**CS423 Lab 2: Databases and Genomes**

**Authors: Charles Bonefeole, Caleb Piekstra**

**Due:** Wednesday, Sept 16 [beginning of class]

**Lab Exercises:**

1. Login to a computer and launch a web browser of your choice. Open tabs for the following URLs:

yeastgenome.org

flybase.org

2. (42 points, 2 pts per box) [**Complete the table**] As a pair, choose one yeast gene on the master gene list (located on Moodle). Note the common name, standard name, and systematic name of the gene on your write-up. Also, note its ortholog in the fruit fly. Put the name on the board in the room, so each team chooses a different gene.

Complete the table below as you work your way through the lab.

**Table 1: Information for the gene \_\_\_\_\_\_PFK1\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ in yeast and its orthologs in the fruit fly and the human**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Organism** | **Number of chromo-somes** | **Gene name** | **Gene location (chromo-some #)** | **# of amino acids in protein** | **Upstream gene (left)** | **Downstream gene (right)** | **# of introns in gene** |
| **Yeast** | 16 | PFK1 | VII nuclueotide position: 970,771 through 973,734, C strand | 987 | YGR239C | YGR241C | 0 |
| **Fruit fly** | 8 | Pfk | II, nucleotide position: 10,109,740 through 10,117,457, Right arm | 1044 (Pfk-RE) | 14-3-3Zeta | dmpd | 40 |
| **Human** | 46 | PFKL | 21  Nucleotide position: 44,300,051 through 44,327,376 Forward strand | 780 (PFKL-001) | AIRE | C21orf2 | 21 |

3. (15 points) [**All about yeast**] Complete the questions below for your gene. You should find the information in the yeast database. If you see a word you do not understand, look up its meaning in the textbook or search the web. If you cannot find the information requested and you have looked for 15 minutes, just write that the information was not located (sometimes this information is hard to find). On the yeast database page, click on “Sequence -> Gene/Seq Resources”. In the first box on the left, enter the systematic name for your yeast gene, such as YGR240C. Once your gene is found in the database, click on the “Locus Info” link.

a. What are all the names (standard, systematic, aliases) for your yeast gene?

The names are PFK1, YGR240C, and Pfk respectively.

b. Why might a gene have multiple names?

There could have been something studied before they were systematically named. Those previous names could be maintained.

c. What is the reported function of your gene product?

Our gene product’s reported function is to take part in glycolysis. It is also necessary for anaerobic growth.

d. Where is the gene located in the yeast genome (chromosome number, nucleotide position, and C or W strand)? **[Fill in table.]**

e. How many amino acids does your gene product have? (Go to Protein tab and look for length a. a. (amino acid)) **[Fill in table]**

f. How many chromosomes does yeast have? **[Fill in table]**

g. GBrowse is a tool that displays the linear arrangement of genes on a chromosome. Click on GBrowse on the right of chromosomal location menu on the sequence tab. What are the genes on either side of your gene’s location on the same DNA strand? Make sure the genes are verified (shown in red). You may need to zoom out to see the neighboring genes. **[Fill in table]**

h. Do the adjacent genes have functions related to your gene? (You may need to read the summary pages for the adjacent genes)

Neither of the adjacent genes have functions related to our gene.

i. Why do some genes point left and some point right in GBrowse?

Some genes point left and some point right in GBrowse to indicate if the genes are on opposite strands or not.

j. Find sequences in the annotated sequence features that are not coding for genes. Name some of the sequences that you find.

ARS733, ARS734, ARS375, tT(UGU)G2

k. Go the tab for Interactions on the summary page. What genes (eventually making proteins) have physical interactions with your gene? List 5 of these genes are their functions.

1. ACT1, cell polarization, endocytosis, and other cytoskeletal functions

2. AHA1, helps determine prion variants

3. AIM4, possible association with the nuclear pore complex

4. AKL1, involved in endocytosis and actin cytoskeleton organization

5. ARC18, subunit of the ARP2/3 complex, required for motility and integration of cortical actin patches

4. (11 points) [**Fruit fly and human**] Now, you will compare the genomes of yeast, the fruit fly, and humans. Go to the browser tab containing flybase.org. Search for the fruit fly ortholog of your yeast gene(s) using the ‘jump to gene” search box at the top of the page. In a new tab, open the GBrowse link for your gene.

a. Closely related organisms will have similar gene arrangements on the chromosomes. Regions of the genome that are highly similar between two organisms are said to share synteny. Which two genes are on either side of the same strand of the fly version of your gene? Use GBrowse in FlyBase. **[Fill in table]**

b. What can you conclude about the degree of synteny between the yeast genome and fly genome that surrounds your gene? (Are the genes surrounding your gene in yeast similar to the genes surrounding your gene in the fruit fly?)

The genes surrounding Pfk are not similar between yeast and fruit fly.

c. Is the degree of synteny a surprise? Why or why not?

No, yeast and fruit flies are very far apart in evolution.

d. How many chromosomes does *Drosophila melanogaster* (fruit fly) have? **[Fill in table.]**

e. On which chromosome and on which arm of that chromosome is your fruit fly ortholog located? **[Fill in table.]**

f. Go back to the page that you had prior to opening GBrowse. Scroll down the page for your fruit fly ortholog and click the “+” sign in the blue tab labeled “Orthologs”. Then click on + for Human Orthologs. Choose one of the human orthologs (if there is more than one listed). Open this link in a new tab or window to find the human ortholog of your gene in a new database called ensembl. What chromosome and which strand of that chromosome is the human ortholog of your yeast and fly genes located on? **[Fill in table.]**

g. What is the name of this human ortholog? **[Fill in table.]**

h. How many chromosomes do humans have? **[Fill in table.]**

i. What are the adjacent genes on either side of the same strand of your ortholog on the human chromosome? Click on the region in detail to find them. **[Fill in table.]**

4. (12 points) [**Introns and Multiple Splicing**] You may have noticed in the “mRNA” schematic in GBrowse that some of the fruit fly genes have multiple blocks interrupted by thin lines. This indicates that some sequences in the initial RNA transcript are removed (spliced out) before the protein is translated. The pieces of RNA that are removed from the original transcript are called introns, while the pieces of the original transcript that remain in the final mRNA are called exons. Many of the genes in fruit flies have introns, while only a small proportion of yeast genes have introns (for example, the yeast gene EFB1 has one intron).

Examine the schematics of both your yeast gene and your fruit fly gene and the two verified genes on either side of them (either strand) using GBrowse (you may have to adjust the zoom setting to see five genes simultaneously).

a. Including your gene, how many of the 5 neighboring yeast genes have introns?

None of the 5 neighboring yeast genes have introns.

b. How many of the 5 neighboring fruit fly genes have introns? You may find information about introns in your gene on the Summary page in FlyBase. Click on the drop-down menu for introns and scan the FastA file to see the intron coding regions in order to count the number of introns.

All 5 of the neighboring fruit fly genes have introns.

c. Go back to the yeast genome tab. Look at your gene in GBrowse for any breaks. If there are no breaks, the gene has no introns. How many introns are in your **yeast** gene? **[Fill in table.]**

d. How many introns are in the fly ortholog? **[Fill in table.]**

e. mRNAs are not always spliced in one way; sometimes a gene has multiple possible translations of configurations of exons (alternative splicing can produce related versions of a protein called isoforms). Different isoforms can have different functions. How many isoforms can your fruit fly ortholog encode? (Look for different transcripts in GBrowse.)

Our fruit fly ortholog can encode 5 different isoforms.

f. How many isoforms can the human ortholog encode? (known isoforms are listed as transcripts under “about this gene” on the gene page. You can view the transcript table for more details about the isoforms.)

The human ortholog can encode 12 isoforms.

g. Using the first isoform listed for the fruit fly, record the number of amino acids. **[Fill in table.]**

h. Using the first isoform listed for the human (under transcripts), record the number of amino acids. **[Fill in table.]**

i. How many introns does the first isoform listed in the human have? Click on the transcript ID and scroll down the page to see the number of exons. That should help you determine the number of introns (or you can count the thin lines between the bigger boxes in the schematic). **[Fill in table.]**

j. The number of introns and exons in genes is generally correlated with genome size and complexity. How do the numbers of introns in the fly ortholog compare to the number of introns in the human ortholog?

The fly ortholog has nearly twice the number of introns than the human ortholog.

k. What does this suggest about the relative genome complexity of fruit flies and humans?

This suggests that fruit flies have a relatively more complex genome than humans. However, we noticed that since the human ortholog has 12 isoforms and yet the fruit fly ortholog has only 5, it would make sense that the human genome is still far more complex.

5. (5 points) [**Conservation**] Is it surprising that your yeast gene is conserved in fruit flies and humans? Based on what you have learned about your yeast gene’s function, why or why not?

No, it is a gene that is important for glycolysis which makes ATP and is found in pretty much all living things due to the fact that it is a very common way to make energy in any organism.