

Annotation of the gene *ci* on contig 3 dot chromosome of *Drosophila eugracilis*



Introduction

- Dot chromosome (Chromosome 4 or Element F; largely heterochromatic)
- Gene name: *cubitus interruptus* (*ci*). Segment polarity gene; cytoplasmic protein coding gene.
- Protein features: GLI C2H2-type zinc-finger protein family involved in Hedgehog signaling pathway
- Evidence for 17 unique biological process terms, ex: cuticle pattern formation; epidermis, heart, and neuron development; dendrite, eye morphogenesis

Workflow

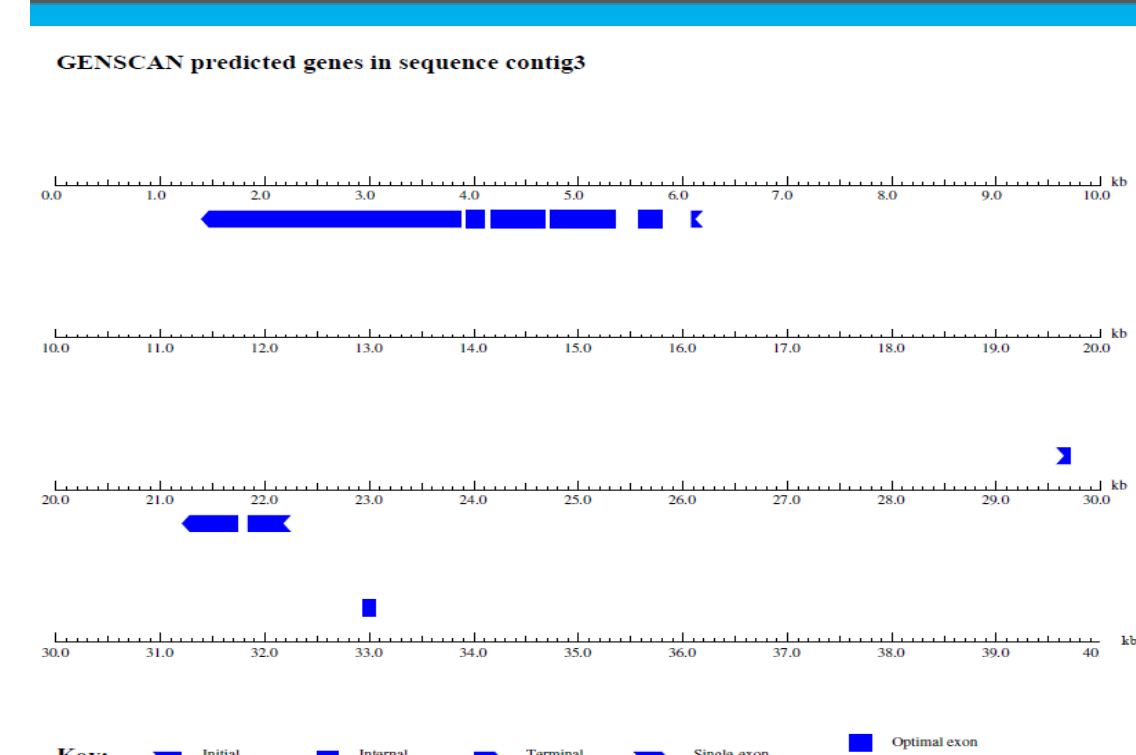


Figure 1: GenScan prediction of contig 3

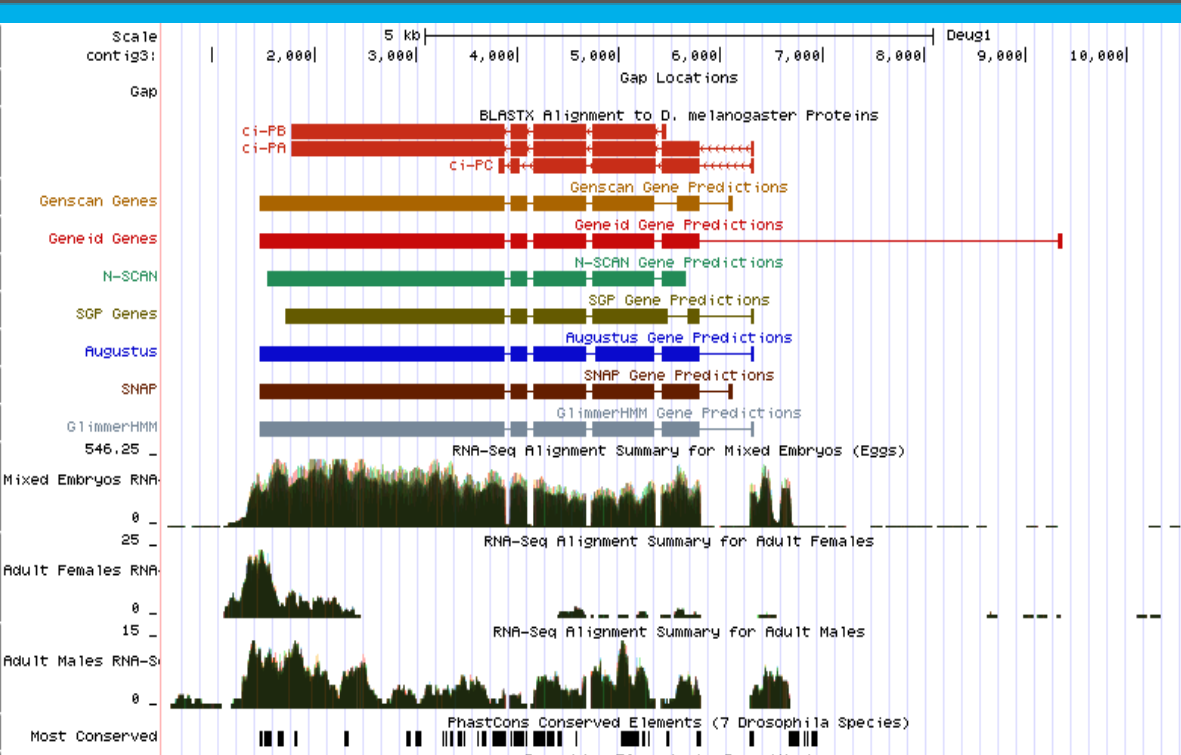


Figure 2: Identifying contig3 using Genome Browser

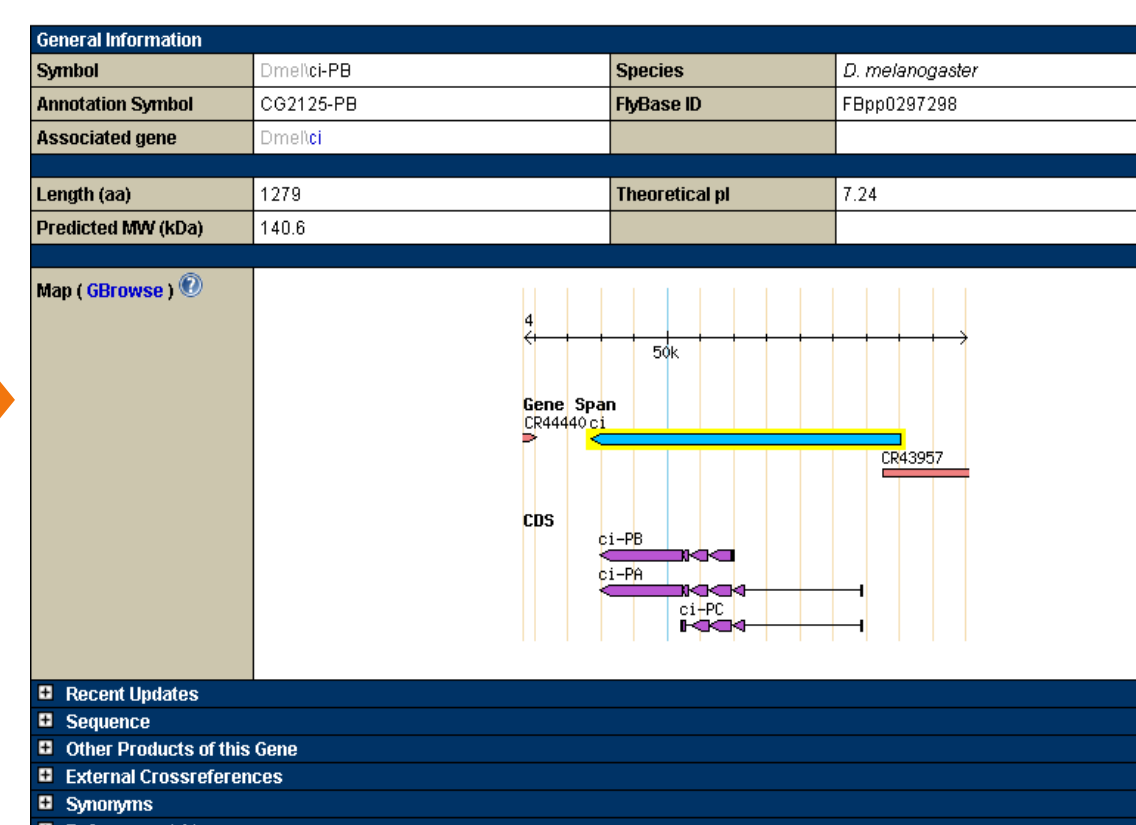


Figure 3: Checking for correct chromosome location using FlyBase ID

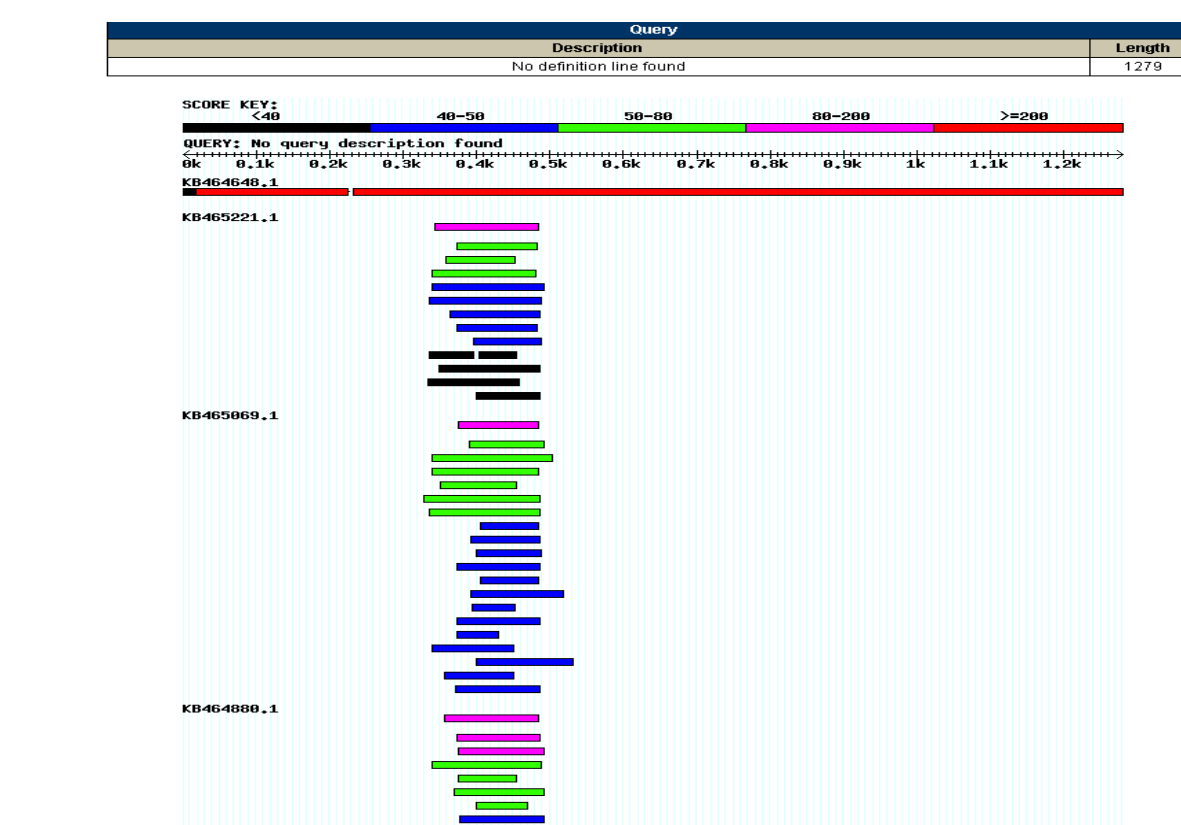


Figure 4: Identifying the beginning and end of the contig and scaffold using blast

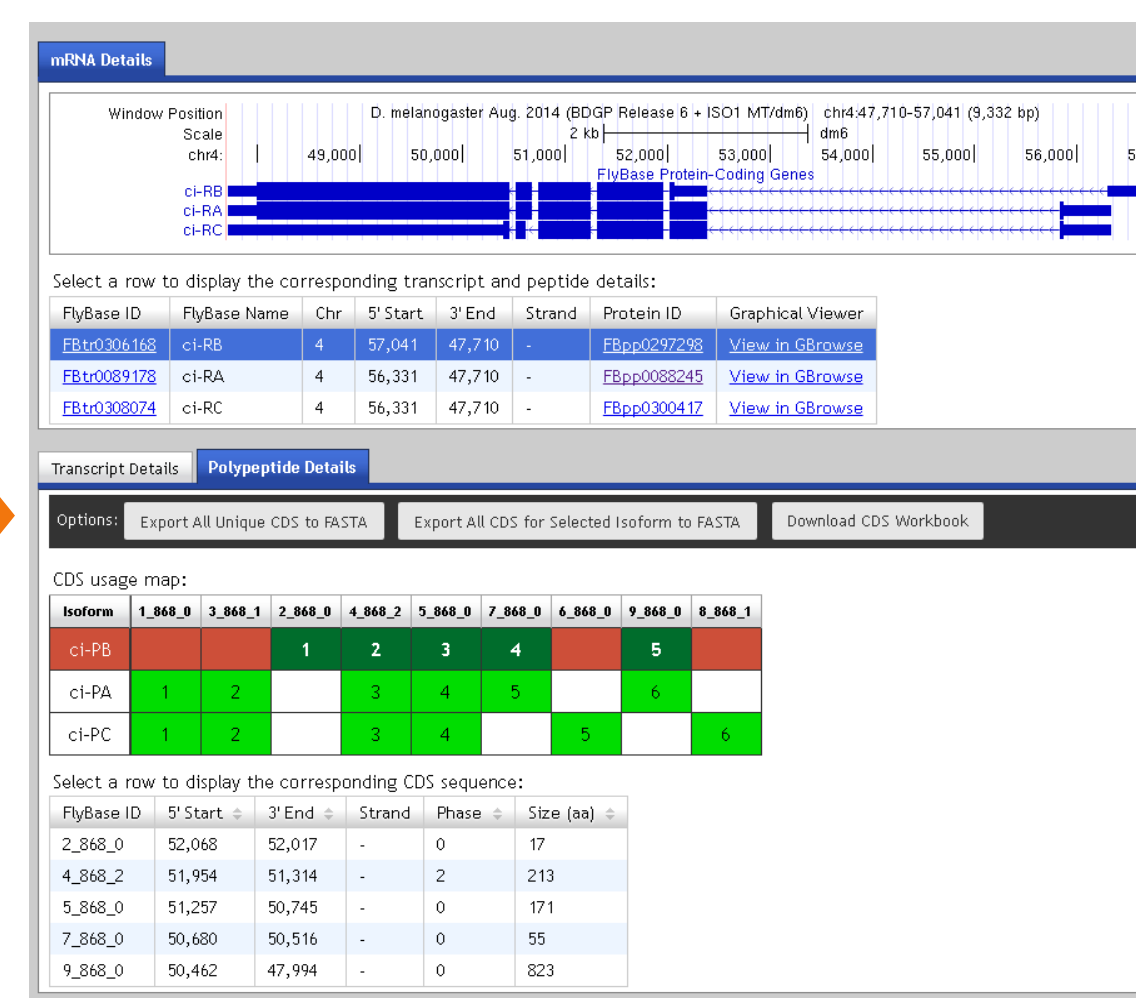


Figure 5: Identifying isoforms using Gene Finder and obtain CDS workbook

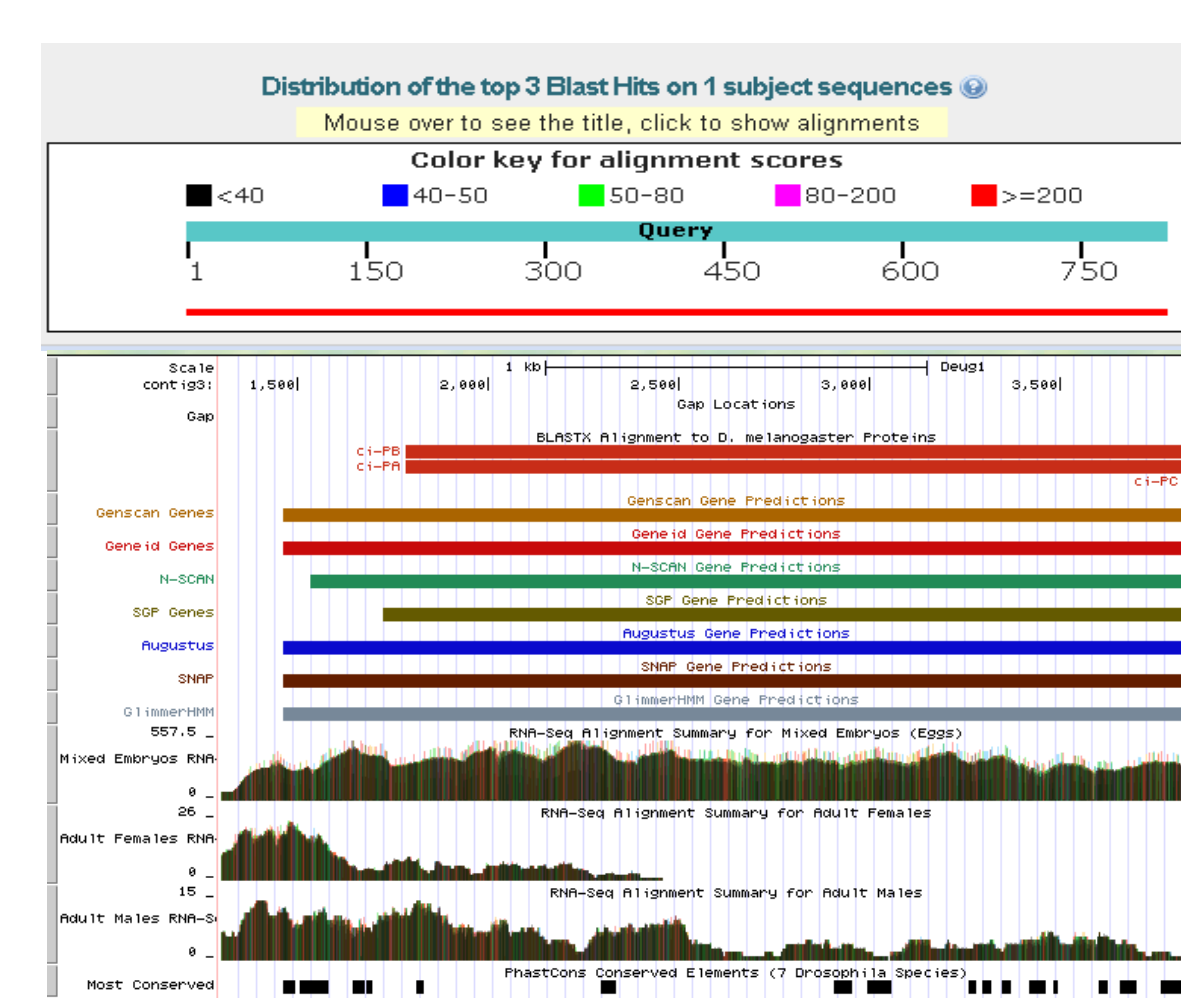


Figure 6: Annotating the last exon using tblast

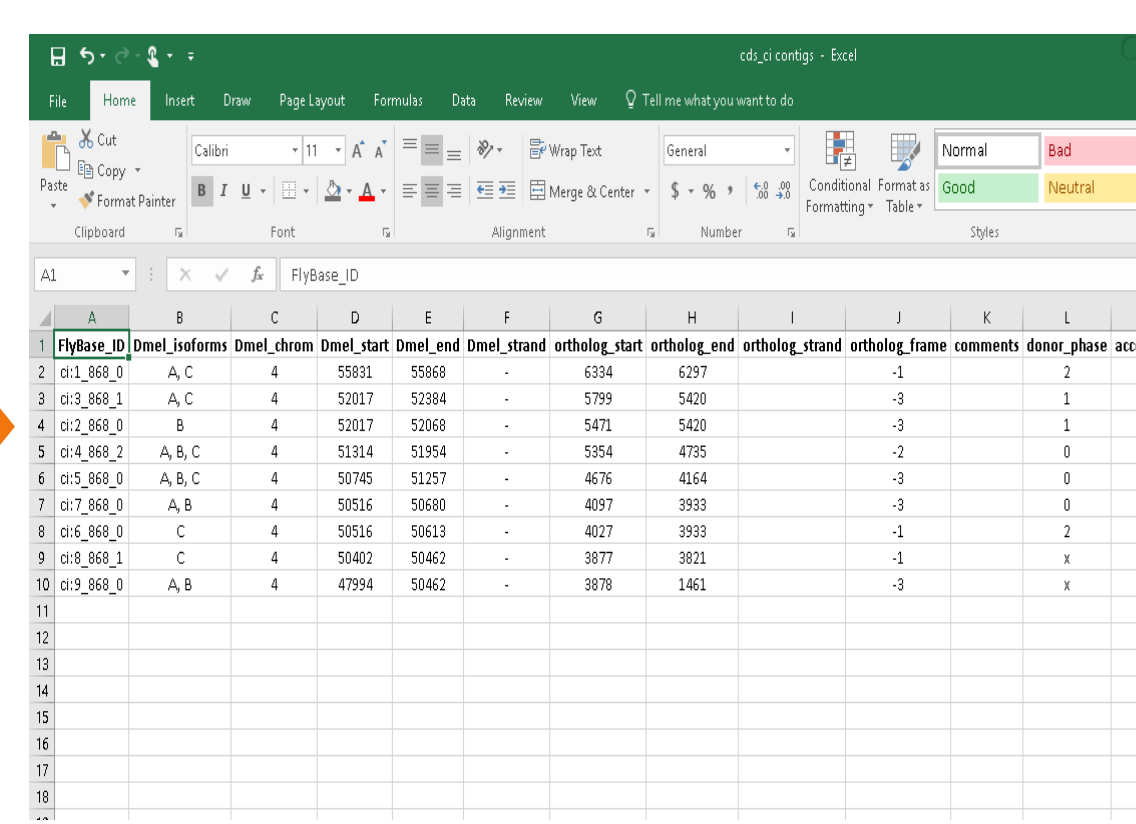


Figure 7: Checking corrected frame, boundaries, and phase in WDS workbook

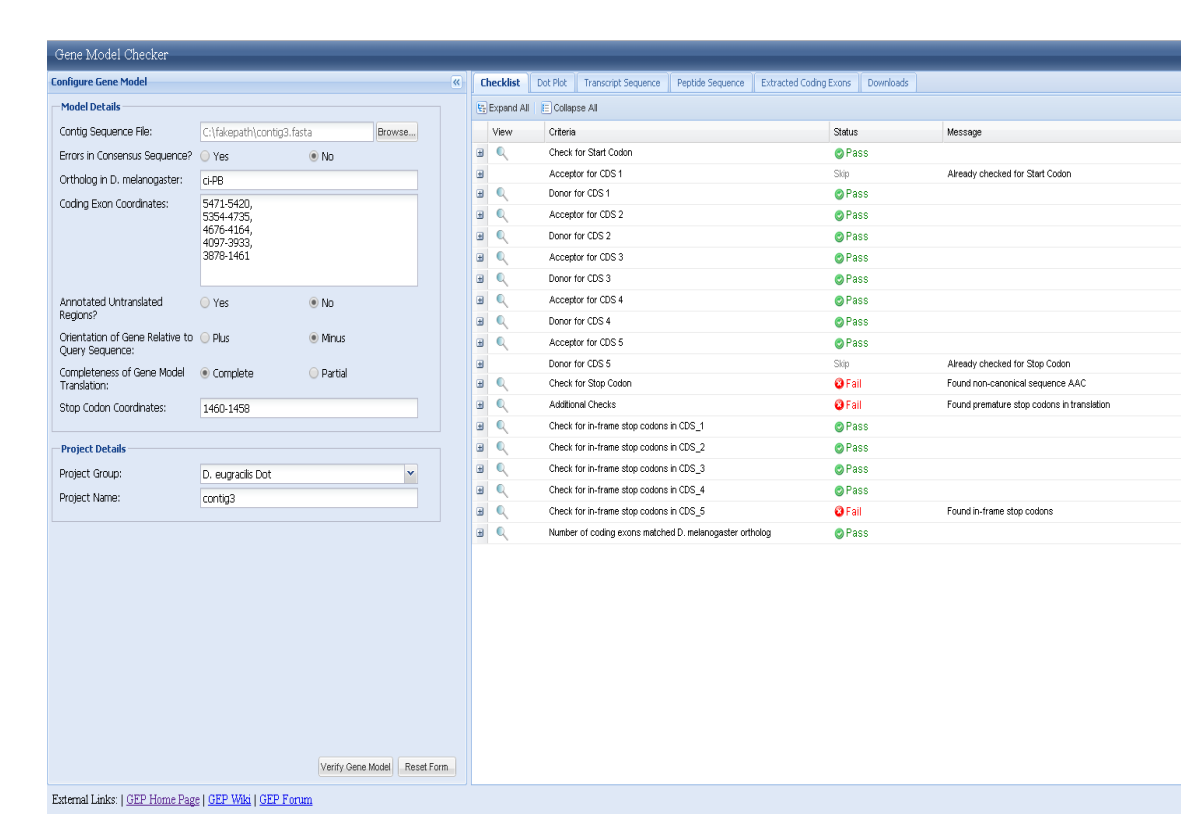


Figure 8: Checking exon coordinates for corresponding isoform in Gene Model Checker and fixing errors

Results

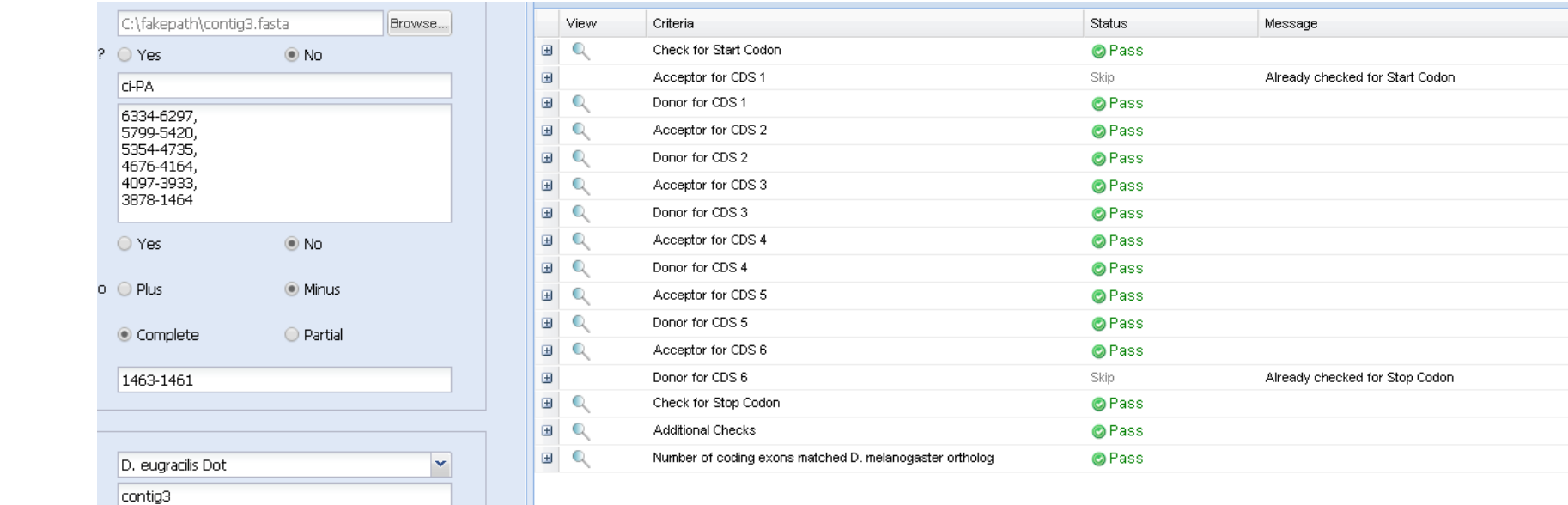


Figure 9: *ci-PA*

- Gene Model checker indicates pass for coding exon coordinates of isoform *ci-PA*.
- Out of total of 6 exons, gaps can be seen in middle of exon 2, both ends of exon 3, beginning of exon 4, and middle of exon 6.
- Protein alignment shows mismatches (ex: 61-115 and 300-345).
- Small segments of repetition in exons 2, 5, and 6.

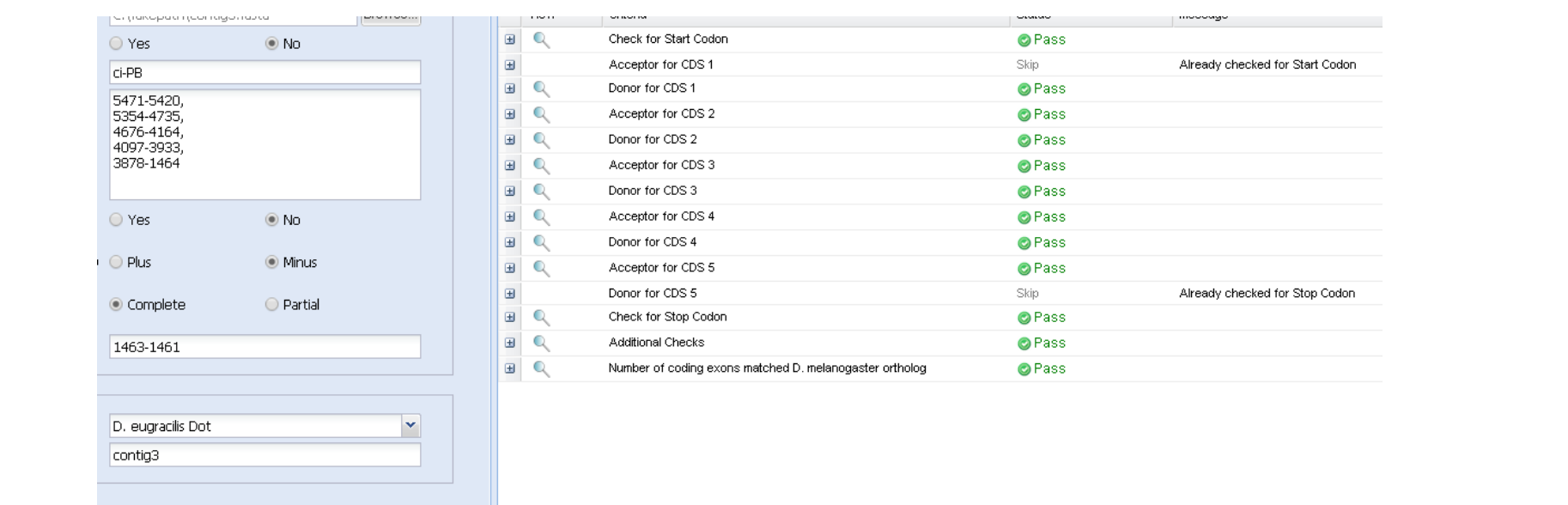
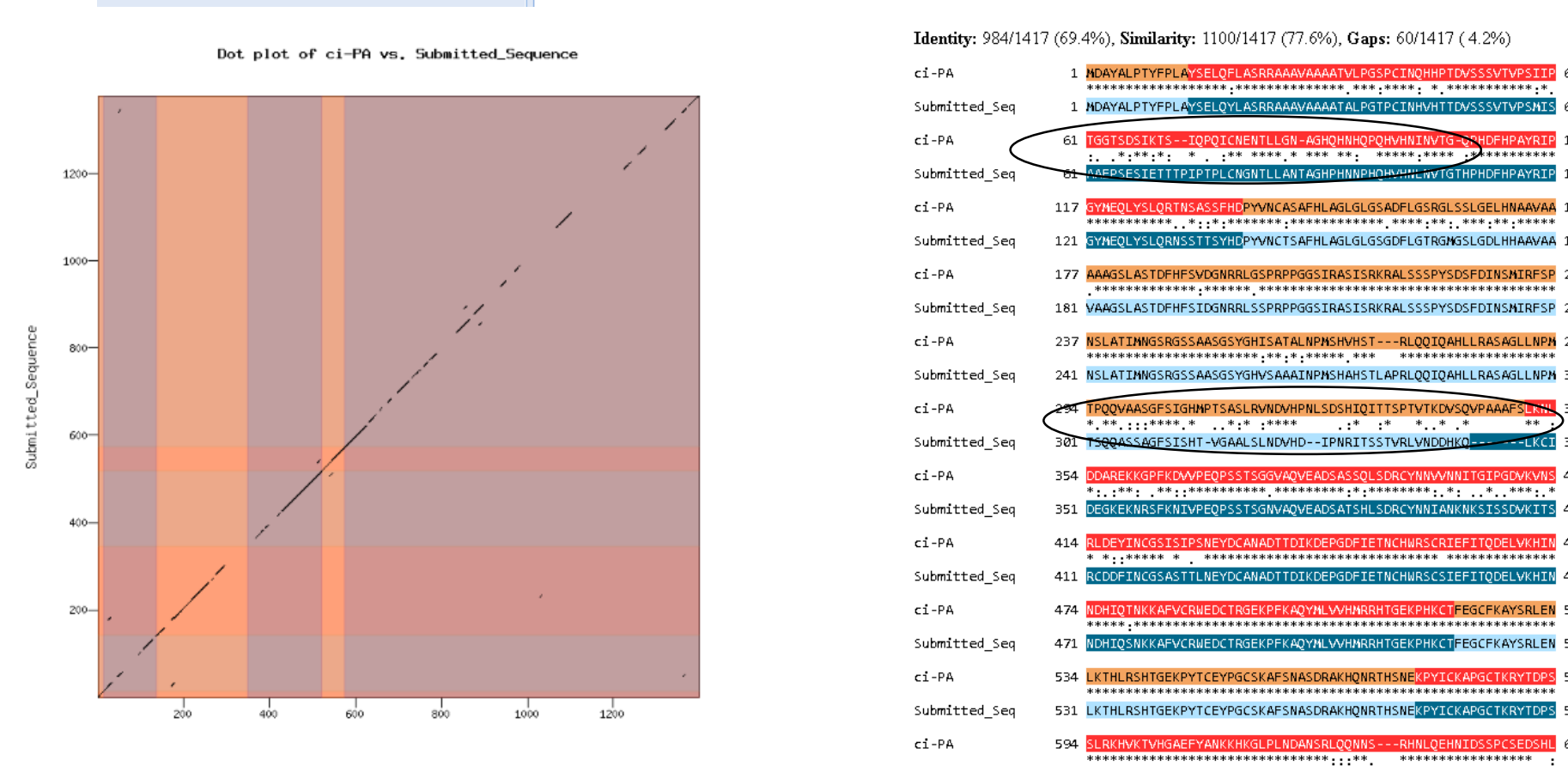


Figure 10: *ci-PB*

- Gene Model Checker indicates pass for the coding exon coordinates of isoform *ci-PB*.
- From total of 5 exons, gaps exist at the middle and end of exon 2, beginning and middle of exon 3, and middle of exon 5.
- The gaps are shown in the corresponding alignment. (ex: sequences 180-230)
- Small segments are shown in exons 4 and 5.

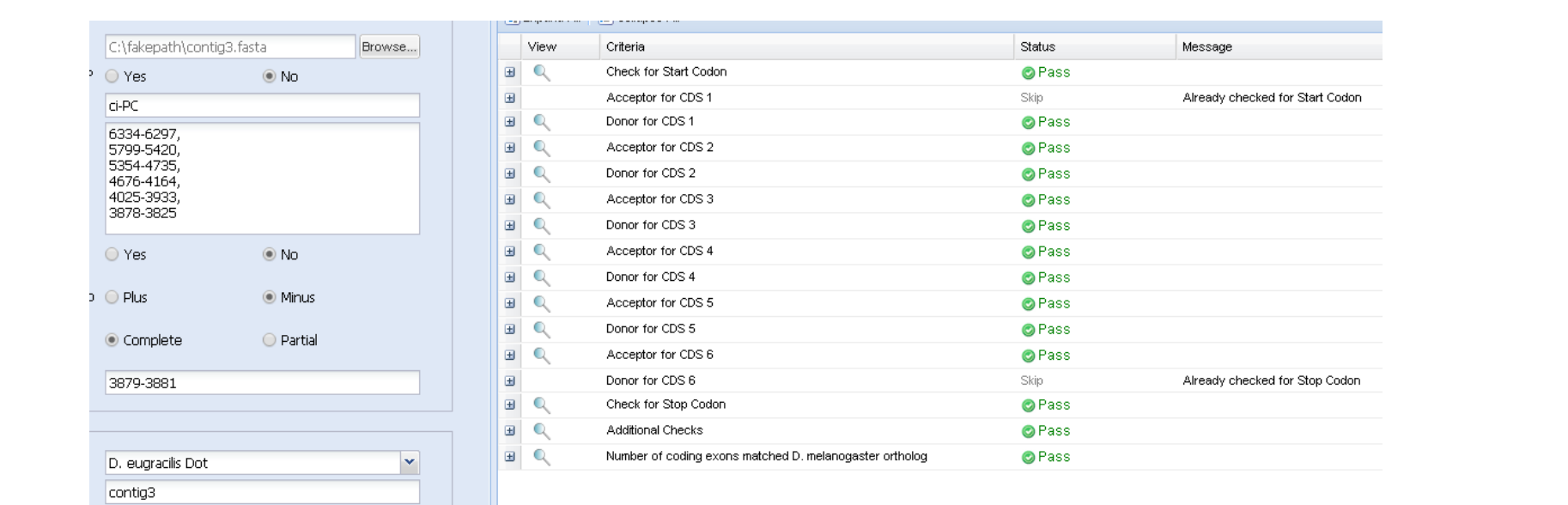
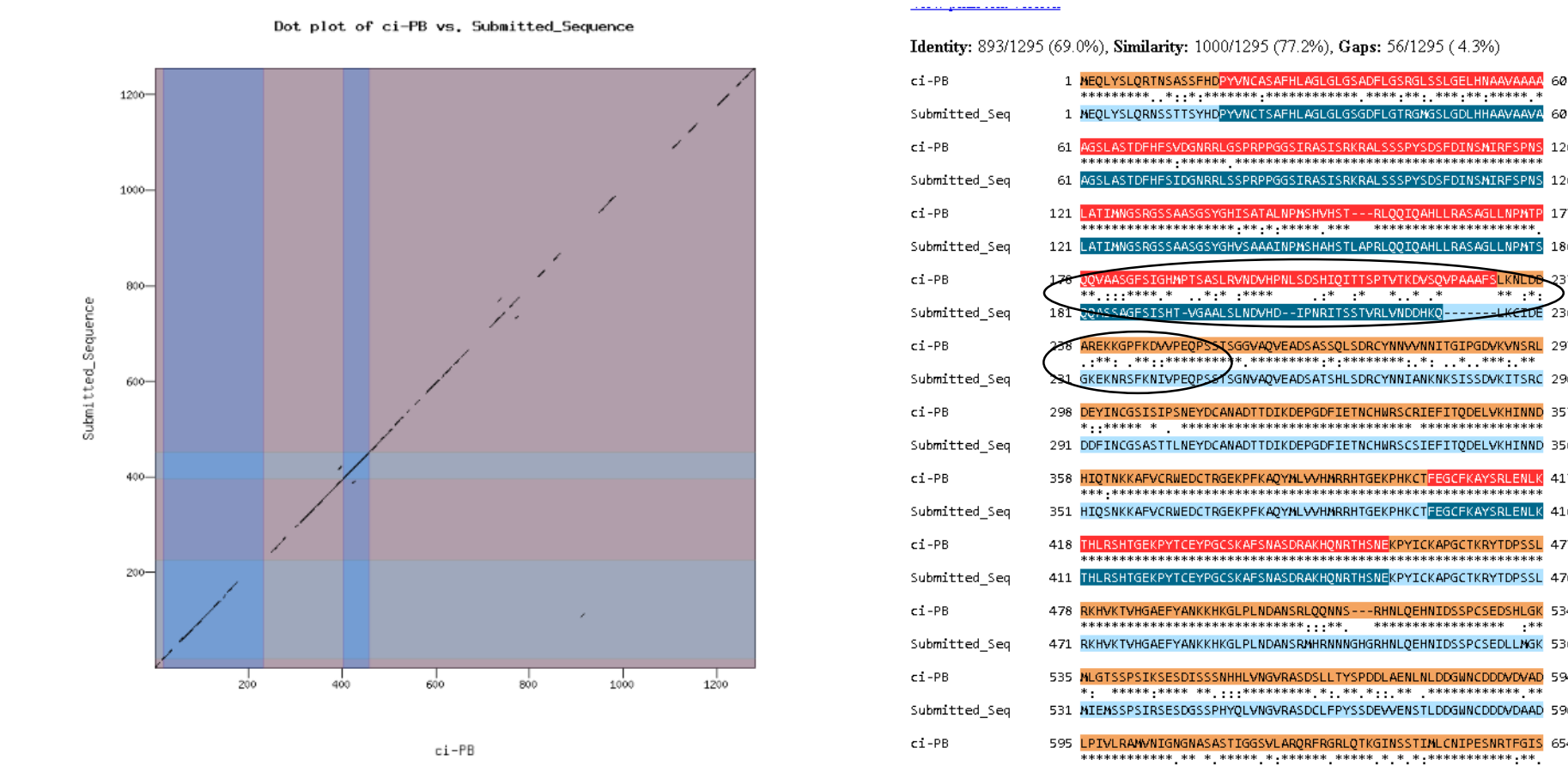
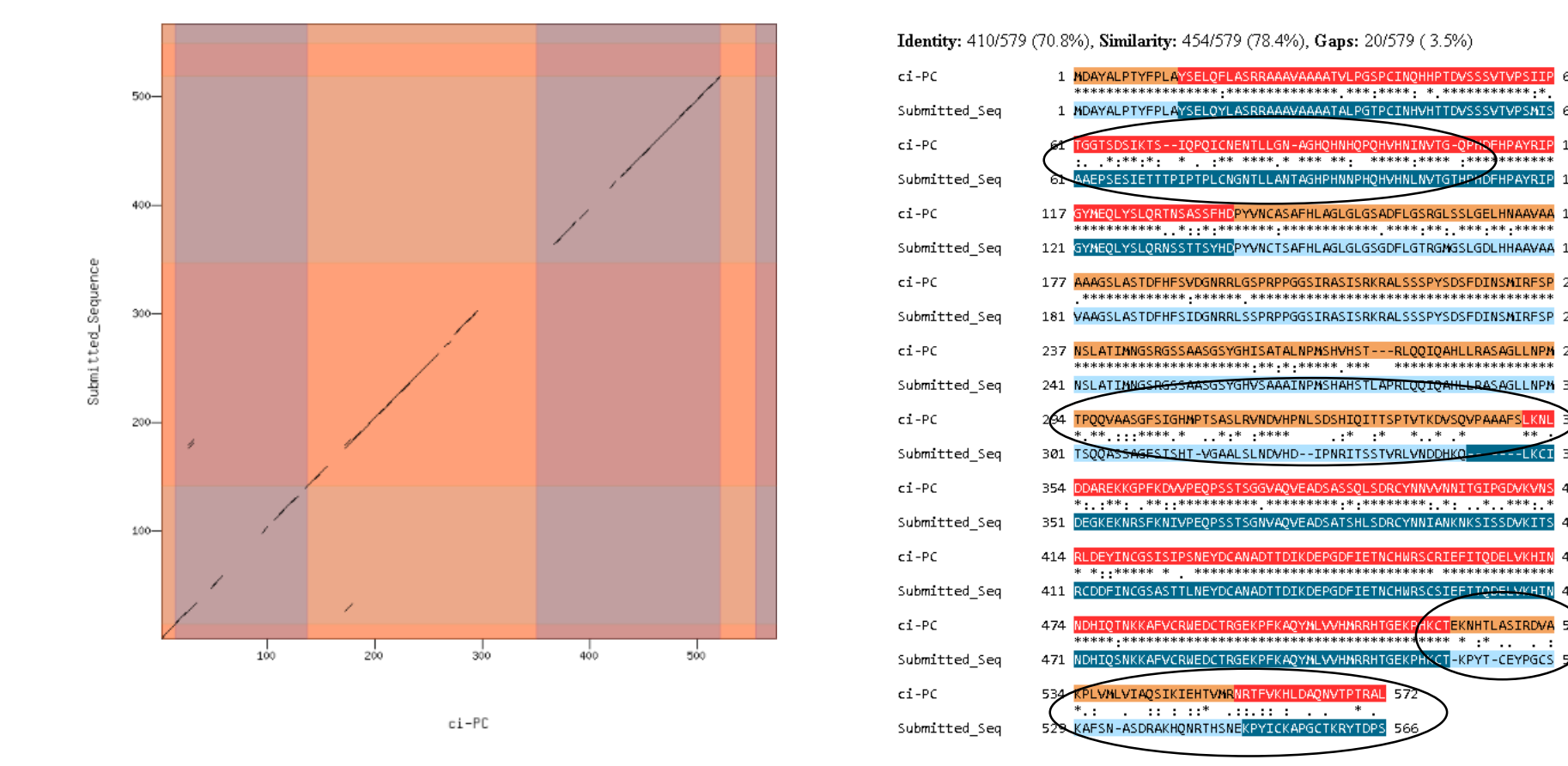


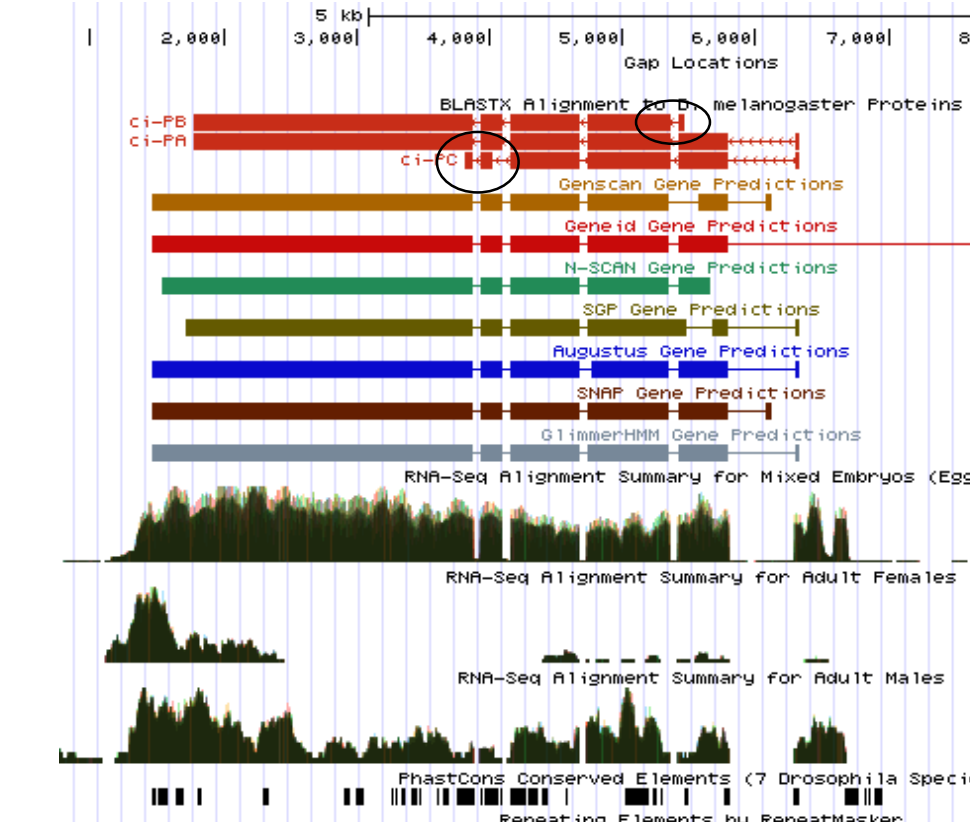
Figure 11: *ci-PC*

- Gene Model Checker indicates all pass for the coding exon coordinates of isoform *ci-PC*.
- Out of total of 6 exons, gaps exist in middle of exon 2, in between and at the end of exon 3, beginning and middle of exon 4, and missing in exons 5 and 6.
- Alignments indicate mismatches from 60 – 110, 300 – 450, and 500 – 570.
- Small segments exist in exon 3.



Discussion

- Annotation of the contig 3 of chromosome 4 indicates somewhat importance of the gene, *ci*.
- *ci* gene is associated with development of organ system, adult segment, integumentary specialization, and peripheral nervous system.
- Moderate selection for *ci* expected due to the heterochromatic characteristic of the dot chromosome and the function of *ci*.
- **Deviations:**
- *ci-PA* and *ci-PB*: numerous gaps in exon 6 but without error
- *ci-PC*: no match for exons 5 and 6 in dot plot and mismatches in protein alignment. Stop codon was moved to 3379-3881 because exons 5 and 6 may be not exist in *D. eugracilis*



Exon 1 of *ci-PA* is a partial exon compared to that of *ci-PB* and *ci-PC*. Similarly for *ci-PC*, exon 5 and 6 are partial exons compared to that of *ci-PA* and *ci-PB*. This may indicate that these exons may not really exist that predicted.

- The gaps in the middle of exons may be due to mutation throughout speciation.
- Gaps at the end of exons are explained mainly by the fact that the genome information may be inaccurate.
- Small segments are just random repetitions in the genome.

Is the gene conserved?

- Dot plot shows moderate solid and overall linear line
- Only 77% amino acid similarity overall.
- These results suggests relaxed selection and lower functional significance in this portion of the protein.
- There is evolutionary divergence between *D. melanogaster* and *D. eugracilis* lineages in these exons

Works Cited

- 1) <http://gep.wustl.edu/>
- 2) <http://flybase.org/reports/FBgn0004859.html>
- 3) <http://www.sdbonline.org/sites/fly/segment/cubitus.htm>