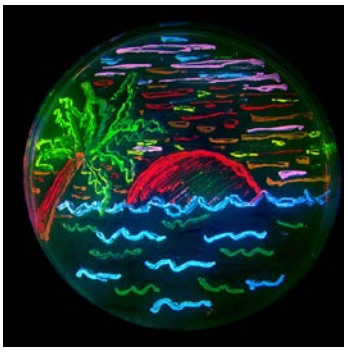
Comparing Sequences of Fluorescent Proteins Using Basic Local Alignment Search Tool (BLAST)

Mice expressing GFP under UV light (left & right), compared to normal mouse (center). Source: Wikipedia.

Researcher Background:
Fluorescent proteins have become a valuable tool in recent years among scientists in many different fields of biology. Often, these glowing proteins are linked to other proteins to identify where specific proteins exist in the cell, to track cell migrations, or to confirm the success of genetic modifications when fluorescent proteins are used as a ***reporter of expression*** (see image at left).

Green fluorescent protein (GFP) is comprised of 238 amino acids that fold into a barrel-shaped protein that exhibits bright green fluorescence when exposed to light in the blue or ultraviolet range, and has a major excitation peak at a wavelength of 395 nm (1, 2). The protein was first isolated from the jellyfish *Aequorea victoria*.



The diversity of genetic mutations is illustrated by this San Diego beach scene drawn with living bacteria expressing 8 different colors of fluorescent proteins. Source: Wikipedia.

Intentional mutation of the *gfp* gene has led to a **rainbow of fluorescent proteins**, dramatically increasing their utility in scientific research (see image at right). This diversity of colors among fluorescent proteins has sometimes been referred to as the "mFruits," referring to the names given to these fluorescent proteins, such as:

- mBlueberry (Blue Fluorescent Protein, or BFP)
- mLemon (Yellow Fluorescent Protein, or YFP)
- mGrape1 (Cyan Fluorescent Protein, or CFP)
- and many others, all with similarly 'fruity' names…

While GFP remains the most "well known" fluorescent protein, there are limitations to the colors that can be generated by mutating the *gfp* gene. Other sources of fluorescent proteins have been sought – and found – resulting in even greater diversity of fluorescent proteins.

Research Questions: The cloning and protein purification experiments you have been conducting in the laboratory involve mTomato (related to mCherry), also called ***red fluorescent protein (RFP).***

(1) Is red fluorescent protein (RFP) related to its famous cousin, GFP, or is from a different source entirely?
(2) What other fluorescent proteins, if any, are closely related to RFP?

Updated June 19, 2013 by NWABR

Instructions: After learning how to detect mutations in the *BRCA1* gene and the BRCA1 protein using the Student Handout: "Instructions for Aligning Sequences with BLAST,"[1] you will use the same skills to compare various fluorescent proteins. These comparisons will help you better understand the origin and diversity of fluorescent proteins used in biological research.

PART I: Aligning DNA Sequences

1. Obtain your copy of the *Lesson Four* Student Handout: "Instructions for Aligning Sequences with BLAST."

2. Open a new web browser tab or window and open the file, "DNA Sequences from Various Fluorescent Proteins" from the page below:
http://www.nwabr.org/teacher-center/introductory-bioinformatics-genetic-testing#resources

3. Perform a nucleotide BLAST alignment as explained in the Student Handout, "Instructions for Aligning Sequences with BLAST," Steps 1-13. Use "Euk-Green-Fluorescent-Protein-eGFP" as the **reference sequence** (i.e., your **Query sequence**, as described in Step 6) and "pLemon-Yellow-Fluorescent-Protein-YFP" as the **subject sequence** (similar to the family member's sequence in the BRCA1 analysis, as described in Step 7).

Do these two sequences appear to be closely related to one another? Why or why not? Your answer should include the "Query Coverage" and "Max Identity" values obtained from your BLAST alignment.

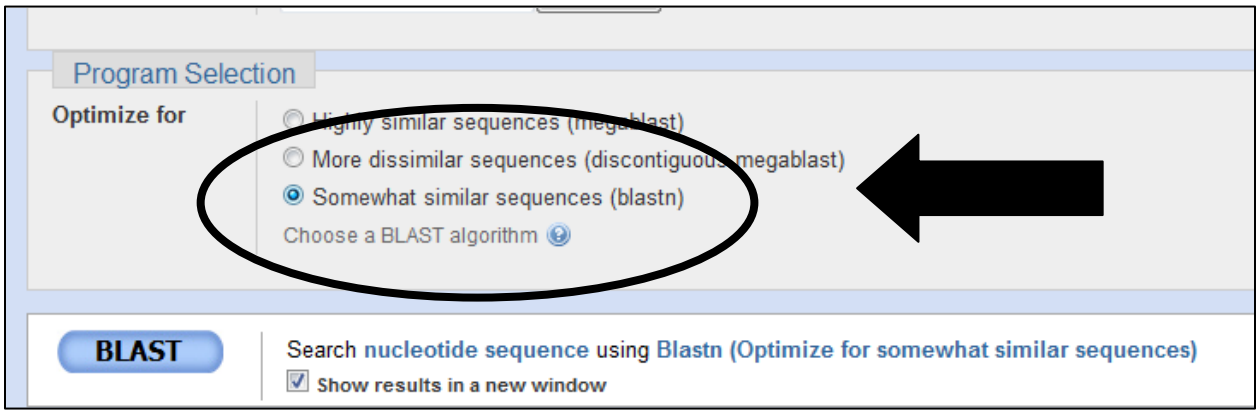
4. Perform another nucleotide BLAST alignment as explained in the Student Handout, "Instructions for Aligning Sequences with BLAST," Steps 1-13. Use ">mTomato-Modified-Monomer-red-fluorescent-protein-RFP" as the **reference sequence** (i.e., your **Query sequence**) and "mGrape1-Cyan-Fluorescent-Protein-CFP" as the **subject sequence**.

Do these two sequences appear to be closely related to one another? Why or why not? Your answer should include the "Query Coverage" and "Max Identity" values obtained from your BLAST alignment.

[1] Found in NWABR's introductory bioinformatics curriculum, "Using Bioinformatics: Genetic Testing," *Lesson Four: Understanding Genetic Tests to Detect* BRCA1 *Mutations,*" available at http://www.nwabr.org/teacher-center/introductory-bioinformatics-genetic-testing#lessons

5. Perform another nucleotide BLAST alignment as explained in the Student Handout, "Instructions for Aligning Sequences with BLAST," Steps 1-13. Use "Euk-Green-Fluorescent-Protein-eGFP" as the **reference sequence** (i.e., your **Query sequence**) and ">mTomato-Modified-Monomer-red-fluorescent-protein-RFP" as the **subject sequence**.

SPECIAL NOTE: BEFORE clicking the blue "BLAST" button, choose "Somewhat similar sequences (blastn) from the "Program Selection" menu.



Do these two sequences appear to be closely related to one another? Why or why not? Your answer should include the "Query Coverage" and "Max Identity" values obtained from your BLAST alignment. You may wish you compare your answer to this question with your answers to Questions 3 and 4.

PART II: Aligning Protein Sequences

6. Perform a protein BLAST alignment as explained in the Student Handout, "Instructions for Aligning Sequences with BLAST," Steps 29-42. Use "Euk-Green-Fluorescent-Protein-eGFP" as the **reference sequence** (i.e., your **Query sequence**, as described in Step 29) and "pLemon-YFP" as the **subject sequence** . [Note: You are using only one subject sequence here, not six as in the BRCA1 analysis.]

Do these two sequences appear to be closely related to one another? Why or why not? Your answer should include the "Query Coverage" and "Max Identity" values obtained from your BLAST alignment.

7. Are your results similar to your nucleotide comparison of these two sequences in Question 3? Explain how they are or are not similar.

8. Perform a protein BLAST alignment as explained in the Student Handout, "Instructions for Aligning Sequences with BLAST," Steps 29-42. Use ">mTomato-Modified-Monomer-RFP" as the **reference sequence** (i.e., your **Query sequence**) and "mGrape1- CFP" as the **subject sequence**. [Note: You are using only one subject sequence here, not six as in the BRCA1 analysis.]

Do these two sequences appear to be closely related to one another? Why or why not? Your answer should include the "Query Coverage" and "Max Identity" values obtained from your BLAST alignment.

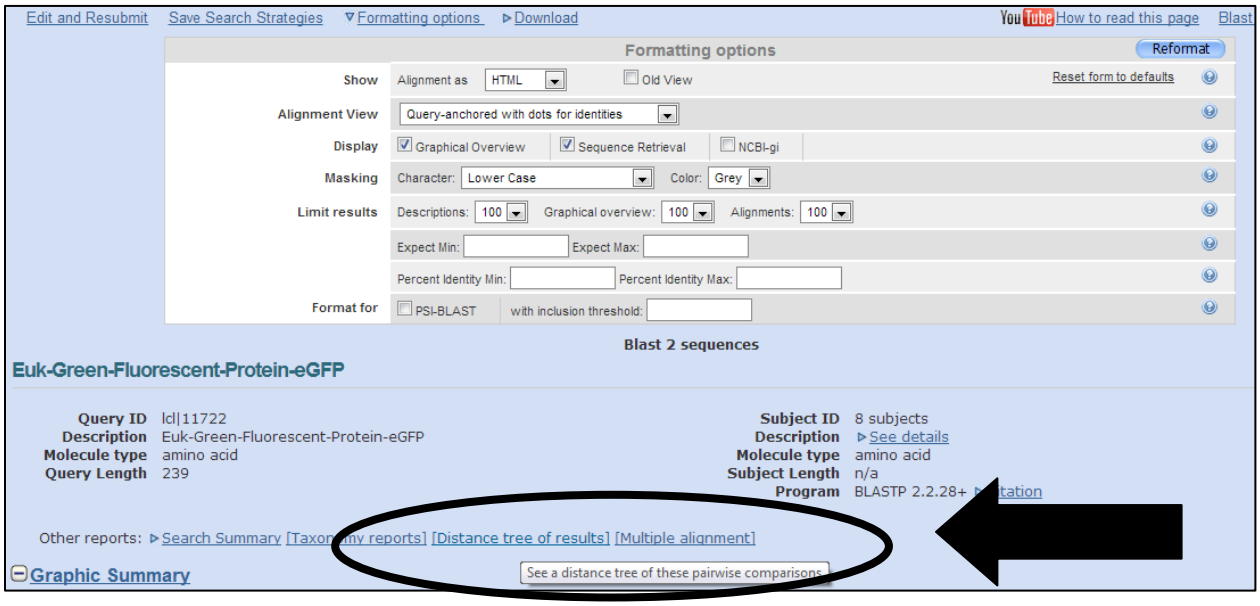
9. Are your results similar to your nucleotide comparison of these two sequences in Question 4? Explain how they are or are not similar.

10. Perform a protein BLAST alignment as explained in the Student Handout, "Instructions for Aligning Sequences with BLAST," Steps 29-42. Use "Euk-Green-Fluorescent-Protein-eGFP" as the **reference sequence** (i.e., your **Query sequence**) and all of the remaining protein sequences as the **subject sequences** [as explained in Step 35]: pLemon-YFP, mTomato-RFP, mGrape-CFP, pLime-GFP, pBlueberry-BFP, mTangerine1.5, mCherry-RFP, mOrange.

Based on the Query coverage and Max identify values you obtained, which fluorescent proteins appear to be most closely related to eGFP?

11. Based on the Query coverage and Max identify values you obtained, which fluorescent proteins appear to be most closely related to RFP?

12. Another way to visualize your results is by using a **phylogenetic tree**. BLAST can perform this analysis for you using the comparisons that you have already performed. Click "Distance Tree of Results" to view your tree.



Draw a rough sketch of your tree in the space below.

13. Does the phylogenetic tree support the conclusions that you made in Questions 10 and 11 about the relatedness of various fluorescent proteins? Why or why not?

References:

1. Prendergast F, Mann K (1978). "Chemical and physical properties of aequorin and the green fluorescent protein isolated from Aequorea forskålea". *Biochemistry* 17 (17): 3448–53. doi:10.1021/bi00610a004. PMID 28749.
2. Tsien R (1998). "The green fluorescent protein" (PDF). *Annu Rev Biochem* 67: 509–44. doi:10.1146/annurev.biochem.67.1.509. PMID 9759496