

The Complexity of eIF4E1, Its Impact on Ortholog Identification, and Improvements for Student Success with Gene Annotation

Katherine Free, Dr. Jack Vincent

Gene annotations provide information that is important for recognizing how regulatory regions of particular genes evolve within a network of a certain species. The Genomics Education Partnership (GEP) Pathways Project provides undergraduate students with the opportunity to participate in genomics research by annotating genes in *Drosophila* species. The Pathways Project provides a walkthrough to students participating in gene annotations, requiring that the identification of the ortholog be found for successful gene annotation. In *Drosophila melanogaster*, the Eukaryotic Translation Initiation Factor 4E1, eIF4E1, is a crucial component of the genome and has seven homologs. As a Genetics and Genomics student, I experienced extreme difficulty while attempting to annotate eIF4E1 in *D. rhopaloa* due to the multiple homologs this gene possesses. Our research investigated whether annotation of the eIF4E1 gene in three *Drosophila* species, *D. obscura*, *D. navojoa*, *D. takahashii*, posed difficulties in identifying the correct ortholog due to the appearance of multiple homologs based on the Pathways Project walkthrough instructions. By following the Pathways Project walkthrough step-by-step, the *tblastn* results of these three species were examined and recorded in order to determine if the species had a result with the criteria provided in the walkthrough, highest percent identity and lowest E value combined, that identified the correct ortholog. Our findings showed that each *tblastn* result with the highest percent identity and lowest E value for these three species did not yield the ortholog of eIF4E1, but instead one of the seven homologs with different genomic neighborhoods. These results support that the eIF4E1 gene cannot be annotated using the Pathways Project walkthrough as written in *D. obscura*, *D. navojoa*, or *D. takahashii* due to the presence of multiple homologs. We have responded to these findings by compiling a series of suggestions for improvement of the Pathways Project walkthrough for the Genomics Education Partnership.