

Gene Annotation of protein Rich-P in *D. ananassae* using the Genomics Education Partnership (GEP) pipeline



Tia Watkins, Hailey Chang, and Dr. Sarah Justice

Background & Purpose

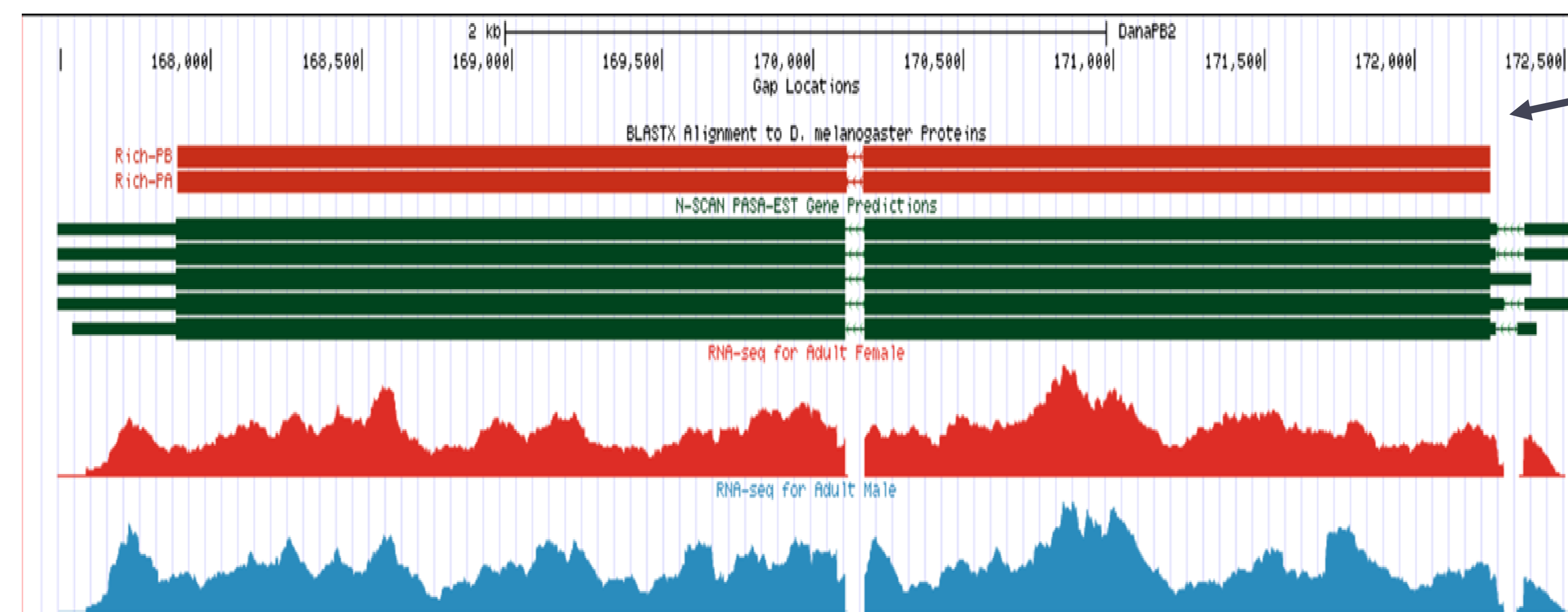
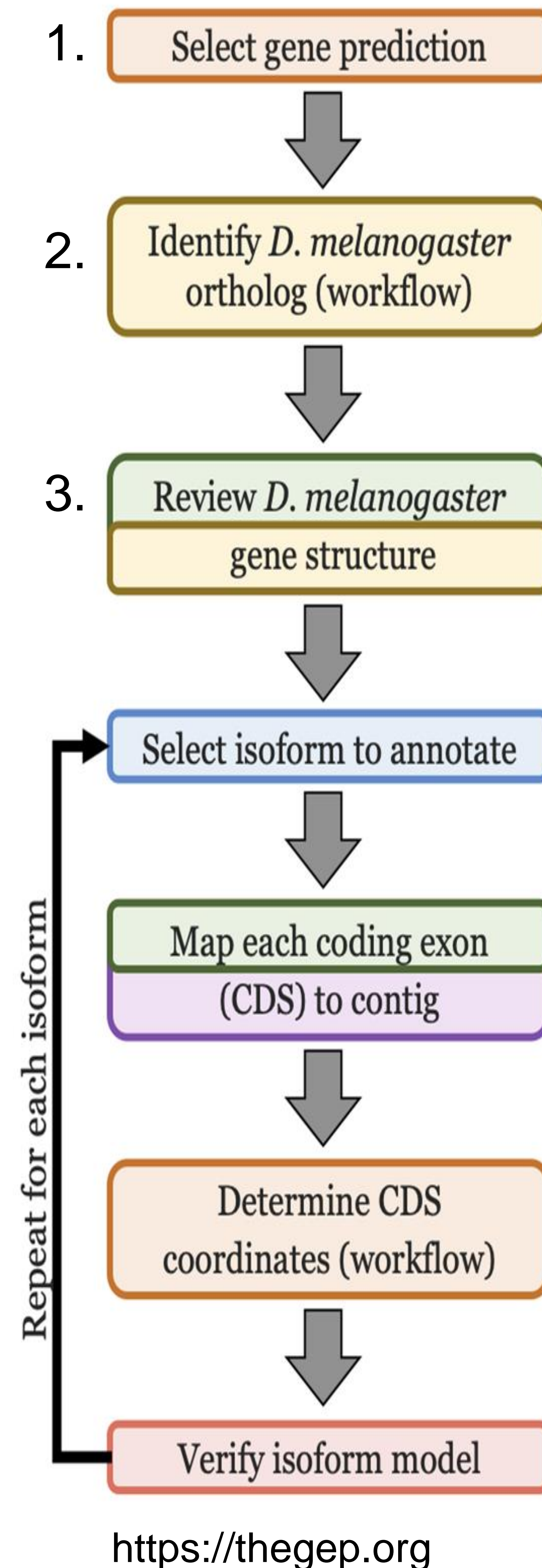
Gene annotation is the process of finding and giving meaning to the sequence of an unknown gene. More specifically, comparative genomics is one way to annotate genes that consists of using a known DNA sequence to characterize that of a similar or an unknown gene. Doing this enables gene prediction and analysis of the potential function of a gene of interest. The Genomics Education Partnership, the GEP, provides training in the needed annotation tools to perform successful comparative genomics of *Drosophila* species using the well characterized *D. melanogaster* genome.

In this research project, comparative genomics is used to assess the evolutionary impact of *Drosophila* genes by annotating coding regions and transcription start sites in *D. ananassae*. This will provide information into the genetic changes across *Drosophila* species and will identify factors that allow genes to function. In particular, we focused on annotation of the protein Rich. Accomplishing this work will lead to a better understanding of genomes and offers hopeful methods to lead to eventual advancements in health and medication due to the efficient way of annotating unknown genes in various species and finding their functions.

Acknowledgements

- TU Biology Dept.
- Sam, Grace, Becca, Kayla, and Landon
- Performed in collaboration with the GEP

Methods & Results



Gene compared in *D. Melanogaster* (Rich-P)

Gene Prediction Selection

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> guanine nucleotide exchange factor subunit Rich [<i>Drosophila ananassae</i>]	<i>Drosophila ananassae</i>	2847	2847	95%	0.0	99.93%	1433	XP_001958491.1
<input checked="" type="checkbox"/> PREDICTED: guanine nucleotide exchange factor subunit Rich [<i>Drosophila bipectinata</i>]	<i>Drosophila bipectinata</i>	2792	2792	95%	0.0	97.90%	1435	XP_017102598.1
<input checked="" type="checkbox"/> PREDICTED: guanine nucleotide exchange factor subunit Rich [<i>Drosophila elegans</i>]	<i>Drosophila elegans</i>	2539	2539	95%	0.0	88.71%	1435	XP_017123001.1
<input checked="" type="checkbox"/> PREDICTED: guanine nucleotide exchange factor subunit Rich [<i>Drosophila rhopaloea</i>]	<i>Drosophila rhopaloea</i>	2532	2532	95%	0.0	88.36%	1435	XP_016974721.1

Selected gene to annotate due to high percent identity

CDS usage map:

Isoform	1_2569_0	2_2569_0
Rich-PA	1	2
Rich-PB	1	2

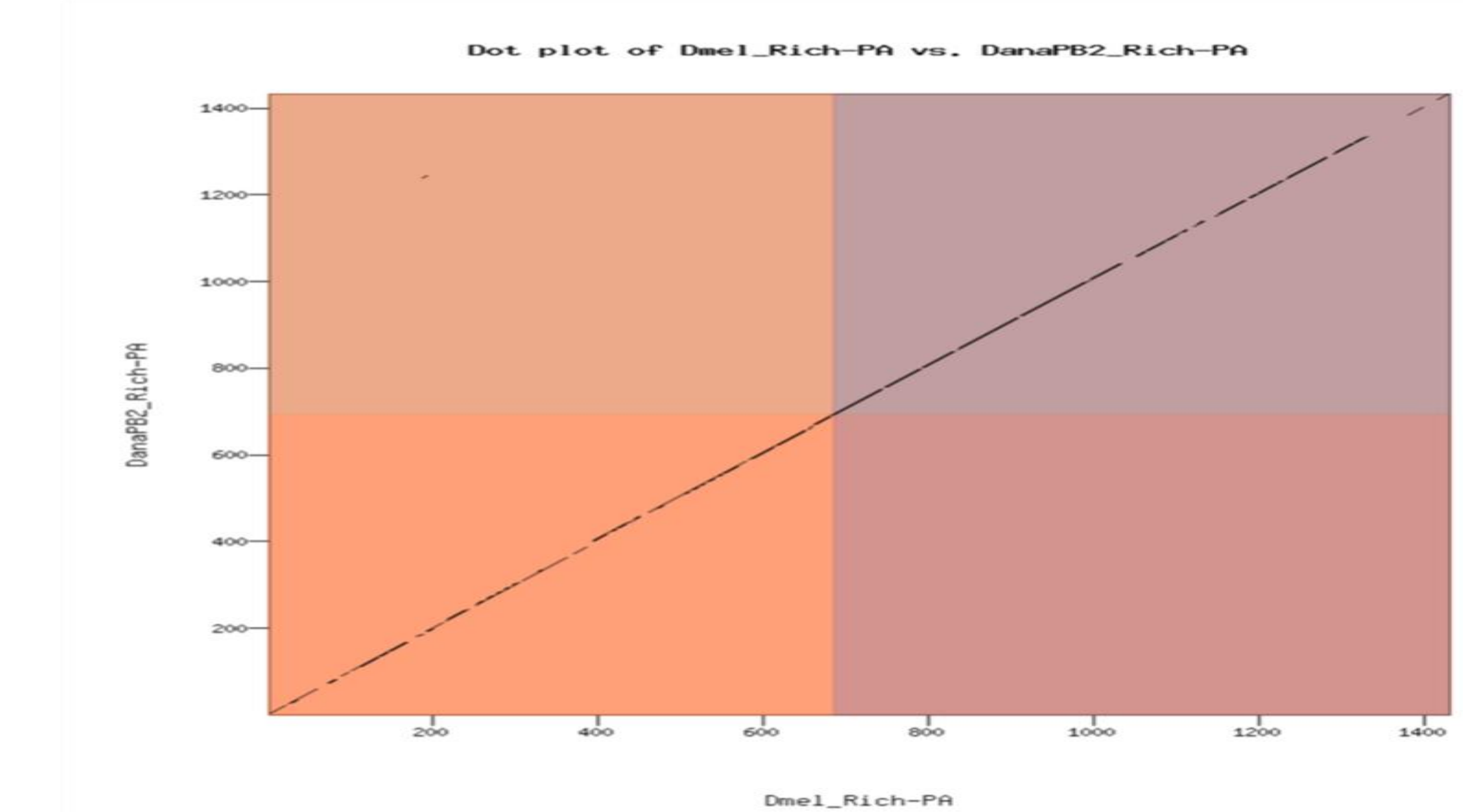
Isoforms with unique coding exons:

Unique isoform(s) based on coding sequence	Other isoforms with identical coding sequences
Rich-PA	Rich-PB

Select a row to display the corresponding CDS sequence:

FlyBase ID	5' Start	3' End	Strand	Phase	Size (aa)
1_2569_0	22,262,678	22,260,624	-	0	685
2_2569_0	22,260,373	22,258,139	-	0	745

Rich has two isoforms and two exons for each isoform



Verified Isoform by inputting coordinates and checking dot plot for positive correlation to the Rich sequence in the *D. melanogaster* genome