BIOL199 BDB
Fall 2022

Lab 9 Pre-lab: Pathways Part D: Determine structure of target gene in *D. melanogaster*

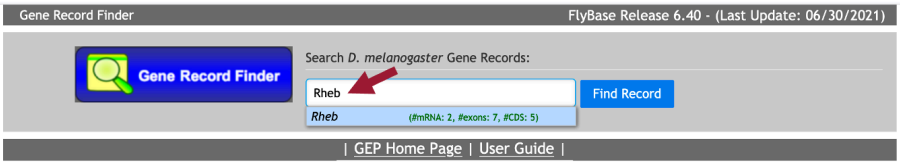
*Adapted by Melinda A. Yang, from "*Pathways Project: Annotation Walkthrough" by *GEP members Katie Sandlin, Wilson Leung, and Laura Reed, and from "Pathways Project Annotation Notebook" by GEP members Katie Sandlin and Alexa Sawa*

The goal of this section is to use the **Gene Record Finder** to better understand the features of the gene we are annotating in *D. melanogaster*. This Finder tool, which you used in the BLAST Lab from Week 6, is a web tool that enables (1) quick identification of the unique set of exons for a given gene and (2) retrieval of their Coding DNA Sequences (CDS's), also referred to as coding exons. The Gene Record Finder will also provide details, such as number of isoforms, exon-intron structure and their coordinates, and transcript and protein information of the gene in question (*Rheb* in examples). It is important to remember that the details provided by the Gene Record Finder are for the gene in the reference species, *D. melanogaster*. We will use the details from *Rheb* in *D. melanogaster* to assist us with creating a gene model for your target gene in your target species.

Instructions

1. Open a new web browser tab and navigate to the Gene Record Finder (http://tiny.cc/generecordfinder).

2. Enter your target gene symbol (case-sensitive) into the text box. Then, click on the 'Find Record' button.



3. Use the description on the next page to understand the results of the Gene Record Finder, and use the available information to answer worksheet questions.

4. Take a screenshot of your Gene Record Finder results:
   - Make sure your screenshot includes the diagram of the gene under 'mRNA details' and the 'CDS usage map' under the 'Polypeptide Details' tab, see Figures 1-2.
   - Save the image into your lab notebook, making sure to indicate your target gene and the words 'Gene Record Finder' near your image.

Comments

Rheb provides an entry in Gene Record Finder, but **rheb** results in the error 'Cannot find the FlyBase gene record: rheb'

Answer question D1.

Fill in columns E1a-E1c with the corresponding information. Note that the CDS column should just be labeled 1,2,3,...

The remaining columns of Table 3 will be completed as part of Week 9's lab

Understanding the Gene Record Finder using the *Rheb* gene

The Gene Record Finder shows that *Rheb* has two isoforms (A and B) in *D. melanogaster* (Figure 1). A graphical overview of the two isoforms is shown in the "mRNA Details" panel. The "CDS usage map" (under the "Polypeptide Details" tab) shows that both isoforms have the same set of coding sequences (CDS's) (i.e., 1_9847_0, 2_9847_2, 3_9847_2, 4_9847_1, and 5_9847_0). (The CDS's are ordered from 5' to 3' (from left to right) in the CDS usage map.) Hence the differences between these two isoforms are limited to the untranslated regions (UTRs).

Based on parsimony (i.e., minimizing the number of changes compared to *D. melanogaster*), we expect to find both the A and B isoforms of *Rheb* in our *D. yakuba* genome sequence. We will only focus on annotating the CDS's, which do not include the UTRs. Consequently, we only need to determine the coordinates of the five CDS's for one of the isoforms (e.g., isoform A) because the set of CDS's for both the A and B isoforms are the same. **If the coding sequences for the two isoforms weren't identical, we'd need to annotate both isoforms. Thus, if you have more than one isoform with unique coding sequences, you will need to repeat everything you do in the Pathways Project from Parts D-G for BOTH unique isoforms.**

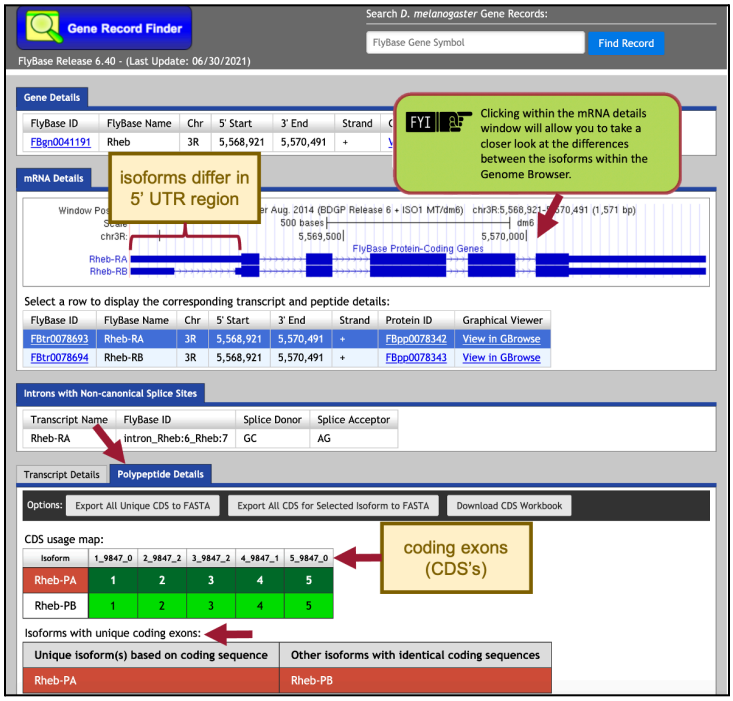


Figure 1 The "mRNA Details" panel of the Gene Record Finder shows that the *Rheb* gene has two isoforms in *D. melanogaster* (i.e., Rheb-RA and Rheb-RB). Under the "Polypeptide Details" tab, the "CDS usage map" indicates that both isoforms have five coding exons (CDS's), and the "Isoforms with unique coding exons" section shows that both isoforms have identical coding sequences.

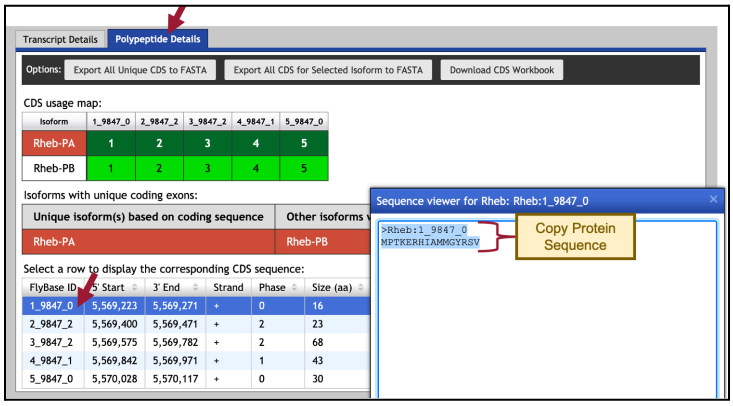


Figure 2 At the bottom of the Gene Record Finder under "Polypeptide Details", a table containing each CDS and information about the CDS are included. Clicking on a row will open a pop-up window with the corresponding amino acid sequence. You will use this section in Part E, but here, note the FlyBase ID column and the size of the coding sequence 'Size (aa) column). Here, 1_9847_0 is the first CDS, its ID is 1_9847_0, and it consists of 16 aa

Complete D1 and E1a-c for each table by the start of this week's lab. This document will eventually be attached to this week's worksheet.

D1. Provide the following information about your target gene:

- # of isoforms total:
- # of isoforms with unique coding sequences:
- List isoforms on the right table (you may not fill out every row), where those with unique coding sequences are on new rows, and those with identical coding sequences as another isoform are on the same row as that isoform but the next column (order does not matter).

Unique Isoform	Identical Isoforms

E1. Fill in the following table based on data from the Gene Record Finder in Part D (pre-lab), and the tblastn searches in Part E. Leave remaining rows blank if you have fewer CDS regions.

TABLE 3: Summary of *tblastn* CDS-by-CDS search results for 1st unique isoform _____

a. CDS	b. FlyBase ID	c. Query Length Size (aa)	d. *D. melanogaster*		e. Target Species		f. Subject Frame
			Query Start	Query End	Subject Start	Subject End	

TABLE 4: Summary of *tblastn* CDS-by-CDS search results for 2nd unique isoform _____

a. CDS	b. FlyBase ID	c. Query Length Size (aa)	d. *D. melanogaster*		e. Target Species		f. Subject Frame
			Query Start	Query End	Subject Start	Subject End	