************::********************************
DbipPB2_loj-PD    1 MQLTWQREQINLLFPIALICCCLLIDVATAQEAQQPWYENLPAVAMDYKVHIDAGKEDCY 60

Dmel_loj-PD      61 HQYVKAGATFYVSFSVVRGGDGMAGFAVRNPAGEVVKPYQWQATADYTDQVSPGGYYSVC 120
                    ****:************************************************************
DbipPB2_loj-PD   61 HQYVQAGATFYVSFSVVRGGDGMAGFAVRNPAGEVVKPYQWQATADYTDQVSPGGYYSVC 120

Dmel_loj-PD     121 IDNQFSRFAGKLVNIYITVVKYDAWDKYAKEIEQLQLNMQNFTATVGTVERNINDMMGYQ 180
                    ********************************************:***************
DbipPB2_loj-PD  121 IDNQFSRFAGKLVNIYITVVKYDAWDKYAKEIEQLQLNMQNFTATIGTVERNINDMMGYQ 180

Dmel_loj-PD     181 AHSRHRESRDYALLLDNNAYIQTFSISQIVVILITCSVQVFFVRKLFEVKSSSKSRI 237
                    ***************************************************.*****
DbipPB2_loj-PD  181 AHSRHRESRDYALLLDNNAYIQTFSISQIVVILITCSVQVFFVRKLFEVKSNSKSRI 237
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): loj-PE

Names of any additional isoforms with identical coding sequences:
loj-PD and loj-PF

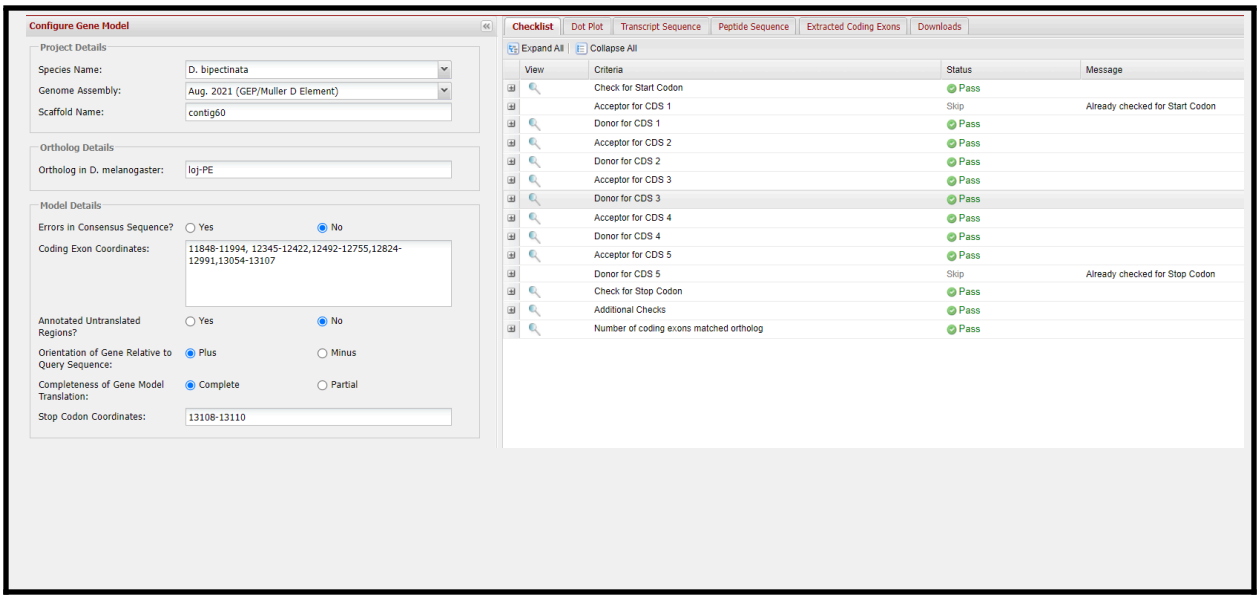
Is the 5' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

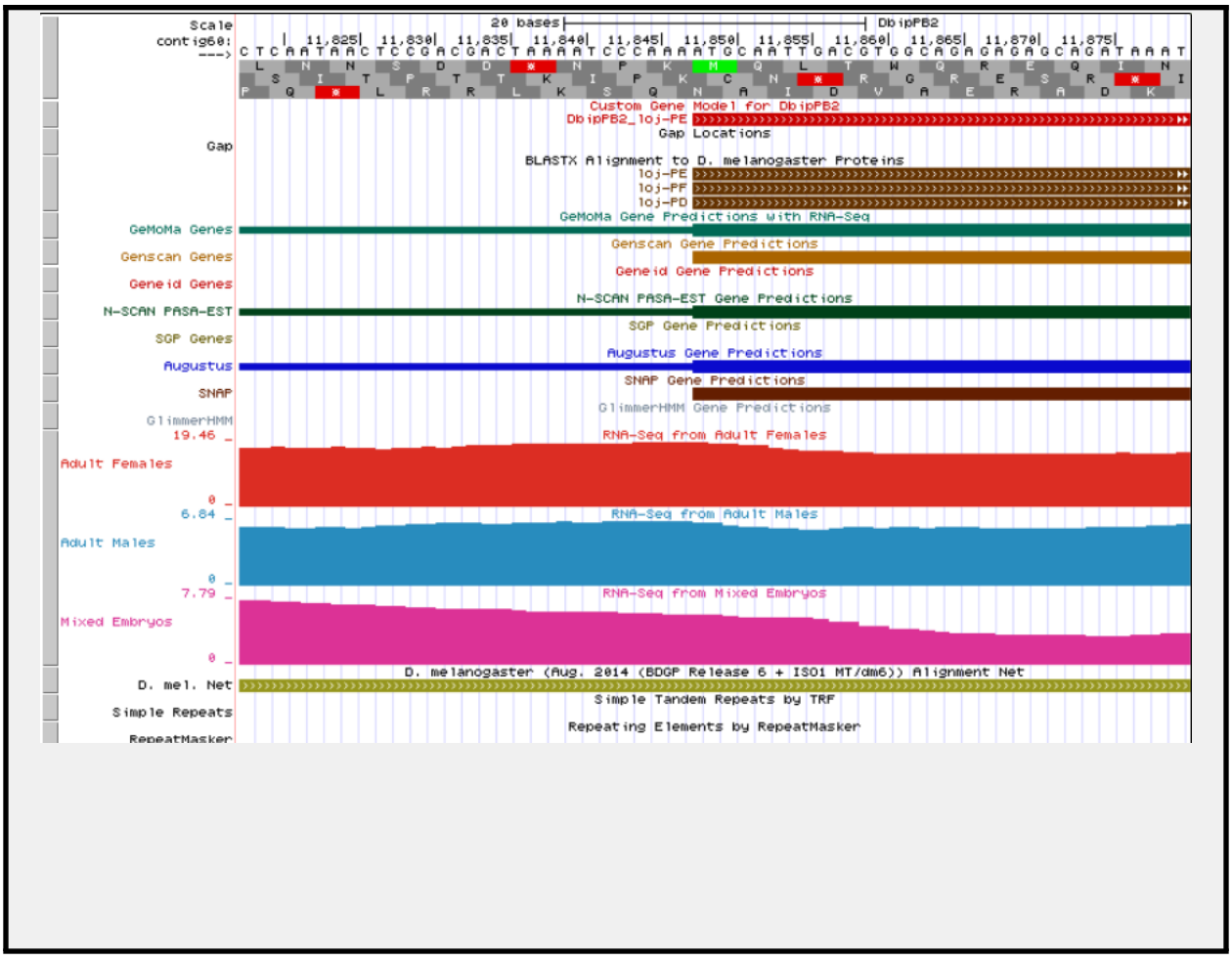


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

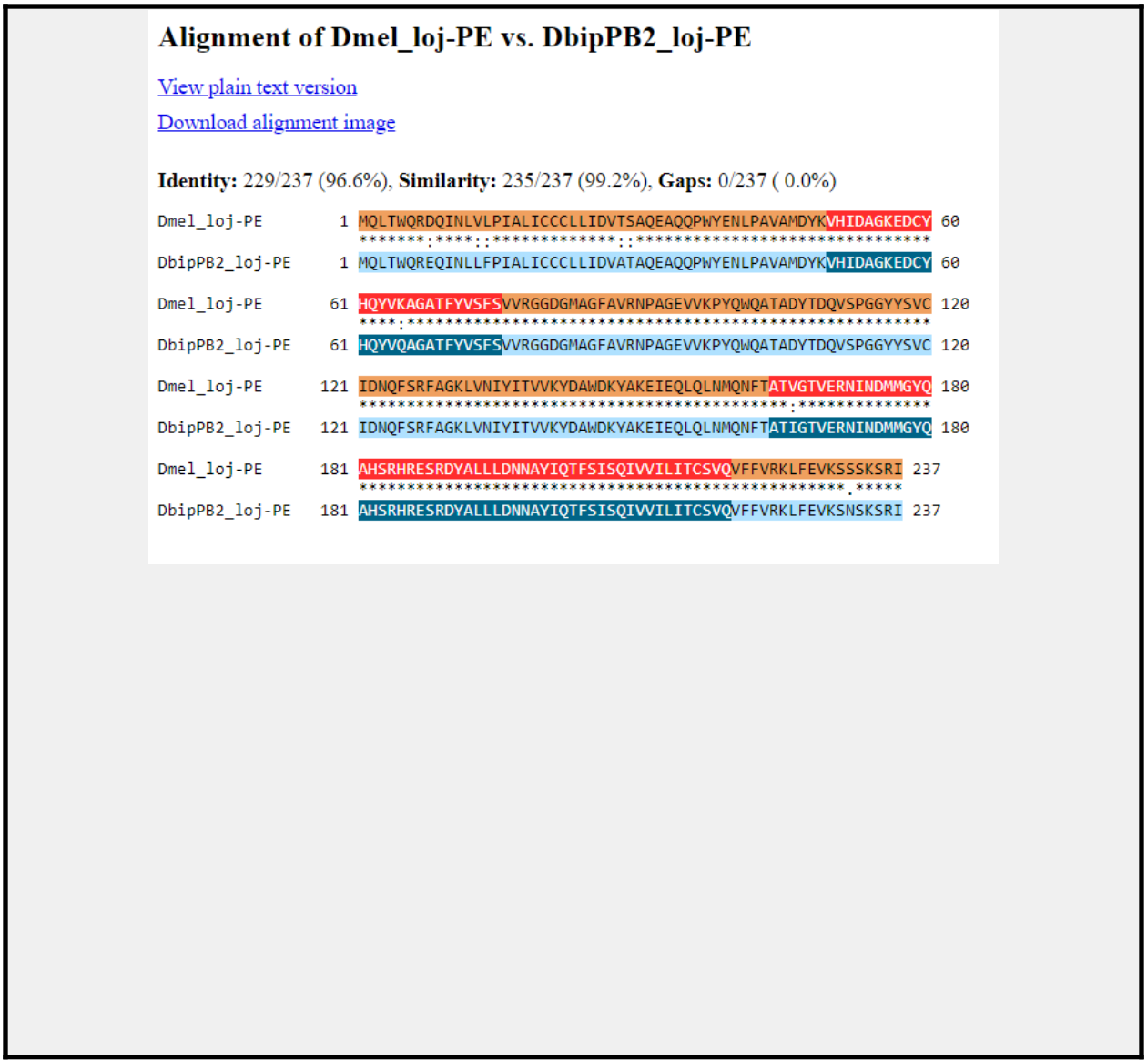
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_loj-PE vs. DbipPB2_loj-PE

View plain text version

Download alignment image

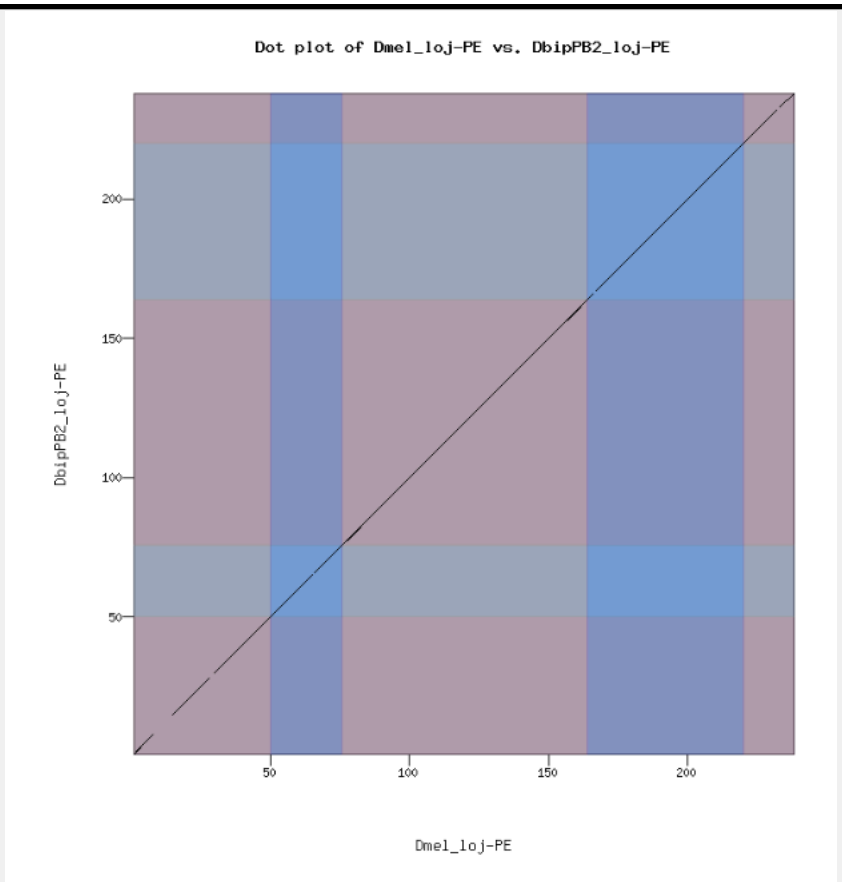
Identity: 229/237 (96.6%), **Similarity:** 235/237 (99.2%), **Gaps:** 0/237 (0.0%)

```
Dmel_loj-PE       1 MQLTWQRDQINLVLPIALICCCLLIDVTSAQEAQQPWYENLPAVAMDYKVHIDAGKEDCY 60
                    *******:****::************::******************************
DbipPB2_loj-PE    1 MQLTWQREQINLLFPIALICCCLLIDVATAQEAQQPWYENLPAVAMDYKVHIDAGKEDCY 60

Dmel_loj-PE      61 HQYVKAGATFYVSFSVVRGGDGMAGFAVRNPAGEVVKPYQWQATADYTDQVSPGGYYSVC 120
                    ****:************************************************************
DbipPB2_loj-PE   61 HQYVQAGATFYVSFSVVRGGDGMAGFAVRNPAGEVVKPYQWQATADYTDQVSPGGYYSVC 120

Dmel_loj-PE     121 IDNQFSRFAGKLVNIYITVVKYDAWDKYAKEIEQLQLNMQNFTATVGTVERNINDMMGYQ 180
                    ************************************************:*************
DbipPB2_loj-PE  121 IDNQFSRFAGKLVNIYITVVKYDAWDKYAKEIEQLQLNMQNFTATIGTVERNINDMMGYQ 180

Dmel_loj-PE     181 AHSRHRESRDYALLLDNNAYIQTFSISQIVVILITCSVQVFFVRKLFEVKSSSKSRI 237
                    ***************************************************.*****
DbipPB2_loj-PE  181 AHSRHRESRDYALLLDNNAYIQTFSISQIVVILITCSVQVFFVRKLFEVKSNSKSRI 237
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): loj-PF

Names of any additional isoforms with identical coding sequences:

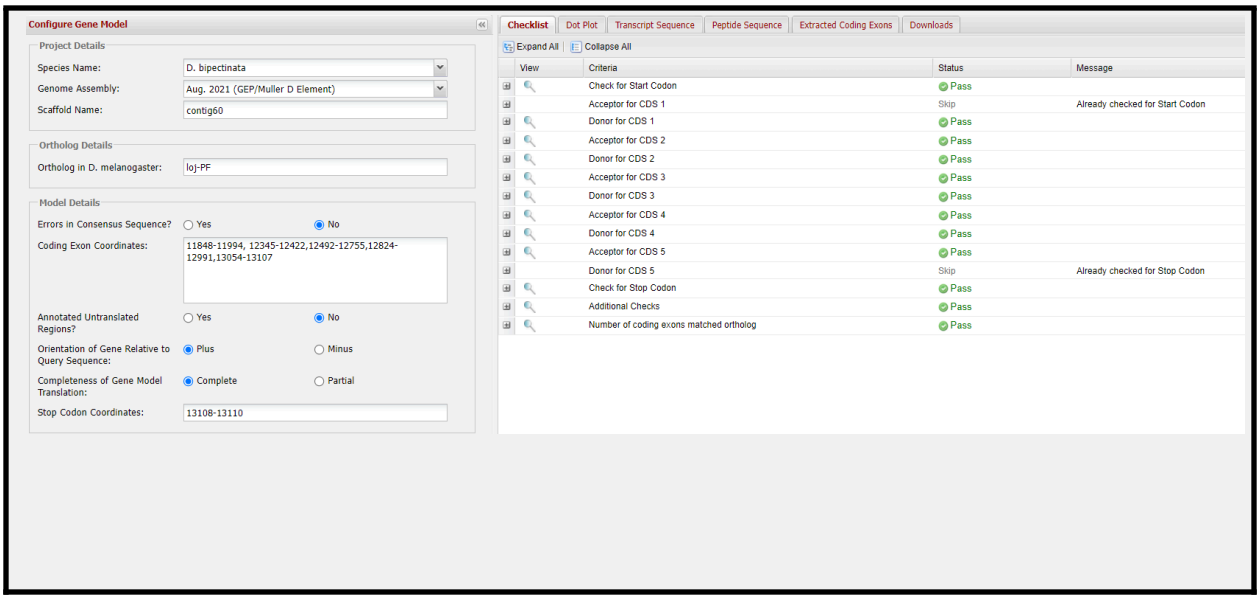
Is the 5' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

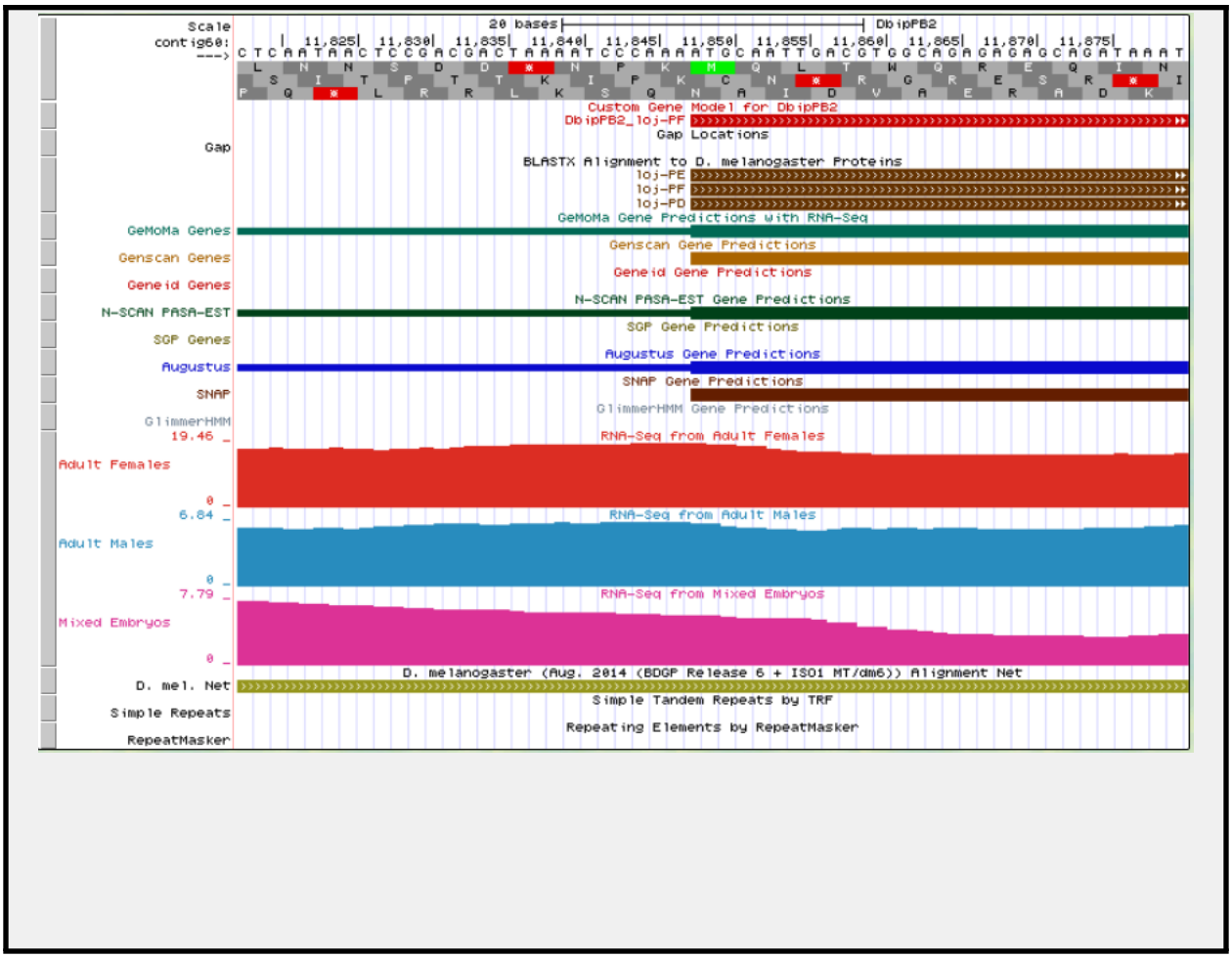


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

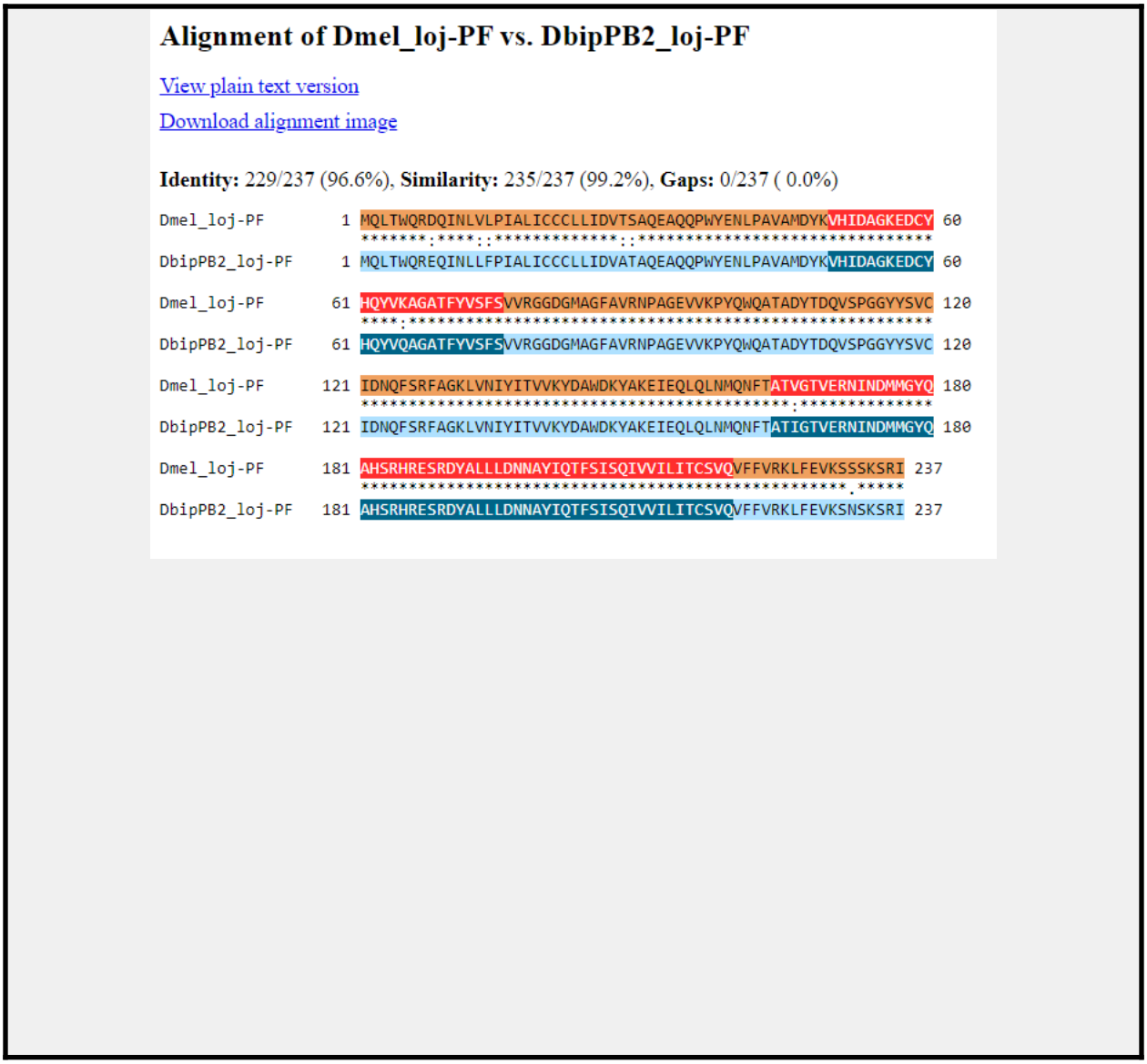
5. A sequence alignment track (*e.g.*, D. mel Proteins)
6. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
7. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
8. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_loj-PF vs. DbipPB2_loj-PF

View plain text version

Download alignment image

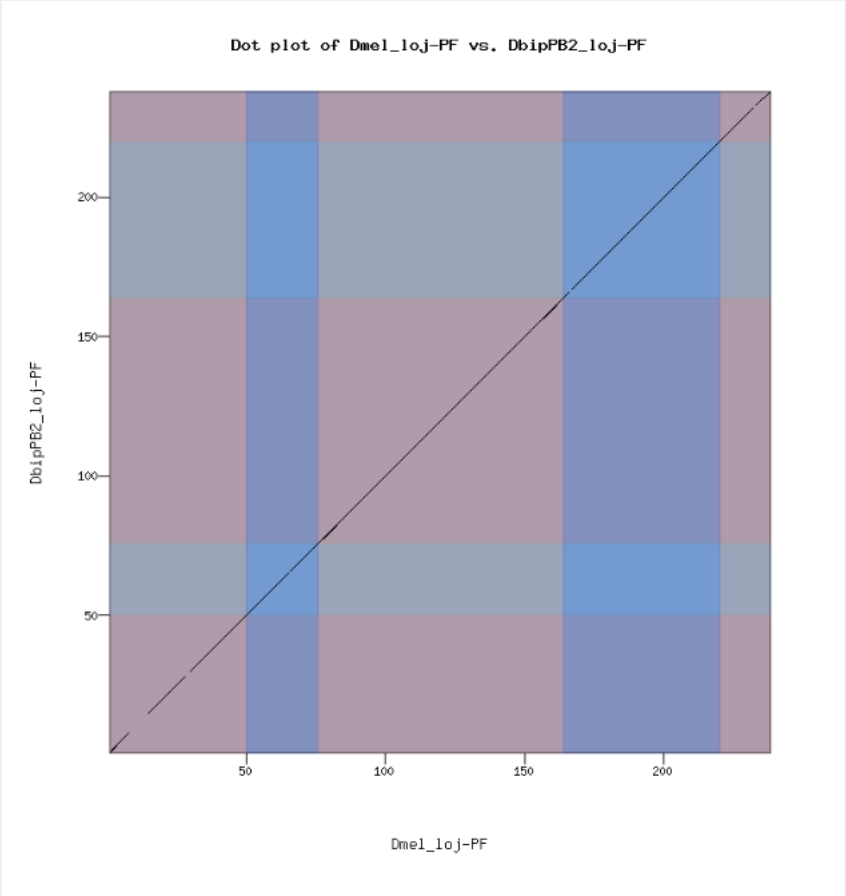
Identity: 229/237 (96.6%), **Similarity:** 235/237 (99.2%), **Gaps:** 0/237 (0.0%)

```
Dmel_loj-PF       1 MQLTWQRDQINLVLPIALICCCLLIDVTSAQEAQQPWYENLPAVAMDYKVHIDAGKEDCY 60
                    *******:****::*************::******************************
DbipPB2_loj-PF    1 MQLTWQREQINLLFPIALICCCLLIDVATAQEAQQPWYENLPAVAMDYKVHIDAGKEDCY 60

Dmel_loj-PF      61 HQYVKAGATFYVSFSVVRGGDGMAGFAVRNPAGEVVKPYQWQATADYTDQVSPGGYYSVC 120
                    ****:************************************************************
DbipPB2_loj-PF   61 HQYVQAGATFYVSFSVVRGGDGMAGFAVRNPAGEVVKPYQWQATADYTDQVSPGGYYSVC 120

Dmel_loj-PF     121 IDNQFSRFAGKLVNIYITVVKYDAWDKYAKEIEQLQLNMQNFTATVGTVERNINDMMGYQ 180
                    ************************************************:*************
DbipPB2_loj-PF  121 IDNQFSRFAGKLVNIYITVVKYDAWDKYAKEIEQLQLNMQNFTATIGTVERNINDMMGYQ 180

Dmel_loj-PF     181 AHSRHRESRDYALLLDNNAYIQTFSISQIVVILITCSVQVFFVRKLFEVKSSSKSRI 237
                    ***************************************************.*****
DbipPB2_loj-PF  181 AHSRHRESRDYALLLDNNAYIQTFSISQIVVILITCSVQVFFVRKLFEVKSNSKSRI 237
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Transcription Start Sites (TSS) Report Form (Optional)

Note: Complete this section if you have annotated the TSS for the gene above. This section is **optional** and you do not need to complete this section to submit the project.

Gene name (*e.g.*, *D. bipectinata eyeless*): ___*logjam*_____
Gene symbol (*e.g.*, *dbip_ey*): _____*log*_____

Name(s) of isoform(s) with unique TSS	List of isoforms with identical TSS
log-PD	log-PE and log-PF

Names of the isoforms with unique TSS in *D. melanogaster* that are absent in this species:

_____

Provide the evidence (text and figures) which support the hypothesis that these isoforms are absent in this species (*e.g.*, changes in canonical splice sites, gene structure, etc.):

Isoform TSS Report

Complete an Isoform TSS report (through page PAGEREF _Ref91436230 \h 17) for each unique TSS listed in the table above. Copy and paste this form to create as many copies as needed.

Gene-isoform name (*e.g.*, *dbip_ey-RA*): loj-PD

Names of the isoforms with the same TSS as this isoform:
loj-PE and loj-PF

Type of core promoter in *D. melanogaster* (see table below):
(Peaked / Intermediate / Broad / Insufficient Evidence)
peaked

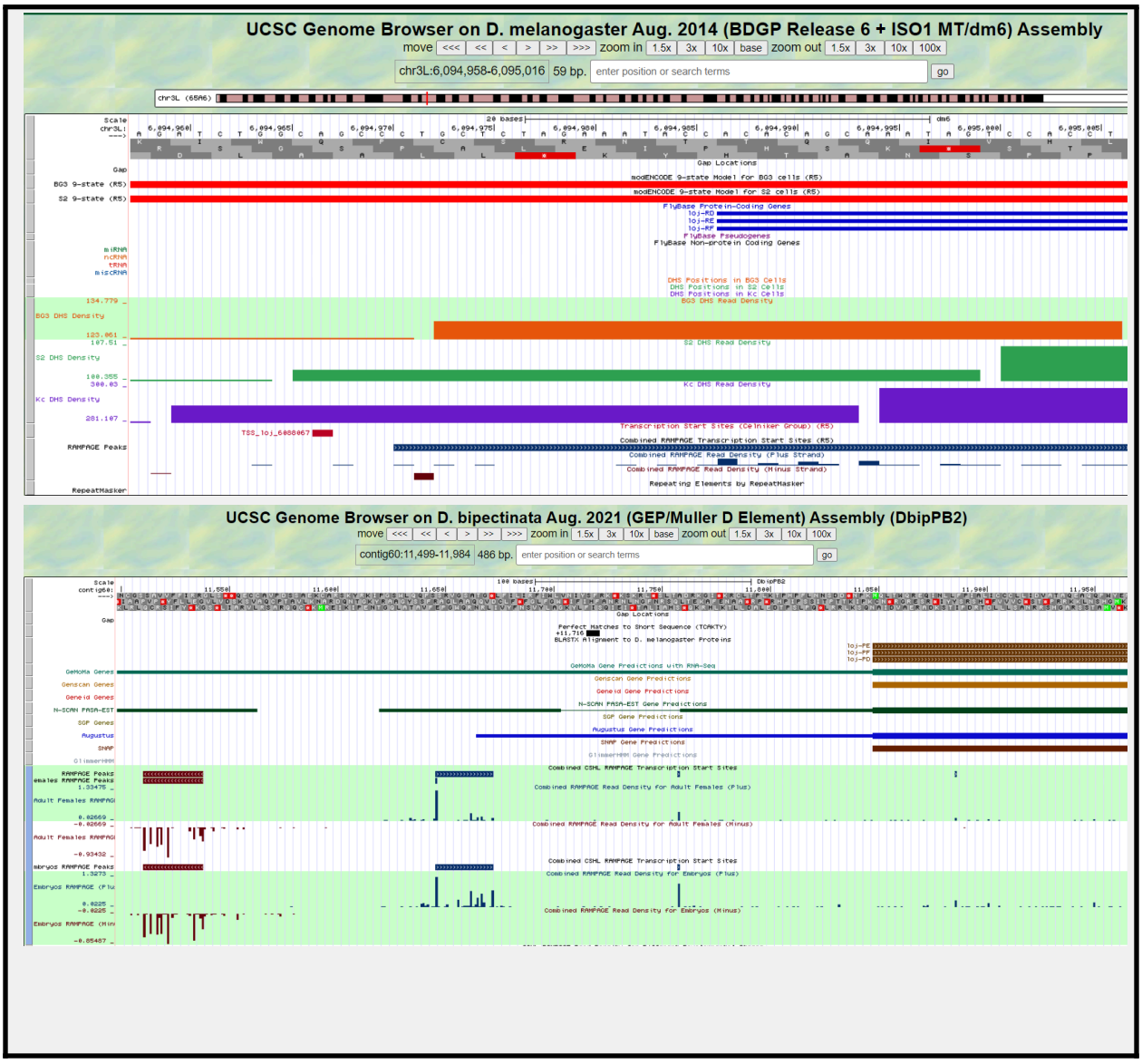
The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

1. Turn on RAMPAGE evidence tracks (Only applies to projects with these tracks)

Coordinates of the TSS position based on position with the highest RAMPAGE read density
D. melanogaster: 6,094,971- 6,095,063
D. bipectinata: 11, 646-11,672

Coordinates of the narrow TSS search region based on RAMPAGE peaks
D. bipectinate:11,646
D. melanogaster: 6,094,987

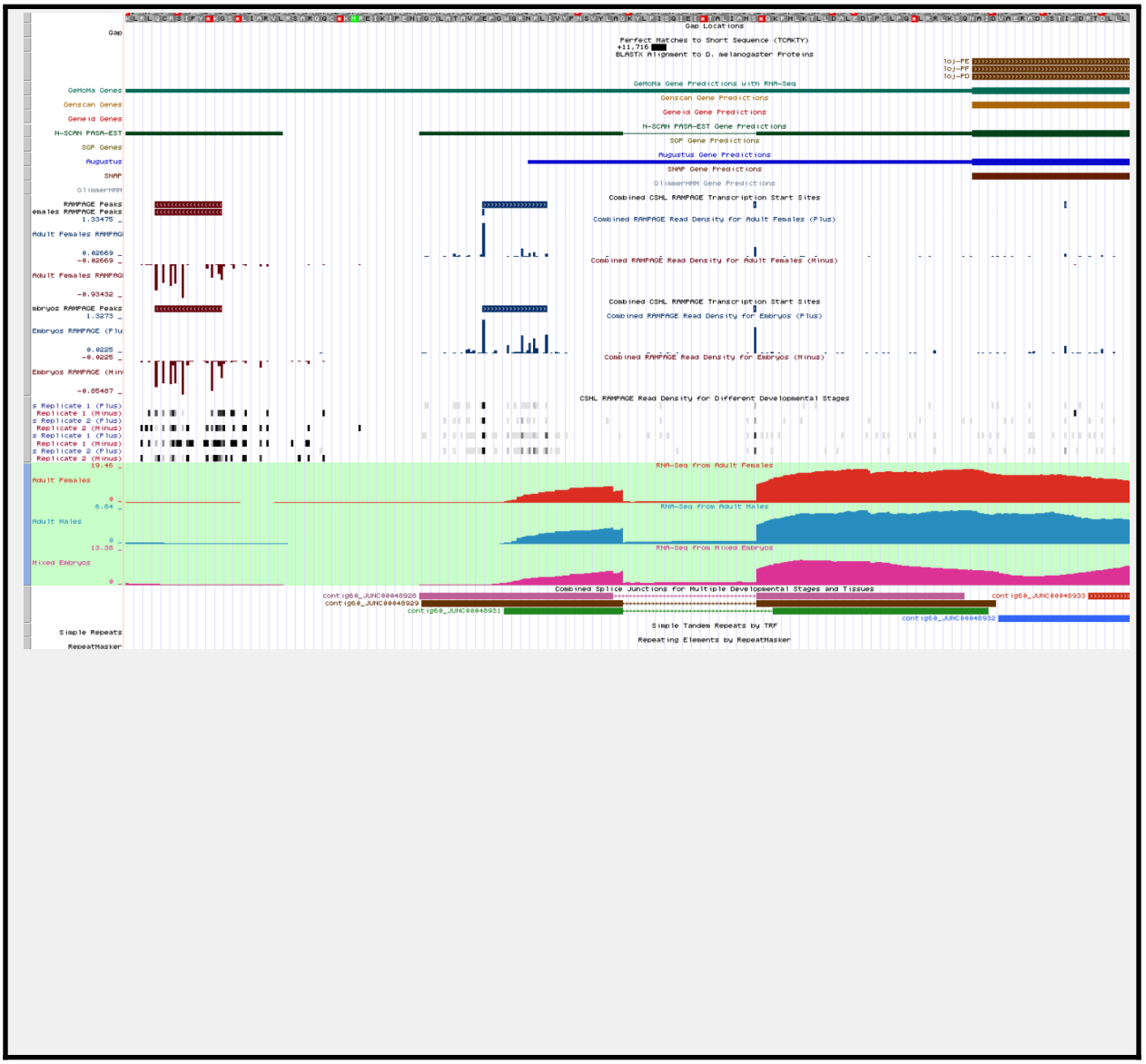
If the TSS position and narrow TSS search region are supported by RAMPAGE data, **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the Combined RAMPAGE TSS evidence track:**



2. Turn on RNA-Seq evidence tracks

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (*e.g.*, *regtools*), **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the following evidence tracks:**

3. RNA-Seq Coverage or RNA-Seq Alignment Summary
4. Combined Splice Junctions or RNA-Seq *TopHat*



If the RNA-Seq evidence tracks indicate the TSS position, list it here: _____N/A_____

If the RNA-Seq evidence tracks indicate a TSS search region, list it here:
11,655-11,703_____

3. Annotate the first transcribed exon

Coordinates of the first transcribed exon based on *BLASTN* alignment:

_____11662- 11994_____

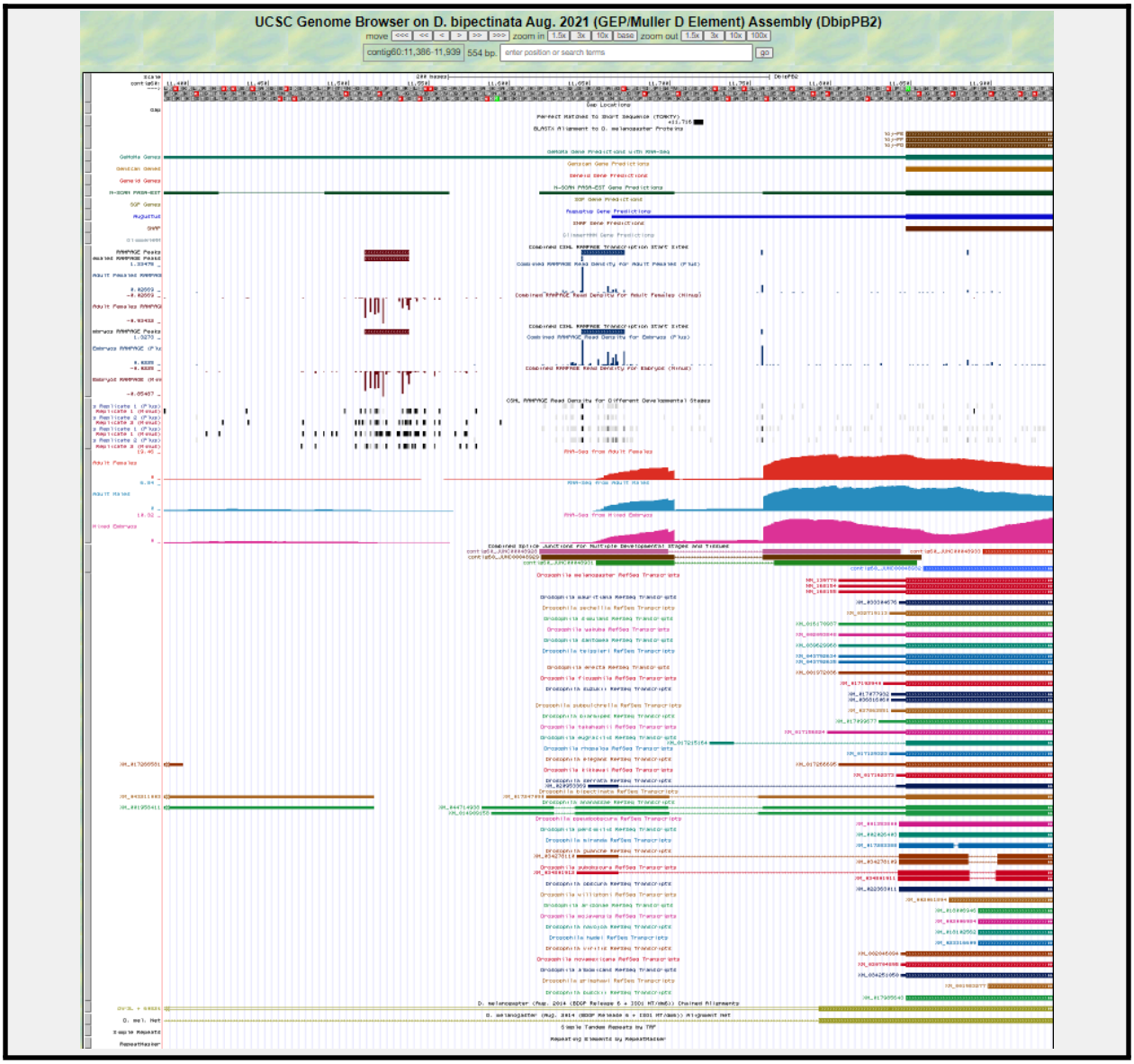
Does the *BLASTN* alignment cover the entire *D. melanogaster* first transcribed exon?
Yes

If not, specify the parts of the *D. melanogaster* exon that are missing from the *BLASTN* alignment.

If the TSS annotation is supported by *BLASTN* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *BLASTN* alignment into the box below:**

4. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, **paste a screenshot of the multiple sequence alignment (*e.g.*, from *Clustal Omega*, *ROAST*) into the box below:**



5. Summarize the evidence that supports the TSS annotation postulated above

Coordinate(s) of the TSS position(s):

Based on RAMPAGE data (if applicable): 11,646

Based on RNA-Seq data: 11,655-11,703

Based on *BLASTN* alignment: 11,662-11,672

Based on other evidence (please specify): 11,656-11,667 based on comparative genomics tracks

Note: If the *BLASTN* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter "Insufficient evidence" if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS position based on the available evidence? yes; 11,662
If so, indicate in the table below the evidence that supports this TSS position

If not, were you able to define a TSS search region? _____
If so, indicate in the table below the evidence that supports the TSS search region(s)

For each evidence type, enter an "X" in the cell to indicate whether the line of evidence supports, refutes, or neither supports nor refutes the TSS annotation:

Evidence type	Support	Refute	Neither
RAMPAGE peaks and read density	yes		
RNA-Seq coverage and splice junctions	yes		
BLASTN alignment of the initial exon from *D. melanogaster*	yes		
Sequence conservation with other *Drosophila* species (*e.g.*, "Conservation" track on the Genome Browser)	yes		
Other (please specify) [*e.g.*, *RefSeq Genes, N-SCAN PASA-EST, Augustus* TSS predictions; histone modifications (ChIP-Seq data)].			

Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against the annotated TSS; check "Refute." In contrast, the lack of RNA-Seq read coverage is a negative result that neither supports nor refutes the TSS annotation; check "Neither."

Gene Report Form

Gene name (*e.g.*, *D. bipectinata eyeless*): *D. bipectinata Ets at 65A*
Gene symbol (*e.g.*, *dbip_ey*): *dbip_Ets65A*

Approximate location in project (from 5' end to 3' end): 18,671
Number of isoforms in *D. melanogaster:* 5
Number of isoforms in this project: 5

Complete the following table, including all of the isoforms in this project:

Name(s) of unique isoform(s) based on coding sequence	**List of isoforms with identical coding sequences**
Ets65A-PA	Ets65A-PC
Ets65A-PE	
Ets65A-PD	
Ets65A-PB	

Names of the isoforms with unique coding sequences in *D. melanogaster* that are absent in this species: N/A

Provide the evidence (text and figures) which support the hypothesis that these isoforms are absent in this species (*e.g.*, changes in canonical splice sites, gene structure, etc.):

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 their coding sequence is identical to that of another isoform.

Consensus Sequence Errors Report Form

Complete this section if you have identified errors in the project consensus sequence that affect the annotation of the gene described above.

All of the coordinates reported in this section should be relative to the coordinates of the original project sequence.

Location(s) in the project sequence with consensus errors:
Ets65A-PB: exon 2 is missing from D. bipectinata, but present in D. melanogaster. Coordinates: 35133-35183

1. Evidence that supports the consensus errors postulated above

Note: Evidence that could be used to support the hypothesis of errors within the consensus sequence includes a CDS alignment with frame shifts or in-frame stop codons, and RNA-Seq reads with discrepant alignments compared to the project sequence.

2. Generate a VCF file which describes the changes to the consensus sequence

Use the *Sequencer Updater* to create a Variant Call Format (VCF) file that describes the changes to the consensus sequence you have identified above. **Paste a screenshot with the list of sequence changes into the box below:**

Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): dbip_Ets65A-PA

Names of any additional isoforms with identical coding sequences:
Dbip Ets65A-PC

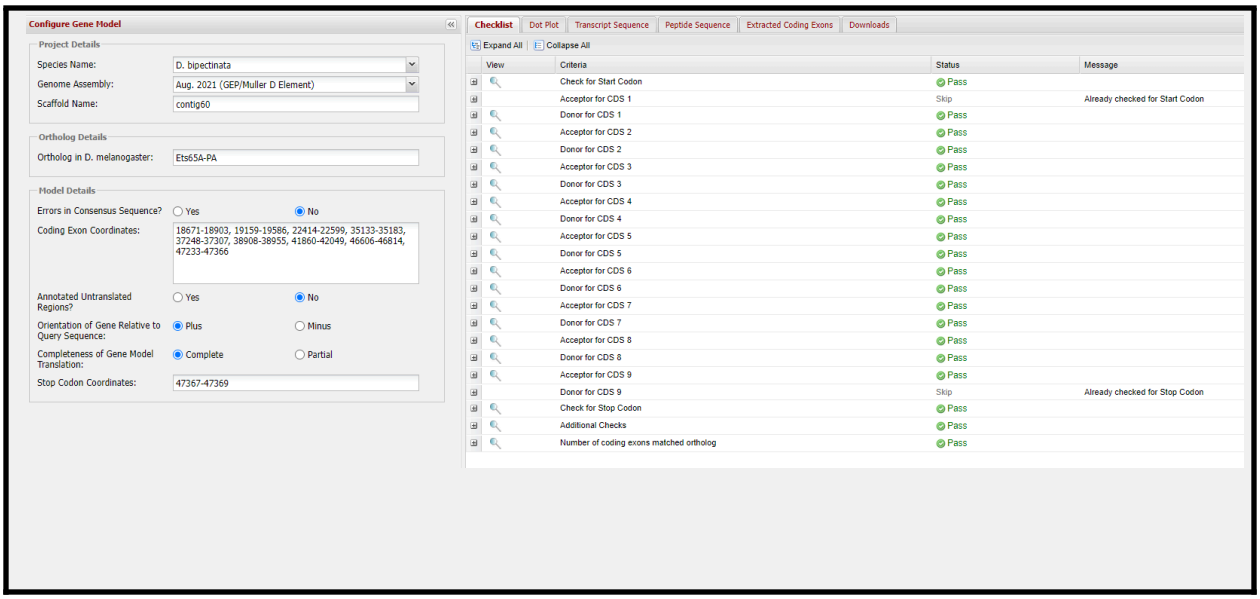
Is the 5' end of this isoform missing from the end of the project? __No_____
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? __No_____
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

> **Note:** For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.



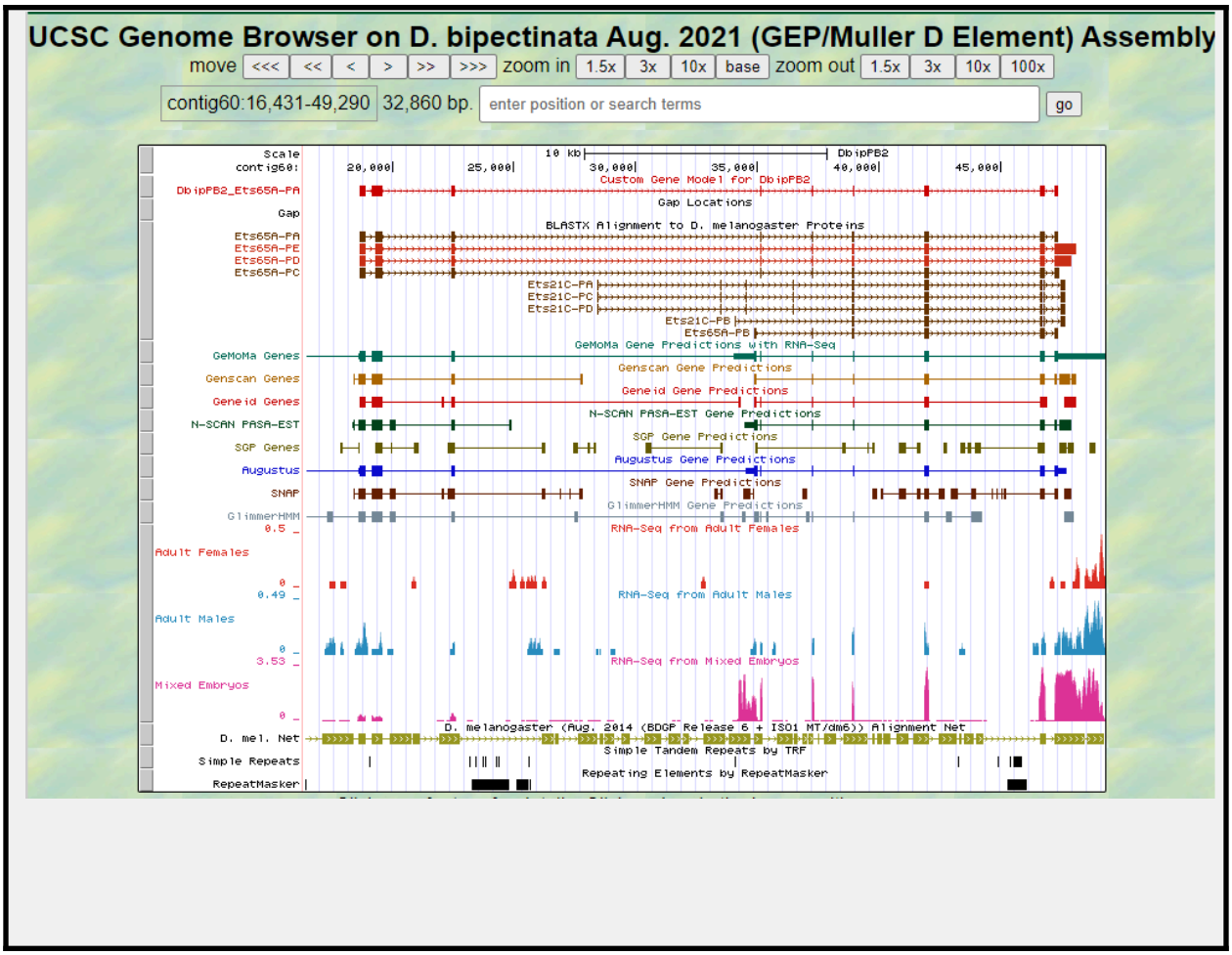
2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

9. A sequence alignment track (*e.g.*, D. mel Proteins)
10. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
11. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)

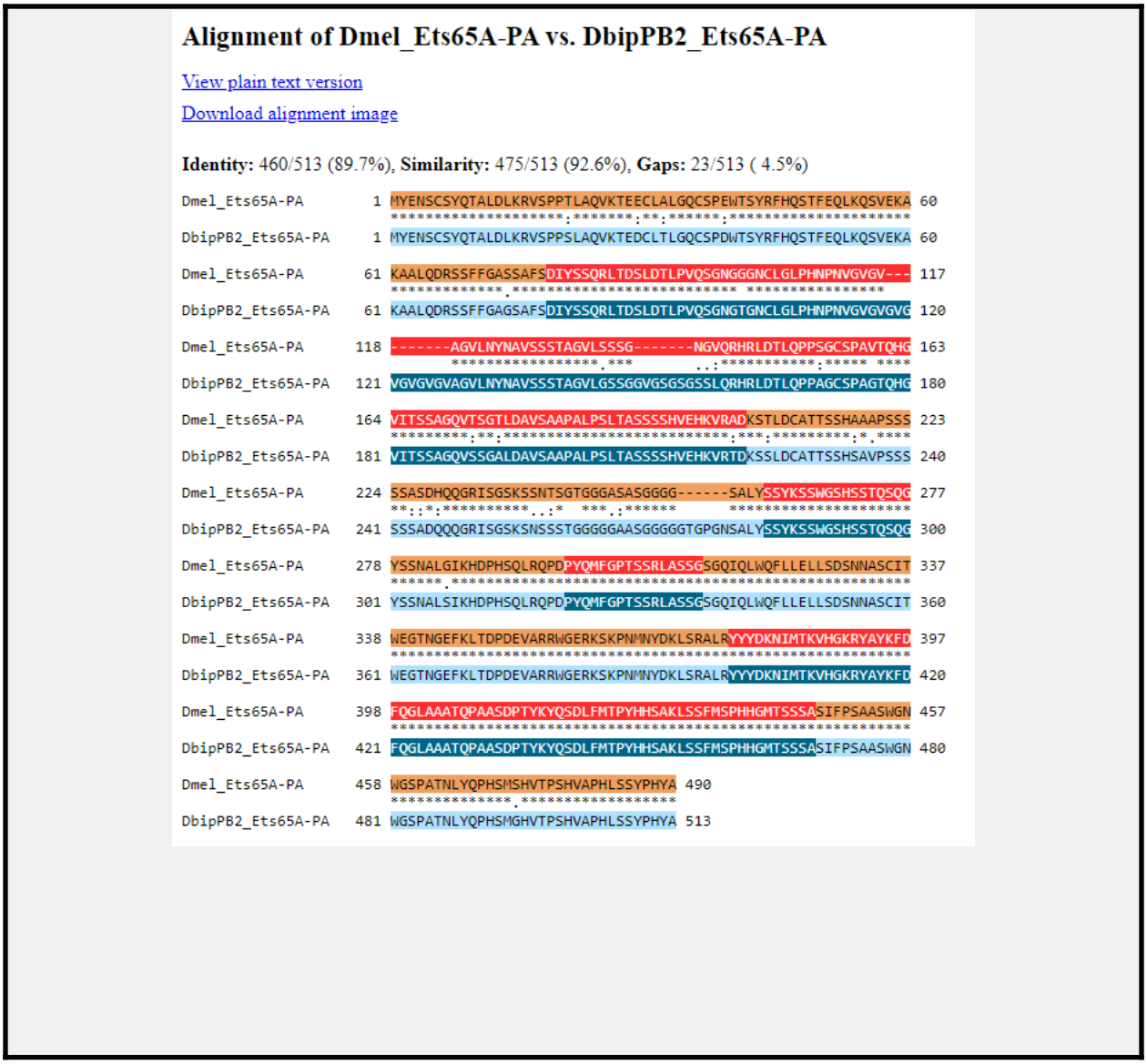
12. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_Ets65A-PA vs. DbipPB2_Ets65A-PA

View plain text version

Download alignment image

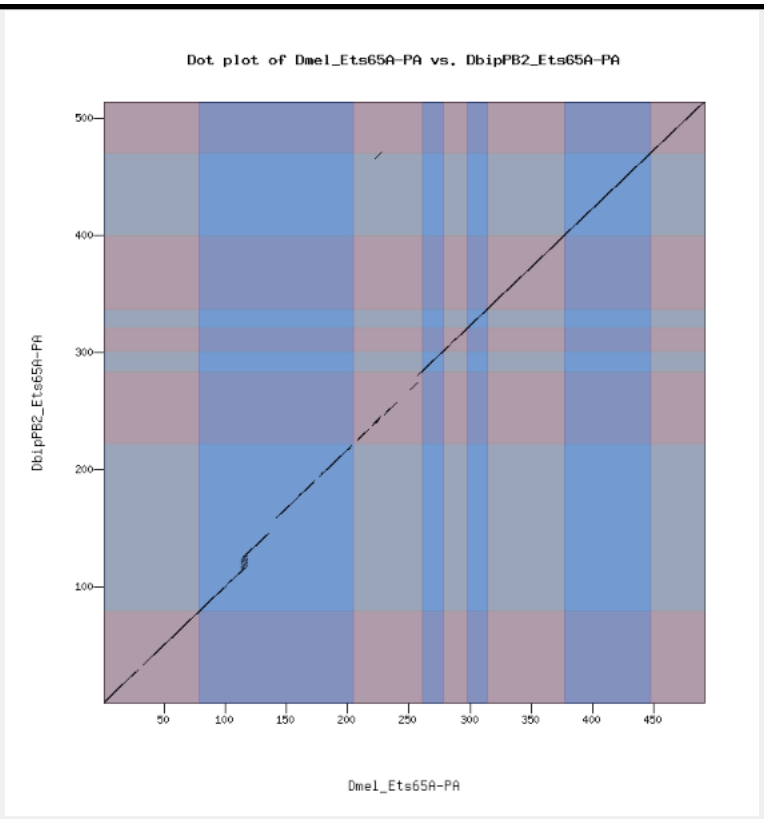
Identity: 460/513 (89.7%), **Similarity:** 475/513 (92.6%), **Gaps:** 23/513 (4.5%)

```
Dmel_Ets65A-PA       1 MYENSCSYQTALDLKRVSPPTLAQVKTEECLALGQCSPEWTSYRFHQSTFEQLKQSVEKA 60
                       ********************:*******:**:******:**********************
DbipPB2_Ets65A-PA    1 MYENSCSYQTALDLKRVSPPSLAQVKTEDCLTLGQCSPDWTSYRFHQSTFEQLKQSVEKA 60

Dmel_Ets65A-PA      61 KAALQDRSSFFGASSAFSDIYSSQRLTDSLDTLPVQSGNGGGNCLGLPHNPNVGVGV--- 117
                       ************.************************ ***************
DbipPB2_Ets65A-PA    61 KAALQDRSSFFGAGSAFSDIYSSQRLTDSLDTLPVQSGNGTGNCLGLPHNPNVGVGVGVG 120

Dmel_Ets65A-PA      118 -------AGVLNYNAVSSSTAGVLSSSG-------NGVQRHRLDTLQPPSGCSPAVTQHG 163
                            *****************.***        ..:***********:***** ****
DbipPB2_Ets65A-PA    121 VGVGVGVAGVLNYNAVSSSTAGVLGSSGGVGSGSGSSLQRHRLDTLQPPAGCSPAGTQHG 180

Dmel_Ets65A-PA      164 VITSSAGQVTSGTLDAVSAAPALPSLTASSSSHVEHKVRADKSTLDCATTSSHAAAPSSS 223
                       *********:**:**************************:***:*********:*.****
DbipPB2_Ets65A-PA    181 VITSSAGQVSSGALDAVSAAPALPSLTASSSSHVEHKVRTDKSSLDCATTSSHSAVPSSS 240

Dmel_Ets65A-PA      224 SSASDHQQGRISGSKSSNTSGTGGGASASGGGG------SALYSSYKSSWGSHSSTQSQG 277
                       **::*:**********..:*  ***.:******      ********************
DbipPB2_Ets65A-PA    241 SSSADQQQGRISGSKSNSSSTGGGGGAASGGGGGTGPGNSALYSSYKSSWGSHSSTQSQG 300

Dmel_Ets65A-PA      278 YSSNALGIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 337
                       ******.***********************************************************
DbipPB2_Ets65A-PA    301 YSSNALSIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 360

Dmel_Ets65A-PA      338 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 397
                       ************************************************************
DbipPB2_Ets65A-PA    361 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 420

Dmel_Ets65A-PA      398 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 457
                       ************************************************************
DbipPB2_Ets65A-PA    421 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 480

Dmel_Ets65A-PA      458 WGSPATNLYQPHSMSHVTPSHVAPHLSSYPHYA 490
                       *************.******************
DbipPB2_Ets65A-PA    481 WGSPATNLYQPHSMGHVTPSHVAPHLSSYPHYA 513
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): dbip Ets65A-PC

Names of any additional isoforms with identical coding sequences:
Dbip Ets65A-PA

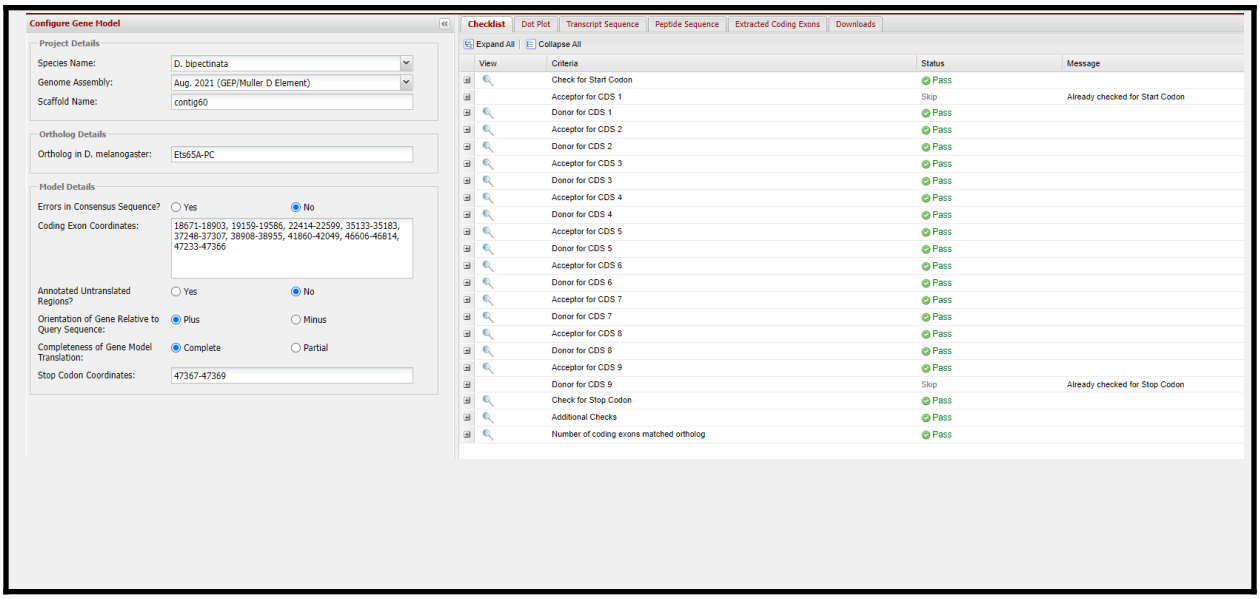
Is the 5' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

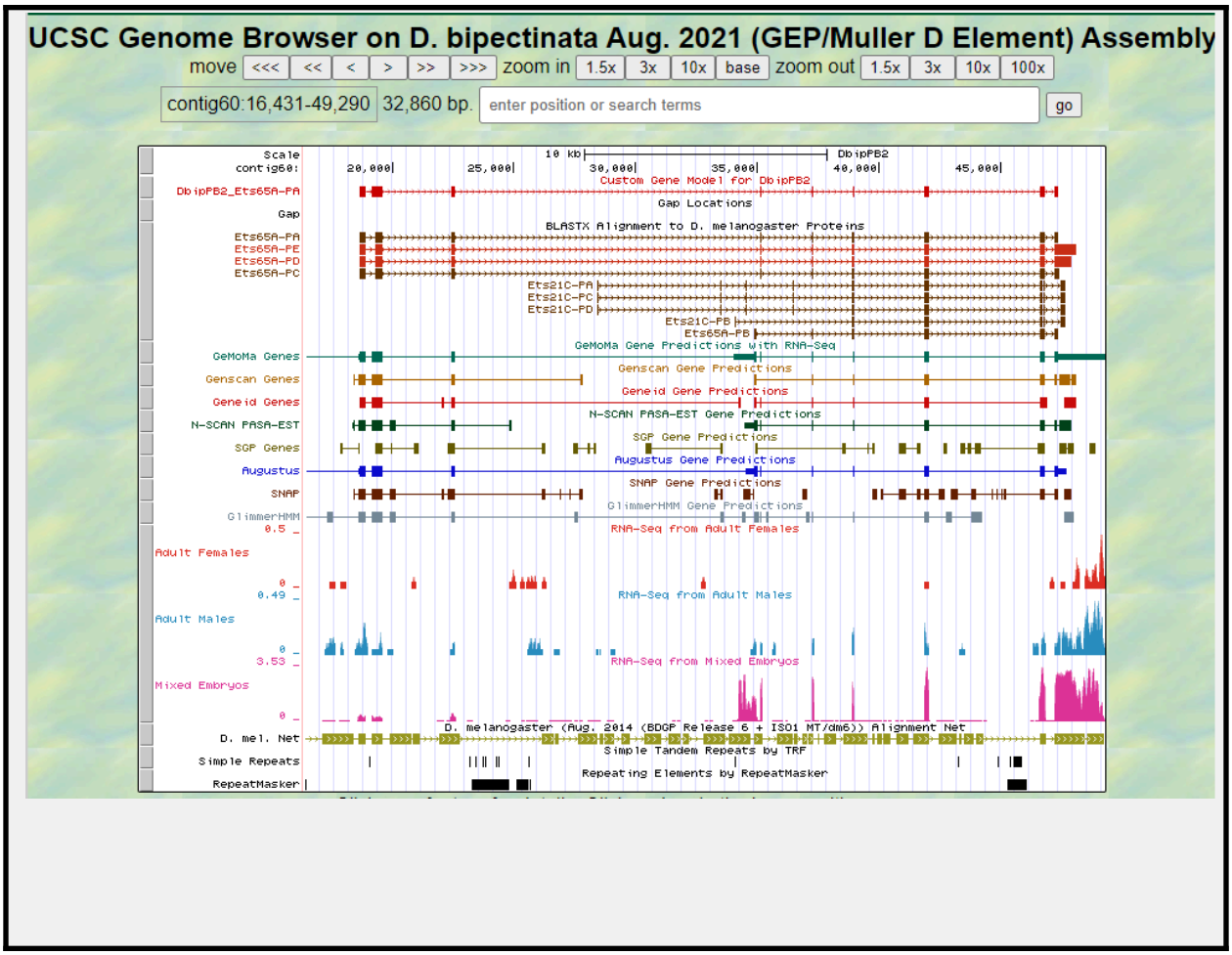


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box below:


UCSC Genome Browser on D. bipectinata Aug. 2021 (GEP/Muller D Element) Assembly
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x
contig60:16,431-49,290 32,860 bp. enter position or search terms go
Scale
contig60:
20,000
25,000
30,000
35,000
40,000
45,000
10 kb
DbipPB2
Custom Gene Model for DbipPB2
DbipPB2_Ets65A-PA
Gap Locations
Gap
BLASTX Alignment to D. melanogaster Proteins
Ets65A-PA
Ets65A-PE
Ets65A-PD
Ets65A-PC
Ets21C-PA
Ets21C-PC
Ets21C-PD
Ets21C-PB
Ets65A-PB
GeMoMa Gene Predictions with RNA-Seq
GeMoMa Genes
Genscan Gene Predictions
Genscan Genes
Geneid Gene Predictions
Geneid Genes
N-SCAN PASA-EST Gene Predictions
N-SCAN PASA-EST
SGP Gene Predictions
SGP Genes
Augustus Gene Predictions
Augustus
SNAP Gene Predictions
SNAP
GlimmerHMM Gene Predictions
GlimmerHMM
RNA-Seq from Adult Females
0.5
Adult Females
0
0.49
RNA-Seq from Adult Males
Adult Males
0
3.53
RNA-Seq from Mixed Embryos
Mixed Embryos
0
D. melanogaster (Aug. 2014 (BDGP Release 6 + ISO1 MT/dm6)) Alignment Net
D. mel. Net
Simple Tandem Repeats by TRF
Simple Repeats
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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**

Alignment of Dmel_Ets65A-PC vs. DbipPB2_Ets65A-PC

View plain text version

Download alignment image

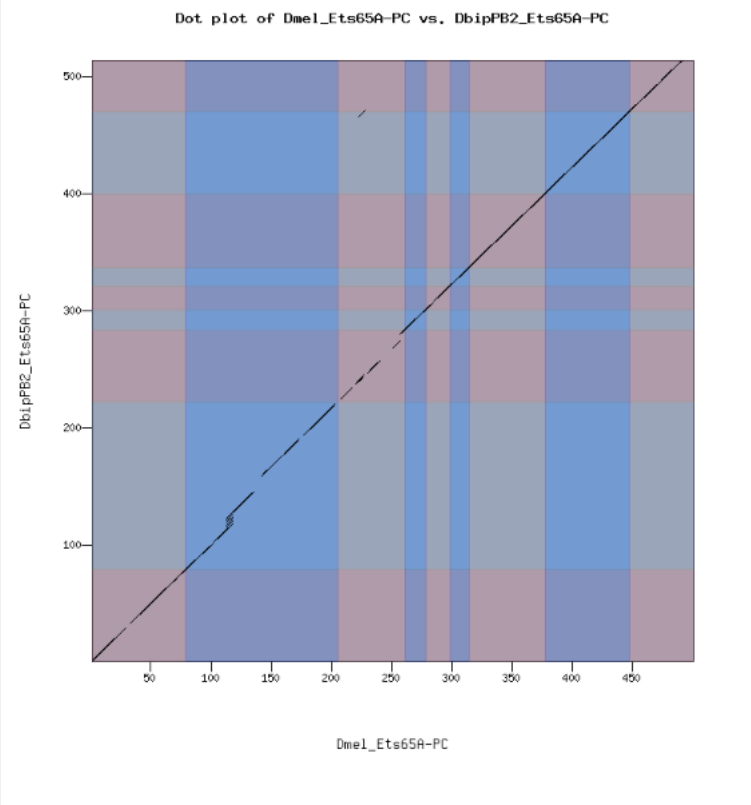
Identity: 460/522 (88.1%), **Similarity:** 475/522 (91.0%), **Gaps:** 32/522 (6.1%)

```
Dmel_Ets65A-PC       1 MYENSCSYQTALDLKRVSPPTLAQVKTEECLALGQCSPEWTSYRFHQSTFEQLKQSVEKA 60
                       ********************:*******:**:******:**********************
DbipPB2_Ets65A-PC    1 MYENSCSYQTALDLKRVSPPSLAQVKTEDCLTLGQCSPDWTSYRFHQSTFEQLKQSVEKA 60

Dmel_Ets65A-PC      61 KAALQDRSSFFGASSAFSDIYSSQRLTDSLDTLPVQSGNGGGNCLGLPHNPNVGVGV--- 117
                       *************.************************ ***************
DbipPB2_Ets65A-PC    61 KAALQDRSSFFGAGSAFSDIYSSQRLTDSLDTLPVQSGNGTGNCLGLPHNPNVGVGVGVG 120

Dmel_Ets65A-PC      118 -------AGVLNYNAVSSSTAGVLSSSG-------NGVQRHRLDTLQPPSGCSPAVTQHG 163
                            ****************.***        ..:***********:***** ****
DbipPB2_Ets65A-PC    121 VGVGVGVAGVLNYNAVSSSTAGVLGSSGGVGSGSGSSLQRHRLDTLQPPAGCSPAGTQHG 180

Dmel_Ets65A-PC      164 VITSSAGQVTSGTLDAVSAAPALPSLTASSSSHVEHKVRADKSTLDCATTSSHAAAPSSS 223
                       *********:**:*************************:***:*********:*.****
DbipPB2_Ets65A-PC    181 VITSSAGQVSSGALDAVSAAPALPSLTASSSSHVEHKVRTDKSSLDCATTSSHSAVPSSS 240

Dmel_Ets65A-PC      224 SSASDHQQGRISGSKSSNTSGTGGGASASGGGG------SALYSSYKSSWGSHSSTQSQG 277
                       **::*:**********..:*  ***.:******      ********************
DbipPB2_Ets65A-PC    241 SSSADQQQGRISGSKSNSSSTGGGGGAASGGGGGTGPGNSALYSSYKSSWGSHSSTQSQG 300

Dmel_Ets65A-PC      278 YSSNALGIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 337
                       ******.*************************************************
DbipPB2_Ets65A-PC    301 YSSNALSIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 360

Dmel_Ets65A-PC      338 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 397
                       ************************************************************
DbipPB2_Ets65A-PC    361 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 420

Dmel_Ets65A-PC      398 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 457
                       ************************************************************
DbipPB2_Ets65A-PC    421 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 480

Dmel_Ets65A-PC      458 WGSPATNLYQPHSMSHVTPSHVAPHLSSYPHYAXPSSSGGAF 499
                       *************.******************
DbipPB2_Ets65A-PC    481 WGSPATNLYQPHSMGHVTPSHVAPHLSSYPHYA--------- 513
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): dbip Ets65A-PD

Names of any additional isoforms with identical coding sequences:
N/A

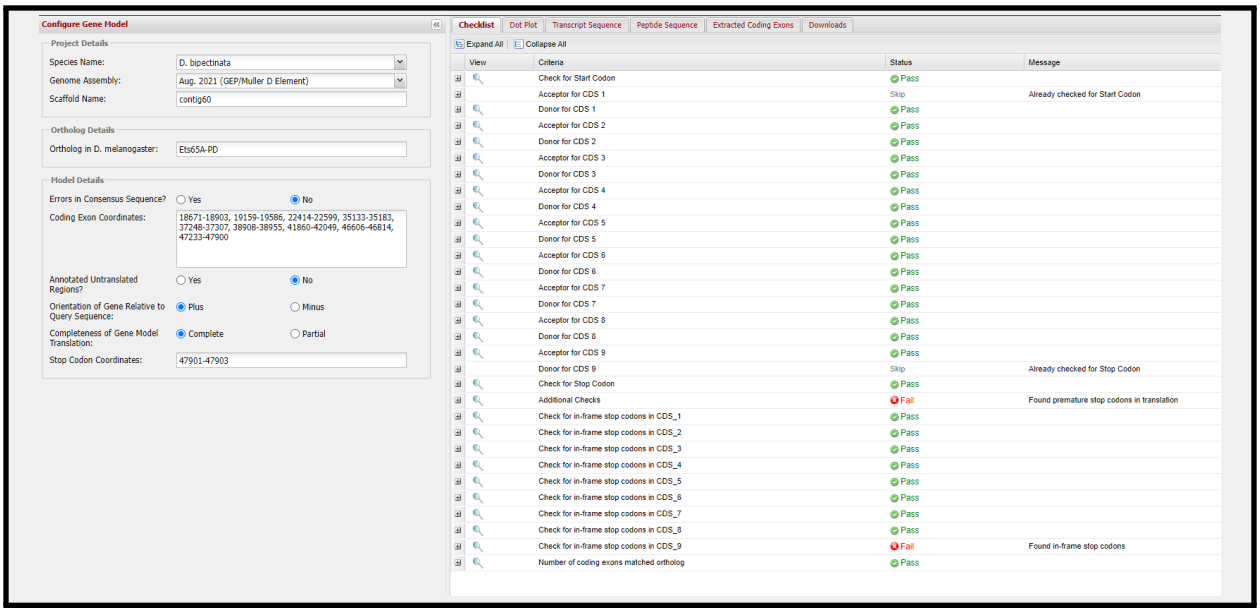
Is the 5' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

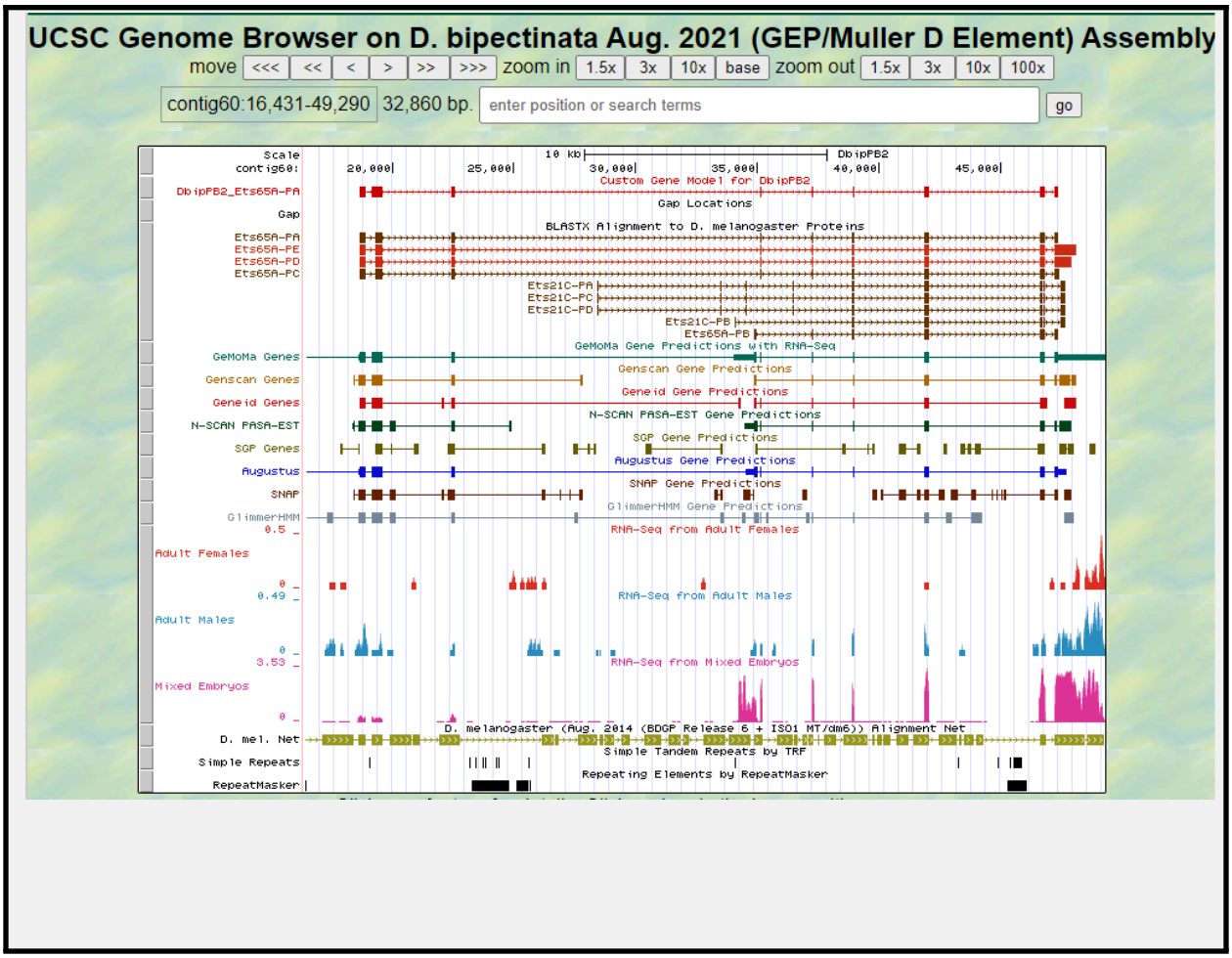


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

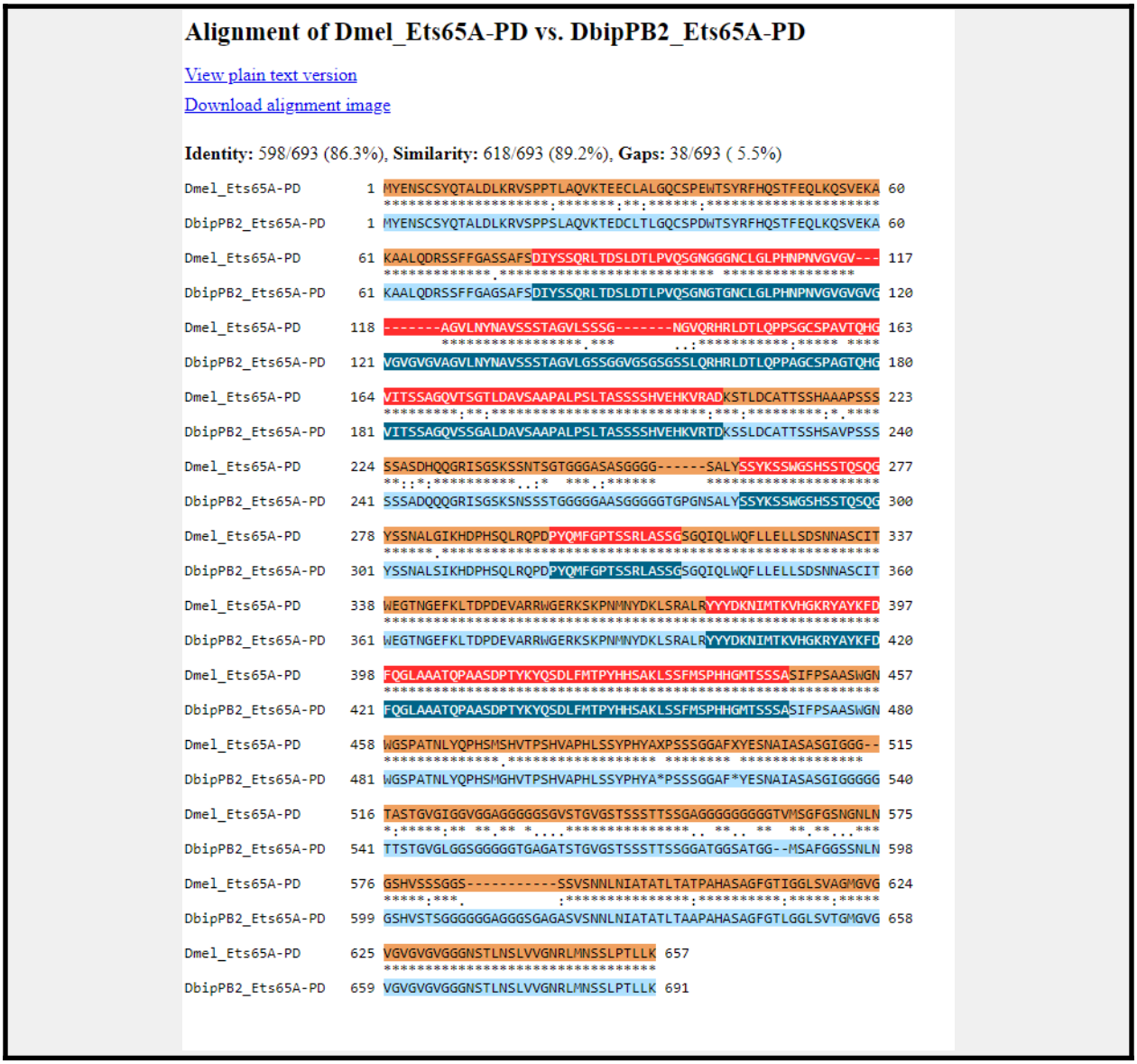
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_Ets65A-PD vs. DbipPB2_Ets65A-PD

View plain text version

Download alignment image

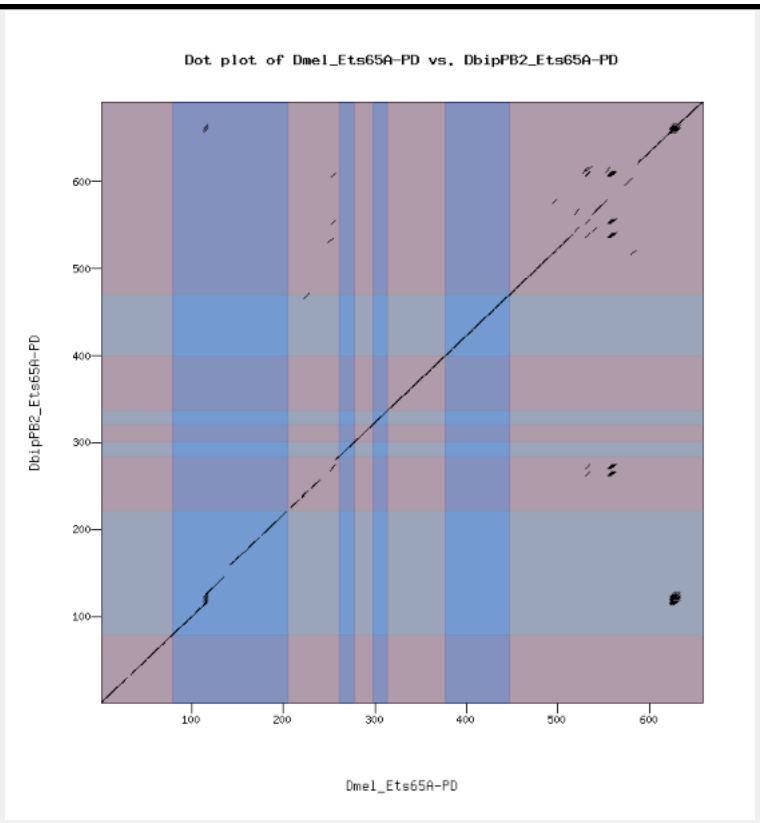
Identity: 598/693 (86.3%), **Similarity:** 618/693 (89.2%), **Gaps:** 38/693 (5.5%)

```
Dmel_Ets65A-PD       1 MYENSCSYQTALDLKRVSPPTLAQVKTEECLALGQCSPEWTSYRFHQSTFEQLKQSVEKA 60
                       ********************:*******:**:******:********************
DbipPB2_Ets65A-PD    1 MYENSCSYQTALDLKRVSPPSLAQVKTEDCLTLGQCSPDWTSYRFHQSTFEQLKQSVEKA 60

Dmel_Ets65A-PD      61 KAALQDRSSFFGASSAFSDIYSSQRLTDSLDTLPVQSGNGGGNCLGLPHNPNVGVGV--- 117
                       ************.********************** ***************
DbipPB2_Ets65A-PD    61 KAALQDRSSFFGAGSAFSDIYSSQRLTDSLDTLPVQSGNGTGNCLGLPHNPNVGVGVGVG 120

Dmel_Ets65A-PD      118 -------AGVLNYNAVSSSTAGVLSSSG-------NGVQRHRLDTLQPPSGCSPAVTQHG 163
                            ****************.***        ..:***********:***** ****
DbipPB2_Ets65A-PD    121 VGVGVGVAGVLNYNAVSSSTAGVLGSSGGVGSGSGSSLQRHRLDTLQPPAGCSPAGTQHG 180

Dmel_Ets65A-PD      164 VITSSAGQVTSGTLDAVSAAPALPSLTASSSSHVEHKVRADKSTLDCATTSSHAAAPSSS 223
                       *********:**:************************:***:*********:*.****
DbipPB2_Ets65A-PD    181 VITSSAGQVSSGALDAVSAAPALPSLTASSSSHVEHKVRTDKSSLDCATTSSHSAVPSSS 240

Dmel_Ets65A-PD      224 SSASDHQQGRISGSKSSNTSGTGGGASASGGGG------SALYSSYKSSWGSHSSTQSQG 277
                       **::*:**********..:*  ***.:******       ********************
DbipPB2_Ets65A-PD    241 SSSADQQQGRISGSKSNSSSTGGGGGAASGGGGGTGPGNSALYSSYKSSWGSHSSTQSQG 300

Dmel_Ets65A-PD      278 YSSNALGIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 337
                       ******.*************************************************
DbipPB2_Ets65A-PD    301 YSSNALSIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 360

Dmel_Ets65A-PD      338 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 397
                       ************************************************************
DbipPB2_Ets65A-PD    361 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 420

Dmel_Ets65A-PD      398 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 457
                       ************************************************************
DbipPB2_Ets65A-PD    421 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 480

Dmel_Ets65A-PD      458 WGSPATNLYQPHSMSHVTPSHVAPHLSSYPHYAXPSSSGGAFXYESNAIASASGIGGG-- 515
                       **************.***************** ******** ***************
DbipPB2_Ets65A-PD    481 WGSPATNLYQPHSMGHVTPSHVAPHLSSYPHYA*PSSSGGAF*YESNAIASASGIGGGGG 540

Dmel_Ets65A-PD      516 TASTGVGIGGVGGAGGGGGSGVSTGVGSTSSSTTSSGAGGGGGGGGGTVMSGFGSNGNLN 575
                       *:*****:** **.** *....***************.. **.. **  **.**...***
DbipPB2_Ets65A-PD    541 TTSTGVGLGGSGGGGGTGAGATSTGVGSTSSSTTSSGGATGGSATGG--MSAFGGSSNLN 598

Dmel_Ets65A-PD      576 GSHVSSSGGS-----------SSVSNNLNIATATLTATPAHASAGFGTIGGLSVAGMGVG 624
                       *****:***.           :***************:**********:******:*****
DbipPB2_Ets65A-PD    599 GSHVSTSGGGGGGAGGGSGAGASVSNNLNIATATLTAAPAHASAGFGTLGGLSVTGMGVG 658

Dmel_Ets65A-PD      625 VGVGVGVGGGNSTLNSLVVGNRLMNSSLPTLLK 657
                       ********************************
DbipPB2_Ets65A-PD    659 VGVGVGVGGGNSTLNSLVVGNRLMNSSLPTLLK 691
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): dbip Ets65A-PE

Names of any additional isoforms with identical coding sequences:
N/A

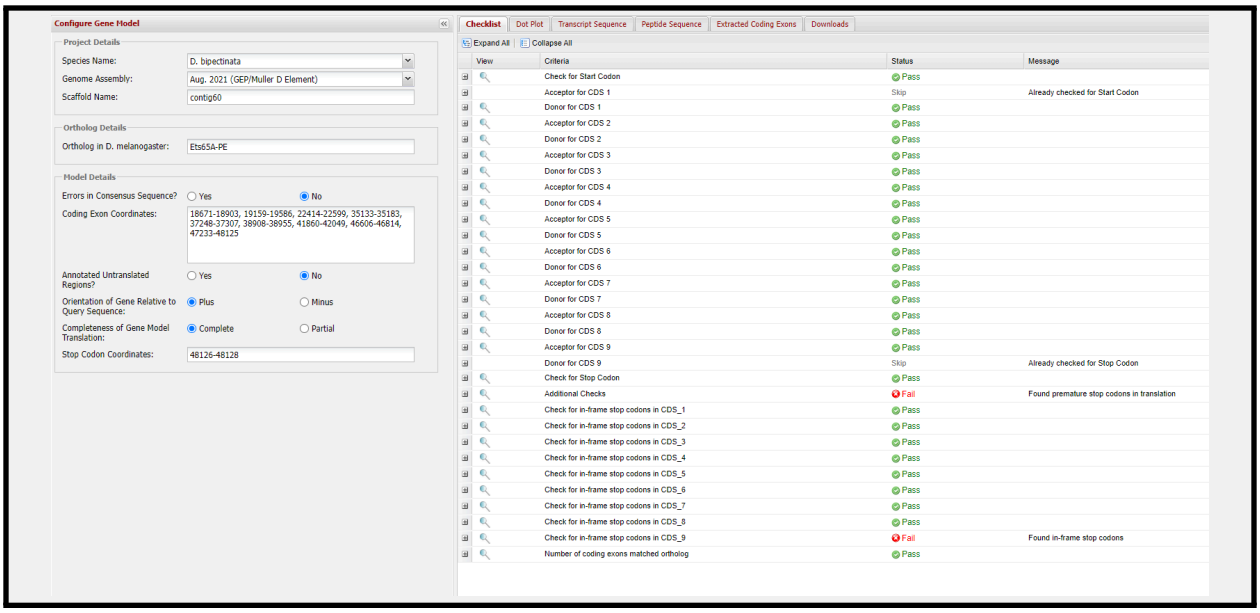
Is the 5' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

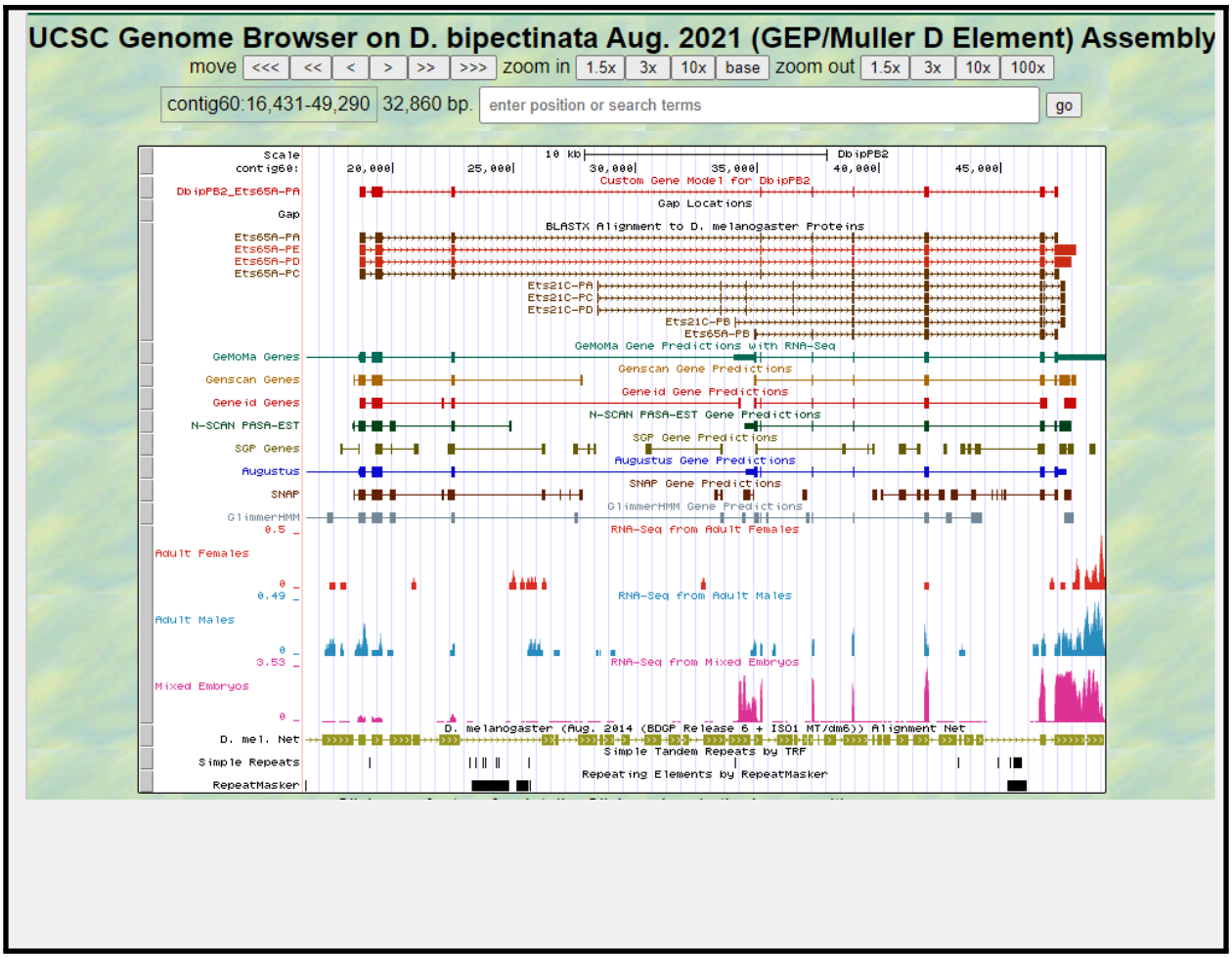


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

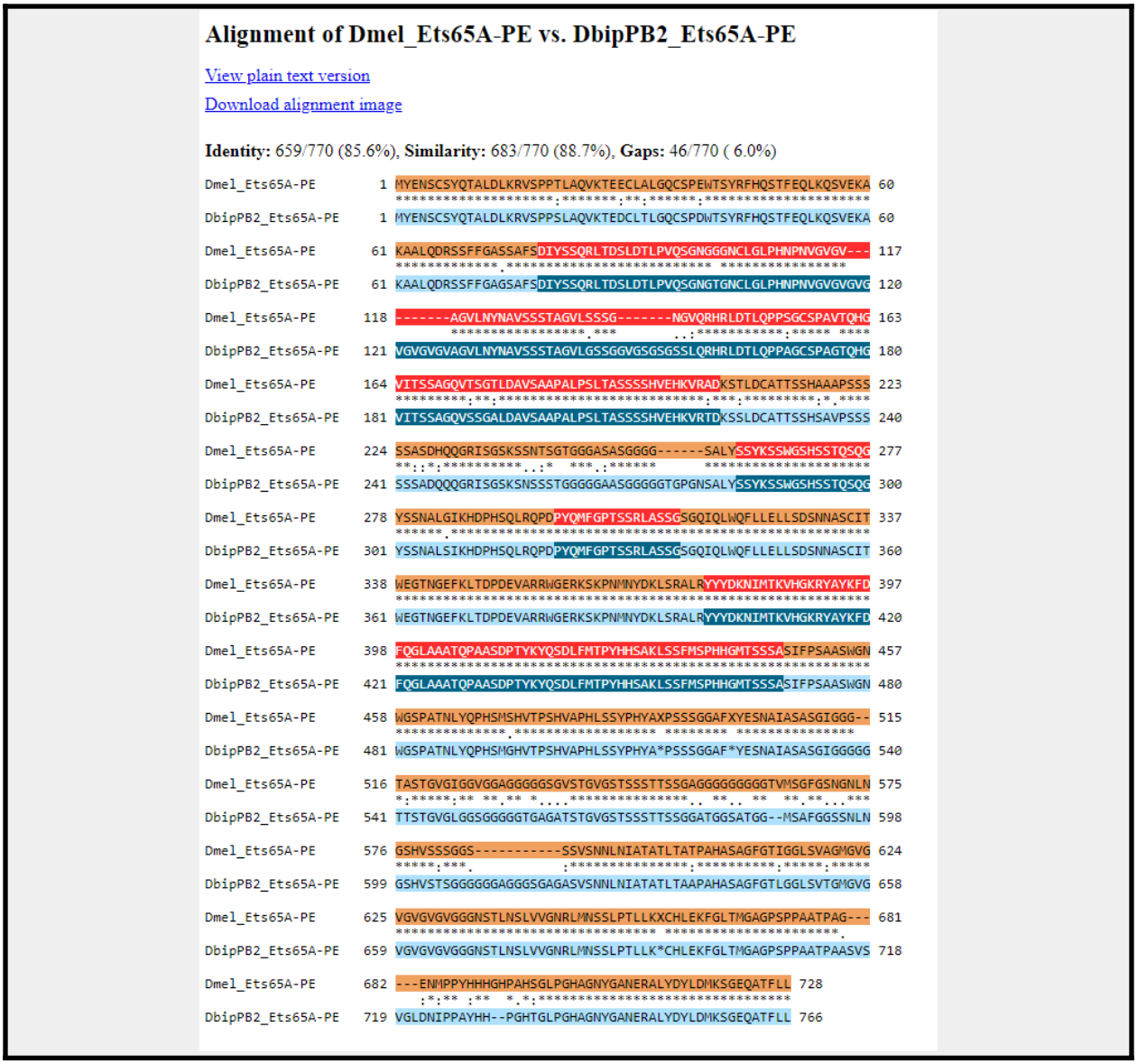
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_Ets65A-PE vs. DbipPB2_Ets65A-PE

View plain text version

Download alignment image

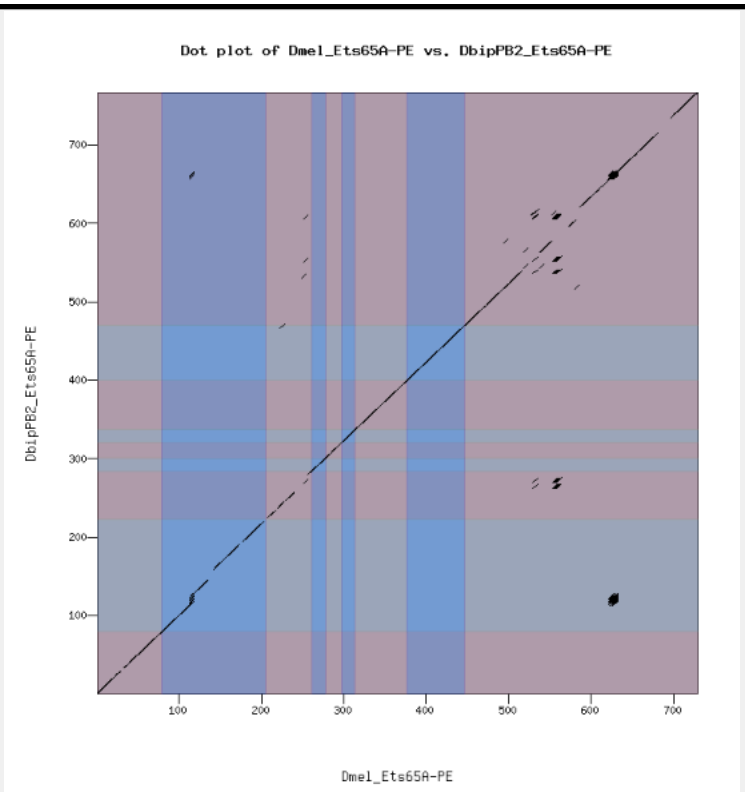
Identity: 659/770 (85.6%), **Similarity:** 683/770 (88.7%), **Gaps:** 46/770 (6.0%)

```
Dmel_Ets65A-PE        1 MYENSCSYQTALDLKRVSPPTLAQVKTEECLALGQCSPEWTSYRFHQSTFEQLKQSVEKA 60
                        ********************:********:**:******:**********************
DbipPB2_Ets65A-PE     1 MYENSCSYQTALDLKRVSPPSLAQVKTEDCLTLGQCSPDWTSYRFHQSTFEQLKQSVEKA 60

Dmel_Ets65A-PE       61 KAALQDRSSFFGASSAFSDIYSSQRLTDSLDTLPVQSGNGGGNCLGLPHNPNVGVGV--- 117
                        ************.************************ ***************
DbipPB2_Ets65A-PE     61 KAALQDRSSFFGAGSAFSDIYSSQRLTDSLDTLPVQSGNGTGNCLGLPHNPNVGVGVGVG 120

Dmel_Ets65A-PE       118 -------AGVLNYNAVSSSTAGVLSSSG-------NGVQRHRLDTLQPPSGCSPAVTQHG 163
                              *****************.***         ..:***********:***** ****
DbipPB2_Ets65A-PE     121 VGVGVGVAGVLNYNAVSSSTAGVLGSSGGVGSGSGSSLQRHRLDTLQPPAGCSPAGTQHG 180

Dmel_Ets65A-PE       164 VITSSAGQVTSGTLDAVSAAPALPSLTASSSSHVEHKVRADKSTLDCATTSSHAAAPSSS 223
                        *********:**:*************************:***:*********:*.****
DbipPB2_Ets65A-PE     181 VITSSAGQVSSGALDAVSAAPALPSLTASSSSHVEHKVRTDKSSLDCATTSSHSAVPSSS 240

Dmel_Ets65A-PE       224 SSASDHQQGRISGSKSSNTSGTGGGASASGGGG------SALYSSYKSSWGSHSSTQSQG 277
                        **::*:*********...*  ***.:******      ********************
DbipPB2_Ets65A-PE     241 SSSADQQQGRISGSKSNSSSTGGGGGAASGGGGGTGPGNSALYSSYKSSWGSHSSTQSQG 300

Dmel_Ets65A-PE       278 YSSNALGIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 337
                        ******.***************************************************
DbipPB2_Ets65A-PE     301 YSSNALSIKHDPHSQLRQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCIT 360

Dmel_Ets65A-PE       338 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 397
                        ************************************************************
DbipPB2_Ets65A-PE     361 WEGTNGEFKLTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFD 420

Dmel_Ets65A-PE       398 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 457
                        ************************************************************
DbipPB2_Ets65A-PE     421 FQGLAAATQPAASDPTYKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGN 480

Dmel_Ets65A-PE       458 WGSPATNLYQPHSMSHVTPSHVAPHLSSYPHYAXPSSSGGAFXYESNAIASASGIGGG-- 515
                        **************.***************** ******** **************
DbipPB2_Ets65A-PE     481 WGSPATNLYQPHSMGHVTPSHVAPHLSSYPHYA*PSSSGGAF*YESNAIASASGIGGGGG 540

Dmel_Ets65A-PE       516 TASTGVGIGGVGGAGGGGGSGVSTGVGSTSSSTTSSGAGGGGGGGGGTVMSGFGSNGNLN 575
                        *:*****:** **.** *....***************.. **.. **  **.**...***
DbipPB2_Ets65A-PE     541 TTSTGVGLGGSGGGGGTGAGATSTGVGSTSSSTTSSGGATGGSATGG--MSAFGGSSNLN 598

Dmel_Ets65A-PE       576 GSHVSSSGGS----------SSVSNNLNIATATLTATPAHASAGFGTIGGLSVAGMGVG 624
                        *****.***.           .*************.**********.*****.*****
DbipPB2_Ets65A-PE     599 GSHVSTSGGGGGGAGGGSGAGASVSNNLNIATATLTAAPAHASAGFGTLGGLSVTGMGVG 658

Dmel_Ets65A-PE       625 VGVGVGVGGGNSTLNSLVVGNRLMNSSLPTLLKXCHLEKFGLTMGAGPSPPAATPAG--- 681
                        ********************************* ********************.
DbipPB2_Ets65A-PE     659 VGVGVGVGGGNSTLNSLVVGNRLMNSSLPTLLK*CHLEKFGLTMGAGPSPPAATPAASVS 718

Dmel_Ets65A-PE       682 ---ENMPPYHHHGHPAHSGLPGHAGNYGANERALYDYLDMKSGEQATFLL 728
                           :*:** :**  *.*:*********************************
DbipPB2_Ets65A-PE     719 VGLDNIPPAYHH--PGHTGLPGHAGNYGANERALYDYLDMKSGEQATFLL 766
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Isoform Report Form

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

Gene-isoform symbol (*e.g.*, dbip_ey-PA): dbip Ets65A-PB

Names of any additional isoforms with identical coding sequences:
N/A

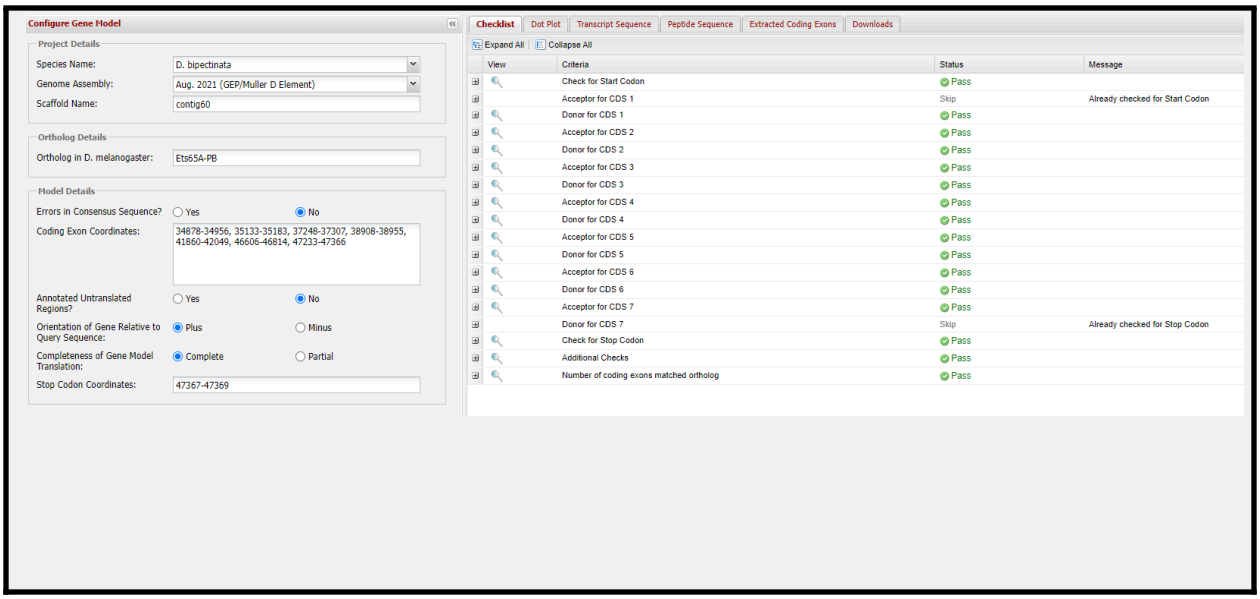
Is the 5' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 5' end: _____
Is the 3' end of this isoform missing from the end of the project? No
If so, how many putative exons are missing from the 3' end: _____

(Define "putative exons" based on the exons present in the *D. melanogaster* ortholog)

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 into the box below:**

Note: For projects with consensus sequence errors, report the exon coordinates relative to the **original project sequence**. Include the VCF file you have generated above when you submit the gene model to the *Gene Model Checker*. The *Gene Model Checker* will use this VCF file to automatically revise the submitted exon coordinates.

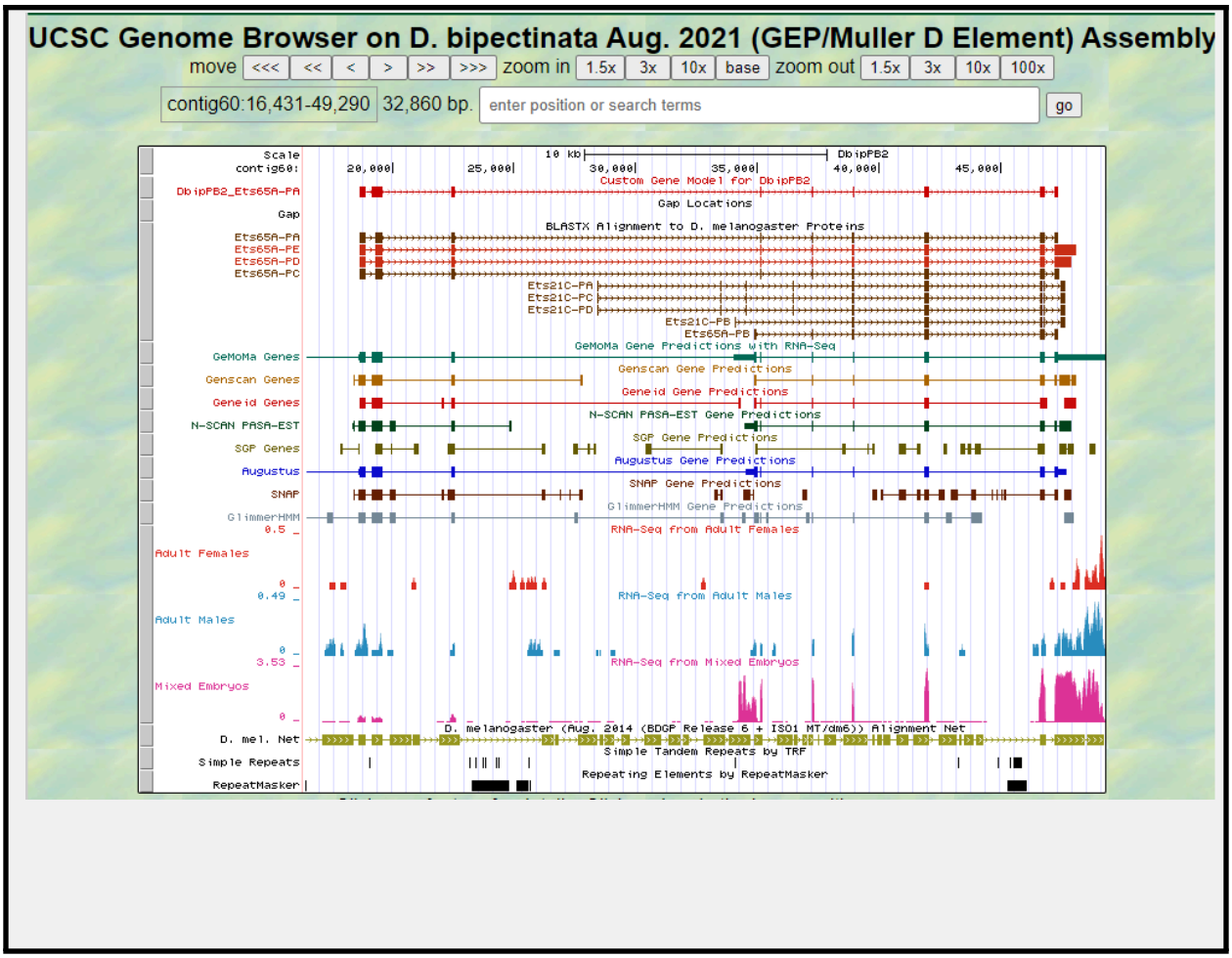


2. View the gene model on the Genome Browser

Click on the magnifying glass icon under the "Checklist" tab of the *Gene Model Checker* to view your gene model on the *GEP UCSC Genome Browser*. Zoom in so that **only this isoform is in the genome browser window, and capture a screenshot that includes the following evidence tracks if they are available:**

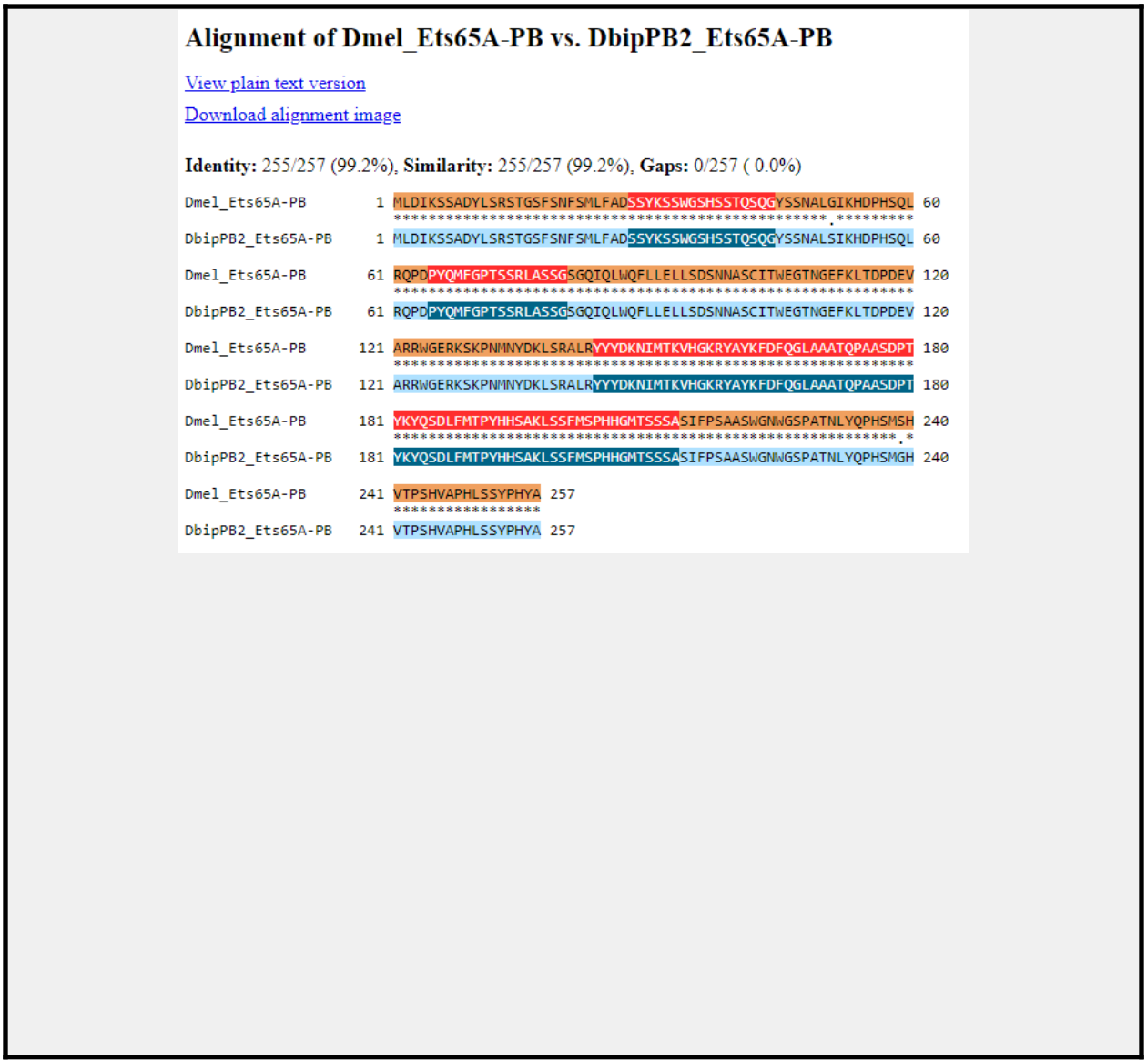
1. A sequence alignment track (*e.g.*, D. mel Proteins)
2. At least one gene prediction track (*e.g.*, *Genscan*, *GeMoMa*)
3. At least one RNA-Seq track (*e.g.*, RNA-Seq Coverage)
4. A comparative genomics track (*e.g.*, D. mel. Net Alignment, Conservation)

Paste a screenshot of your gene model as shown on the *GEP UCSC Genome Browser* into the box 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 either use the protein alignment generated by the *Gene Model Checker* (available through the "**View protein alignment**" link under the "Dot Plot" tab) or you can generate a new alignment using the "Align two or more sequences" feature at the NCBI BLAST website. **Paste a screenshot of the protein alignment into the box below:**



Alignment of Dmel_Ets65A-PB vs. DbipPB2_Ets65A-PB

View plain text version

Download alignment image

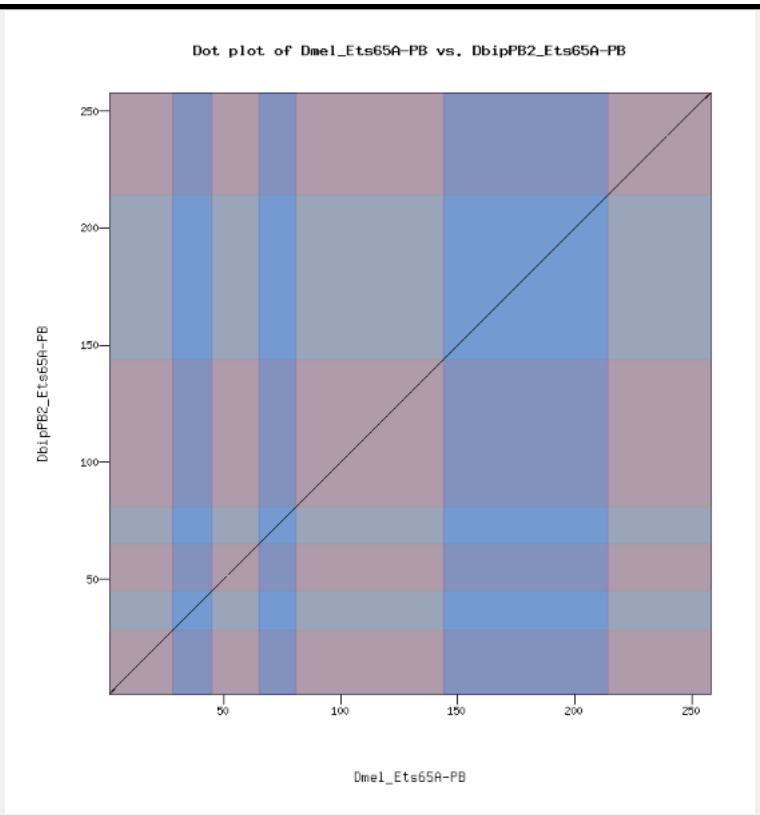
Identity: 255/257 (99.2%), **Similarity:** 255/257 (99.2%), **Gaps:** 0/257 (0.0%)

```
Dmel_Ets65A-PB        1 MLDIKSSADYLSRSTGSFSNFSMLFADSSYKSSWGSHSSTQSQGYSSNALGIKHDPHSQL 60
                        ************************************************.*********
DbipPB2_Ets65A-PB     1 MLDIKSSADYLSRSTGSFSNFSMLFADSSYKSSWGSHSSTQSQGYSSNALSIKHDPHSQL 60

Dmel_Ets65A-PB       61 RQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCITWEGTNGEFKLTDPDEV 120
                        ************************************************************
DbipPB2_Ets65A-PB     61 RQPDPYQMFGPTSSRLASSGSGQIQLWQFLLELLSDSNNASCITWEGTNGEFKLTDPDEV 120

Dmel_Ets65A-PB       121 ARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFQGLAAATQPAASDPT 180
                        ************************************************************
DbipPB2_Ets65A-PB     121 ARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFQGLAAATQPAASDPT 180

Dmel_Ets65A-PB       181 YKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGNWGSPATNLYQPHSMSH 240
                        ************************************************************.*
DbipPB2_Ets65A-PB     181 YKYQSDLFMTPYHHSAKLSSFMSPHHGMTSSSASIFPSAASWGNWGSPATNLYQPHSMGH 240

Dmel_Ets65A-PB       241 VTPSHVAPHLSSYPHYA 257
                        ****************
DbipPB2_Ets65A-PB     241 VTPSHVAPHLSSYPHYA 257
```

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

Paste a screenshot of the dot plot (generated by the *Gene Model Checker*) of your submitted model against the putative *D. melanogaster* ortholog into the box below. Provide an explanation for any anomalies on the dot plot (*e.g.*, large gaps, which would appear as kinks in the diagonal line; regions with no sequence similarity; indications of significant insertions or deletions).

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



Transcription Start Sites (TSS) Report Form 2 (Optional)

Note: Complete this section if you have annotated the TSS for the gene above. This section is **optional** and you do not need to complete this section to submit the project.

Gene name (*e.g.*, *D. bipectinata eyeless*): *D. bipectinate Ets at 65A*
Gene symbol (*e.g.*, *dbip_ey*): *dbip Ets65A*

Name(s) of isoform(s) with unique TSS	List of isoforms with identical TSS
Ets65A-PA	Ets65A-PC, Ets65A-PD, Ets65A-PE
Ets65A-PB	

Names of the isoforms with unique TSS in *D. melanogaster* that are absent in this species:

Provide the evidence (text and figures) which support the hypothesis that these isoforms are absent in this species (*e.g.*, changes in canonical splice sites, gene structure, etc.):

Isoform TSS Report

Complete an Isoform TSS report (through page PAGEREF _Ref91436230 \h 17) for each unique TSS listed in the table above. Copy and paste this form to create as many copies as needed.

Gene-isoform name (*e.g.*, *dbip_ey-RA*): *dbip_Ets65A-PA*

Names of the isoforms with the same TSS as this isoform:
dbip_Ets65A-PC, *dbip_Ets65A-PD*, *dbip_Ets65A-PE*

Type of core promoter in *D. melanogaster* (see table below):
(Peaked / Intermediate / Broad / Insufficient Evidence)
peaked

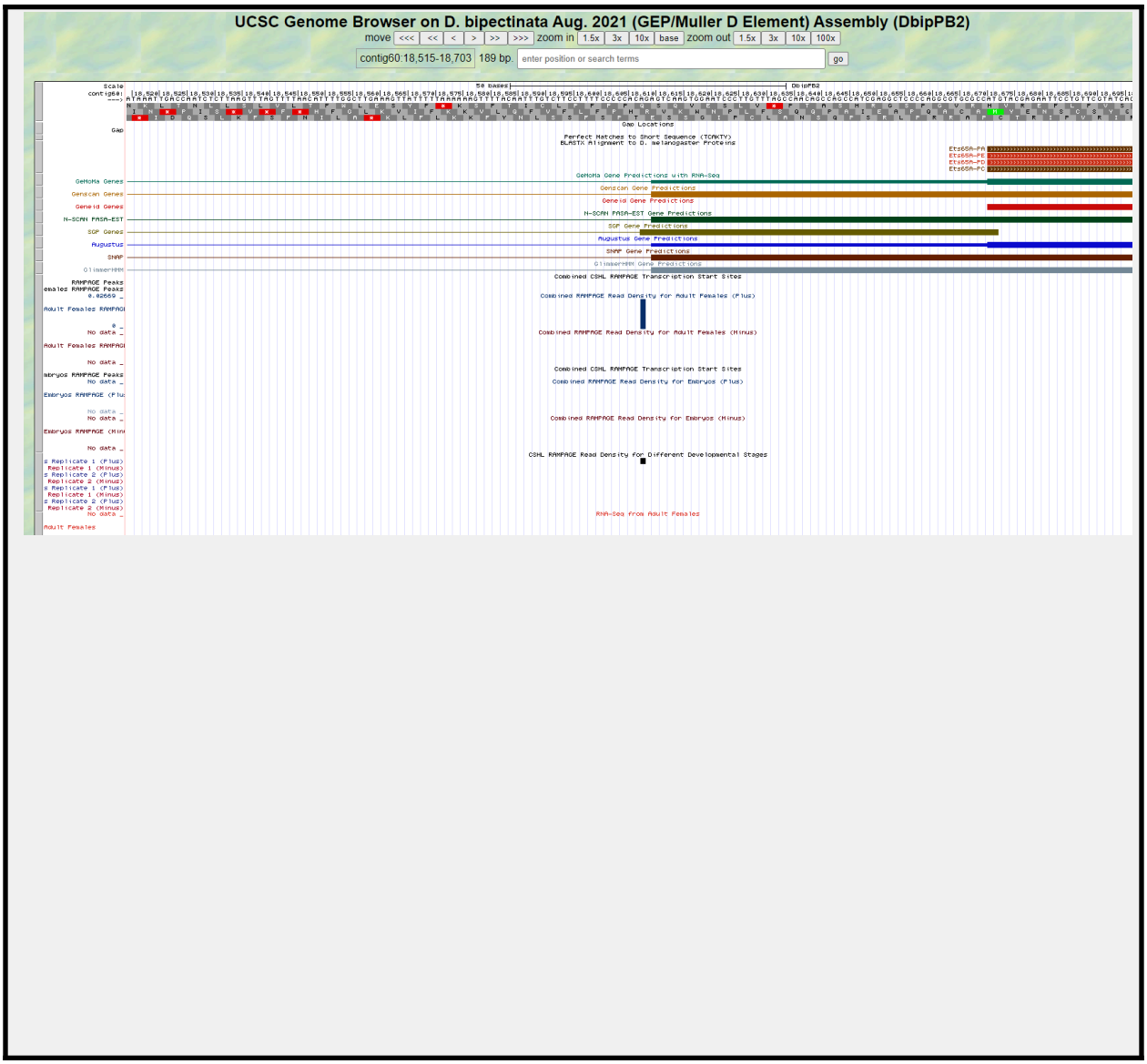
The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

1. Turn on RAMPAGE evidence tracks (Only applies to projects with these tracks)

Coordinates of the TSS position based on position with the highest RAMPAGE read density
18,608

Coordinates of the narrow TSS search region based on RAMPAGE peaks
18,608

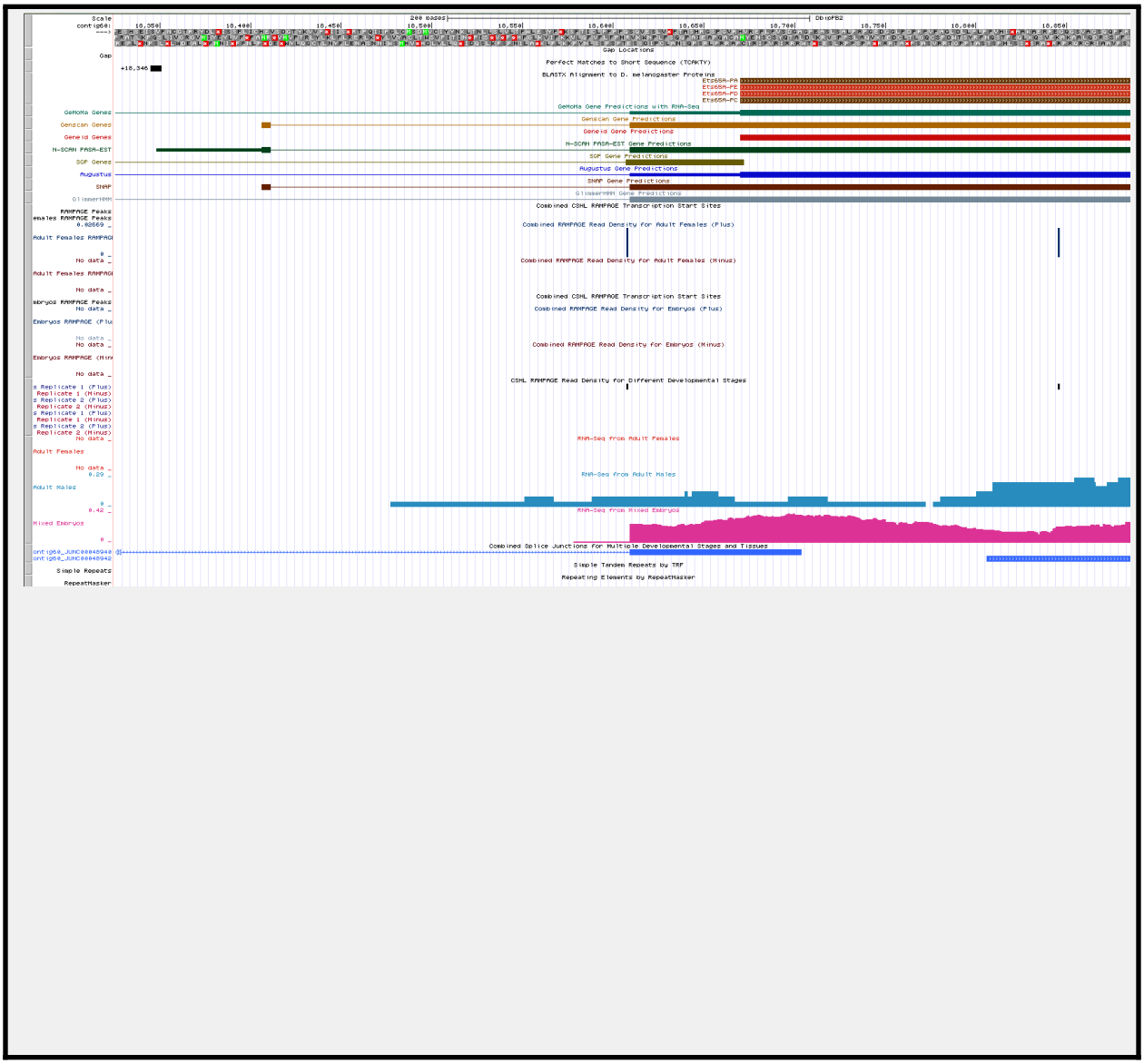
If the TSS position and narrow TSS search region are supported by RAMPAGE data, **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the Combined RAMPAGE TSS evidence track:**



2. Turn on RNA-Seq evidence tracks

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (*e.g.*, *regtools*), **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the following evidence tracks:**

5. RNA-Seq Coverage or RNA-Seq Alignment Summary
6. Combined Splice Junctions or RNA-Seq *TopHat*



If the RNA-Seq evidence tracks indicate the TSS position, list it here:

_____

If the RNA-Seq evidence tracks indicate a TSS search region, list it here:
18,600-18,670_____

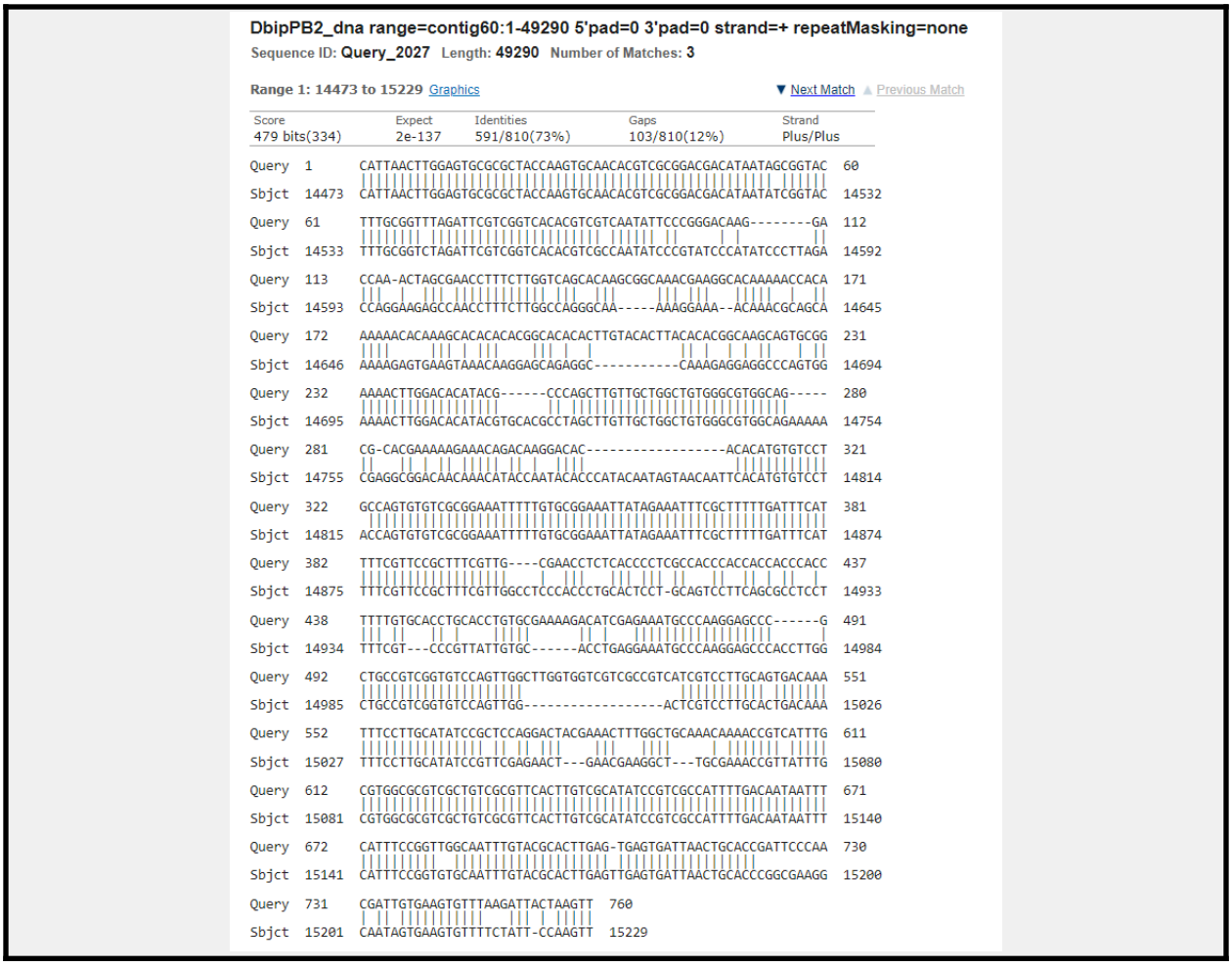
3. Annotate the first transcribed exon

Coordinates of the first transcribed exon based on *BLASTN* alignment:
14473 to 15229

Does the *BLASTN* alignment cover the entire *D. melanogaster* first transcribed exon?
No

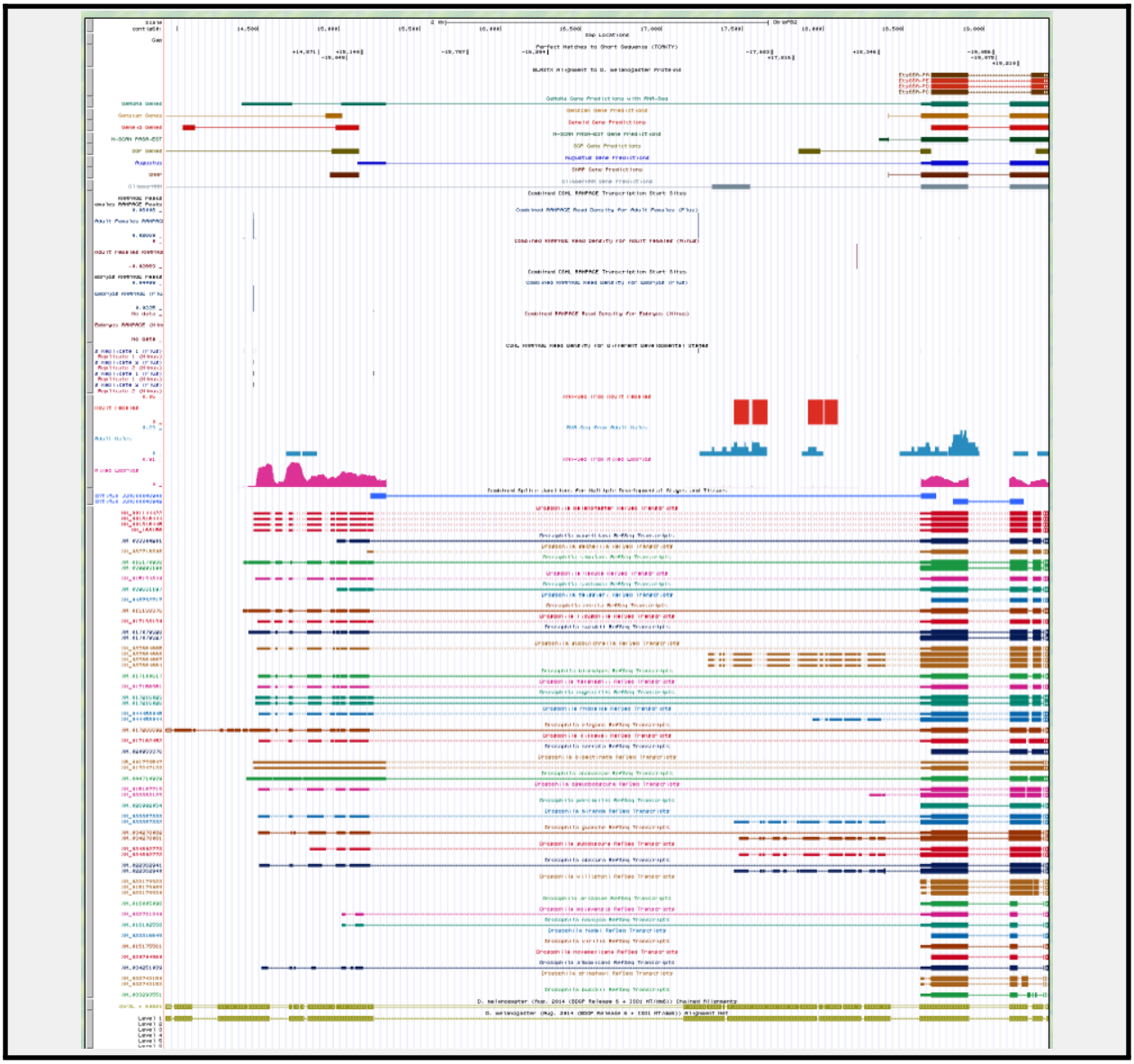
If not, specify the parts of the *D. melanogaster* exon that are missing from the *BLASTN* alignment.
Missing the entire first exon. Coordinates: 18671-18903

If the TSS annotation is supported by *BLASTN* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *BLASTN* alignment into the box below:**



4. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, **paste a screenshot of the multiple sequence alignment (*e.g.*, from *Clustal Omega*, *ROAST*) into the box below:**



5. Summarize the evidence that supports the TSS annotation postulated above

Coordinate(s) of the TSS position(s):

Based on RAMPAGE data (if applicable): 14,472; 14,474

Based on RNA-Seq data: 14,470-15,200

Based on *BLASTN* alignment: 14,474

Based on other evidence (please specify): _____

Note: If the *BLASTN* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter "Insufficient evidence" if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS position based on the available evidence?

_____yes_____

If so, indicate in the table below the evidence that supports this TSS position

If not, were you able to define a TSS search region? _____

If so, indicate in the table below the evidence that supports the TSS search region(s)

For each evidence type, enter an "X" in the cell to indicate whether the line of evidence supports, refutes, or neither supports nor refutes the TSS annotation:

Evidence type	Support	Refute	Neither
RAMPAGE peaks and read density	yes		
RNA-Seq coverage and splice junctions	yes		
BLASTN alignment of the initial exon from *D. melanogaster*	yes		
Sequence conservation with other *Drosophila* species (*e.g.*, "Conservation" track on the Genome Browser)	yes		
Other (please specify) [*e.g.*, *RefSeq Genes, N-SCAN PASA-EST, Augustus* TSS predictions; histone modifications (ChIP-Seq data)].			

Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against the annotated TSS; check "Refute." In contrast, the lack of RNA-Seq read coverage is a negative result that neither supports nor refutes the TSS annotation; check "Neither."

Provide an explanation if the TSS annotation is inconsistent with at least one of the evidence types specified above:

Isoform TSS Report

Complete an Isoform TSS report (through page PAGEREF _Ref91436230 \h 17) for each unique TSS listed in the table above. Copy and paste this form to create as many copies as needed.

Gene-isoform name (*e.g.*, *dbip_ey-RA*): *dbip_Ets65A-PB*

Names of the isoforms with the same TSS as this isoform:
N/A

Type of core promoter in *D. melanogaster* (see table below):
(Peaked / Intermediate / Broad / Insufficient Evidence)
peaked

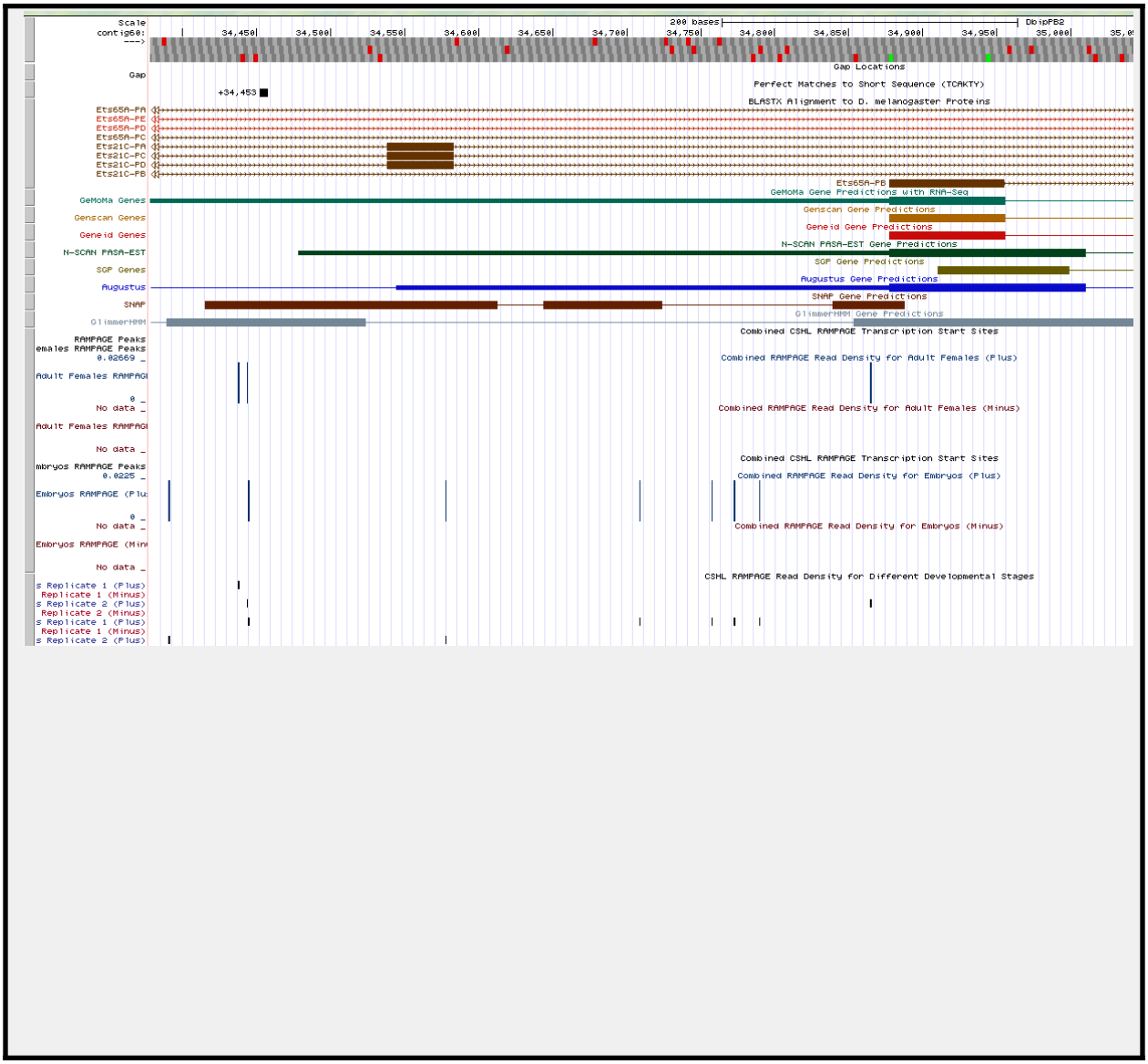
The type of core promoter is defined by the number of TSS annotated by the Celniker group at modENCODE and the number of DHS positions:

1. Turn on RAMPAGE evidence tracks (Only applies to projects with these tracks)

Coordinates of the TSS position based on position with the highest RAMPAGE read density
34,400-34,865

Coordinates of the narrow TSS search region based on RAMPAGE peaks
34,438; 34,444; 34,445; 34,865

If the TSS position and narrow TSS search region are supported by RAMPAGE data, **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the Combined RAMPAGE TSS evidence track:**



2. Turn on RNA-Seq evidence tracks

If the TSS annotation is supported by RNA-Seq read coverage or splice junction predictions (*e.g.*, *regtools*), **paste a Genome Browser screenshot of the region surrounding the putative TSS (±300bp) showing the following evidence tracks:**

1. RNA-Seq Coverage or RNA-Seq Alignment Summary
2. Combined Splice Junctions or RNA-Seq *TopHat*

If the RNA-Seq evidence tracks indicate the TSS position, list it here:

_____

If the RNA-Seq evidence tracks indicate a TSS search region, list it here:
_34,756-34,865_____

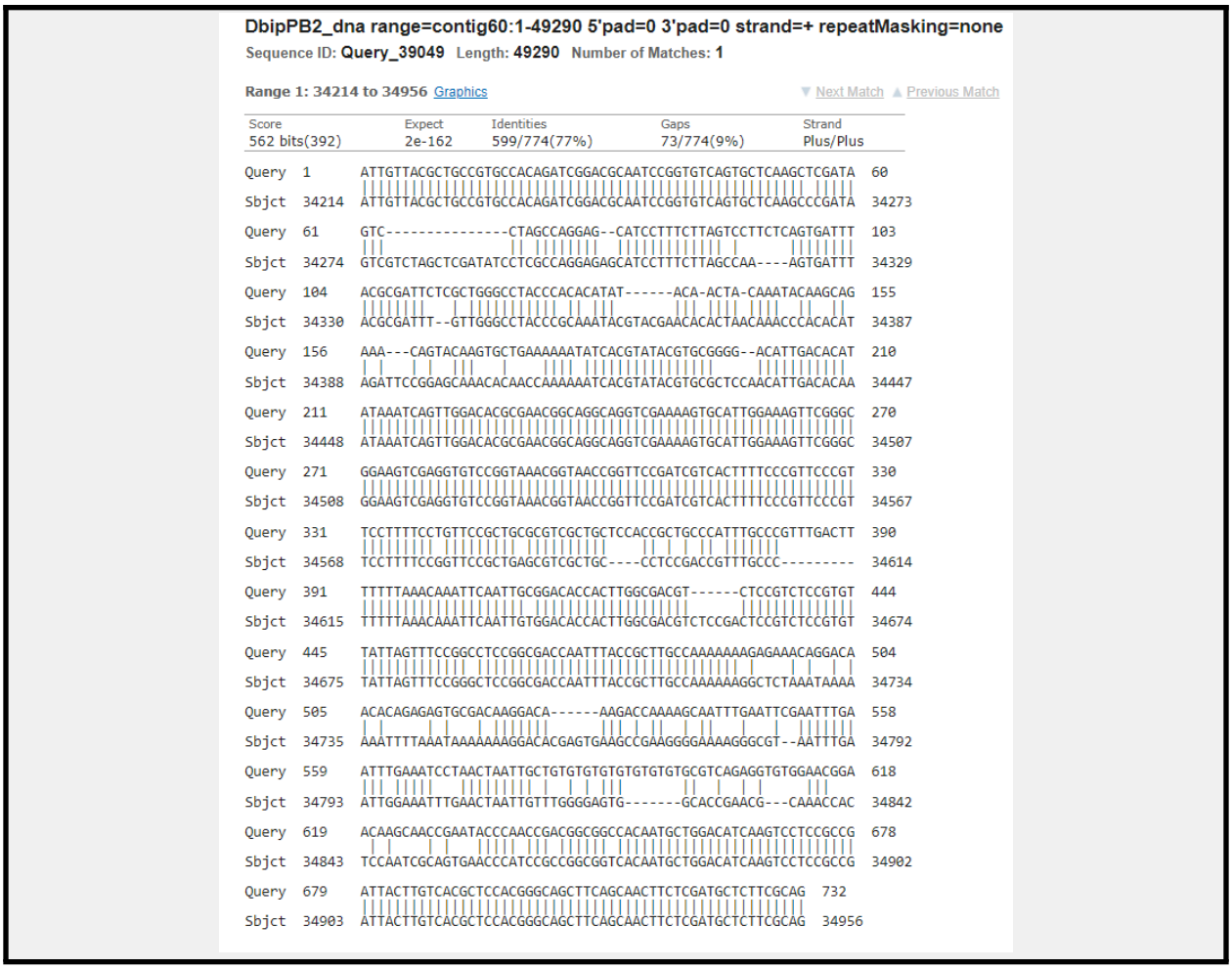
3. Annotate the first transcribed exon

Coordinates of the first transcribed exon based on *BLASTN* alignment:
34214 to 34956

Does the *BLASTN* alignment cover the entire *D. melanogaster* first transcribed exon?
Yes

If not, specify the parts of the *D. melanogaster* exon that are missing from the *BLASTN* alignment.

If the TSS annotation is supported by *BLASTN* alignment of the initial transcribed exon against the contig sequence, **paste a screenshot of the *BLASTN* alignment into the box below:**



DbipPB2_dna range=contig60:1-49290 5'pad=0 3'pad=0 strand=+ repeatMasking=none

Sequence ID: **Query_39049** Length: **49290** Number of Matches: **1**

Range 1: 34214 to 34956 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
562 bits(392)	2e-162	599/774(77%)	73/774(9%)	Plus/Plus

```
Query  1      ATTGTTACGCTGCCGTGCCACAGATCGGACGCAATCCGGTGTCAGTGCTCAAGCTCGATA  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  34214  ATTGTTACGCTGCCGTGCCACAGATCGGACGCAATCCGGTGTCAGTGCTCAAGCCCGATA  34273

Query  61     GTC---------------CTAGCCAGGAG--CATCCTTTCTTAGTCCTTCTCAGTGATTT  103
              |||               || ||||||||  |||||||||||| |      ||||||||
Sbjct  34274  GTCGTCTAGCTCGATATCCTCGCCAGGAGAGCATCCTTTCTTAGCCAA----AGTGATTT  34329

Query  104    ACGCGATTCTCGCTGGGCCTACCCACACATAT------ACA-ACTA-CAAATACAAGCAG  155
              ||||||||  | ||||||||||| || |||       ||| |||| |||| ||  ||
Sbjct  34330  ACGCGATTT--GTTGGGCCTACCCGCAAATACGTACGAACACACTAACAAACCCACACAT  34387

Query  156    AAA---CAGTACAAGTGCTGAAAAAATATCACGTATACGTGCGGGG--ACATTGACACAT  210
              | |   | |  |||   |    |||| ||||||||||||||     ||||||||||
Sbjct  34388  AGATTCCGGAGCAAACACAACCAAAAAATCACGTATACGTGCGCTCCAACATTGACACAA  34447

Query  211    ATAAATCAGTTGGACACGCGAACGGCAGGCAGGTCGAAAAGTGCATTGGAAAGTTCGGGC  270
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  34448  ATAAATCAGTTGGACACGCGAACGGCAGGCAGGTCGAAAAGTGCATTGGAAAGTTCGGGC  34507

Query  271    GGAAGTCGAGGTGTCCGGTAAACGGTAACCGGTTCCGATCGTCACTTTTCCCGTTCCCGT  330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  34508  GGAAGTCGAGGTGTCCGGTAAACGGTAACCGGTTCCGATCGTCACTTTTCCCGTTCCCGT  34567

Query  331    TCCTTTTCCTGTTCCGCTGCGCGTCGCTGCTCCACCGCTGCCCATTTGCCCGTTTGACTT  390
              |||||||| |||||||| |||||||||     || | | || |||||||
Sbjct  34568  TCCTTTTCCGGTTCCGCTGAGCGTCGCTGC----CCTCCGACCGTTTGCCC---------  34614

Query  391    TTTTTAAACAAATTCAATTGCGGACACCACTTGGCGACGT------CTCCGTCTCCGTGT  444
              |||||||||||||||||| ||||||||||||||||||||      ||||||||||||||
Sbjct  34615  TTTTTAAACAAATTCAATTGTGGACACCACTTGGCGACGTCTCCGACTCCGTCTCCGTGT  34674

Query  445    TATTAGTTTCCGGCCTCCGGCGACCAATTTACCGCTTGCCAAAAAAAGAGAAACAGGACA  504
              |||||||||||| |||||||||||||||||||||||||||||||||     | |   | |
Sbjct  34675  TATTAGTTTCCGGGCTCCGGCGACCAATTTACCGCTTGCCAAAAAAGGCTCTAAATAAAA  34734

Query  505    ACACAGAGAGTGCGACAAGGACA------AAGACCAAAAGCAATTTGAATTCGAATTTGA  558
              | |     | |   | |||||||      ||| | ||  | ||    |  |  |||||||
Sbjct  34735  AAATTTTAAATAAAAAAAGGACACGAGTGAAGCCGAAGGGGAAAAGGGCGT--AATTTGA  34792

Query  559    ATTTGAAATCCTAACTAATTGCTGTGTGTGTGTGTGTGTGCGTCAGAGGTGTGGAACGGA  618
              ||| |||||   ||||||||| |  | | |||       ||  |  | |     |||
Sbjct  34793  ATTGGAAATTTGAACTAATTGTTTGGGGAGTG-------GCACCGAACG---CAAACCAC  34842

Query  619    ACAAGCAACCGAATACCCAACCGACGGCGGCCACAATGCTGGACATCAAGTCCTCCGCCG  678
               | |   | |  ||||| ||| |||||| ||||||||||||||||||||||||||||||
Sbjct  34843  TCCAATCGCAGTGAACCCATCCGCCGGCGGTCACAATGCTGGACATCAAGTCCTCCGCCG  34902

Query  679    ATTACTTGTCACGCTCCACGGGCAGCTTCAGCAACTTCTCGATGCTCTTCGCAG  732
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  34903  ATTACTTGTCACGCTCCACGGGCAGCTTCAGCAACTTCTCGATGCTCTTCGCAG  34956
```

4. Turn on comparative genomics tracks

If the TSS annotation is supported by sequence conservation with other *Drosophila* species, **paste a screenshot of the multiple sequence alignment (*e.g.*, from *Clustal Omega*, *ROAST*) into the box below:**

5. Summarize the evidence that supports the TSS annotation postulated above

Coordinate(s) of the TSS position(s):

Based on RAMPAGE data (if applicable): 34,865
Based on RNA-Seq data: 34,865
Based on *BLASTN* alignment: 34,865
Based on other evidence (please specify): _____

Note: If the *BLASTN* alignment for the initial transcribed exon is a partial alignment, you can **extrapolate the TSS position** based on the number of nucleotides that are missing from the beginning of the exon. (Enter "Insufficient evidence" if you cannot determine the TSS position based on the available evidence.)

Were you able to define a TSS position based on the available evidence?
_____yes;34,865_____
If so, indicate in the table below the evidence that supports this TSS position

If not, were you able to define a TSS search region? _____
If so, indicate in the table below the evidence that supports the TSS search region(s)

For each evidence type, enter an "X" in the cell to indicate whether the line of evidence supports, refutes, or neither supports nor refutes the TSS annotation:

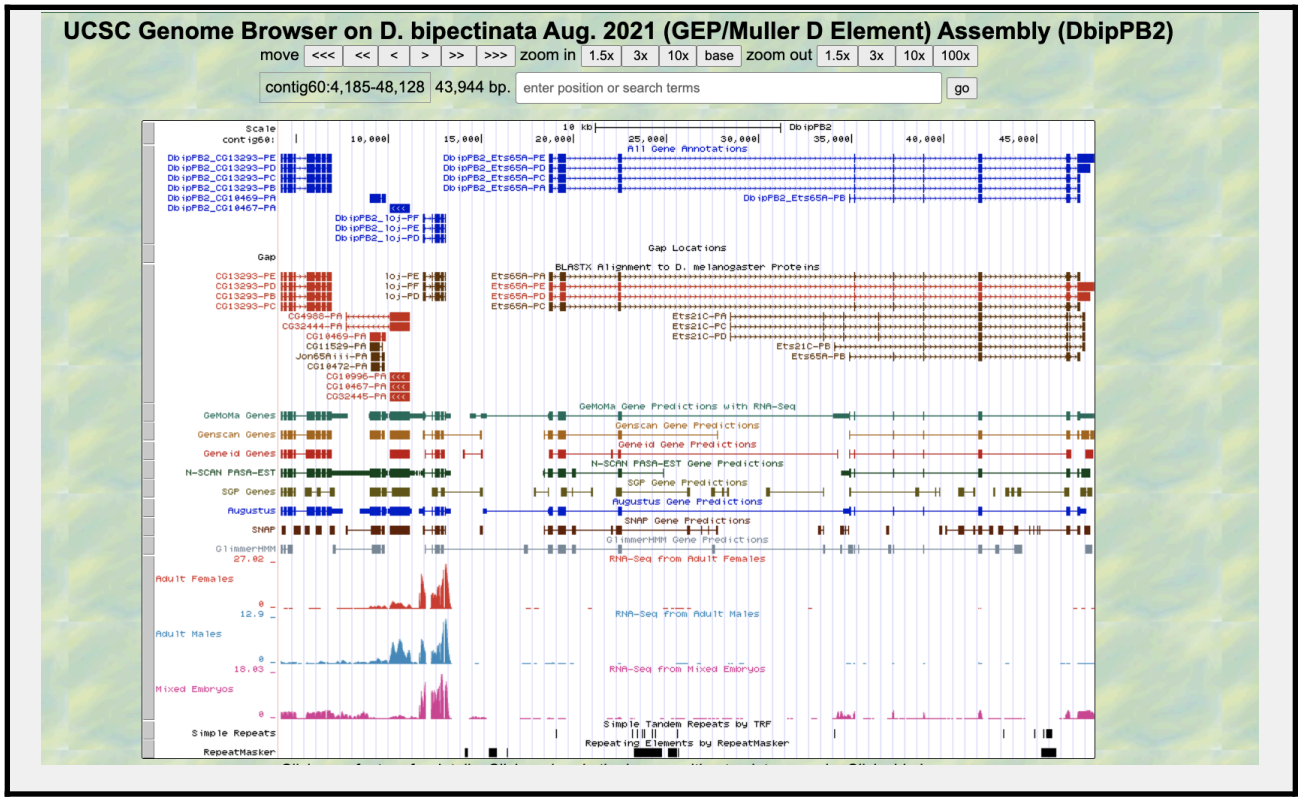
Evidence type	Support	Refute	Neither
RAMPAGE peaks and read density	yes		
RNA-Seq coverage and splice junctions	yes		
BLASTN alignment of the initial exon from *D. melanogaster*	yes		
Sequence conservation with other *Drosophila* species (*e.g.*, "Conservation" track on the Genome Browser)	yes		
Other (please specify) [*e.g.*, *RefSeq Genes, N-SCAN PASA-EST, Augustus* TSS predictions; histone modifications (ChIP-Seq data)].			

Note: The evidence type refutes the TSS annotation only if it **suggests an alternate TSS position**. For example, the presence of RNA-Seq read coverage upstream of the annotated TSS indicates that the TSS is located further upstream and it would be considered to be evidence against the annotated TSS; check "Refute." In contrast, the lack of RNA-Seq read coverage is a negative result that neither supports nor refutes the TSS annotation; check "Neither."

Preparing the Project for Submission

For each project, you should prepare the project GFF, transcript, 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*. Once you have combined the GFF files into a single file, click on the "**Show Track**" button to view all the gene models in the combined GFF file within the Genome Browser.

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



Thank you for your submission, and congratulations on completing your analysis of this region of this genome. Our planned GEP meta-analysis of the genes and genomes in this study depends on the high quality annotations accomplished by GEP students.