

Mystery Gene Identification

Scenario:

You have just been hired as an assistant scientist in a molecular biology lab at UW. While this lab is well-known for its ground-breaking research in genetic diseases, you quickly learn that they are not acclaimed for their organizational skills. It seems that graduate students have been storing tubes of DNA in the freezers without taking the time to carefully label and archive the contents. There are some very important samples of DNA that have been studied over the years, so this mess has to be organized and the samples salvaged. Your first project is to help the lab manager identify the contents of the tubes and write a short report describing what you learn.

You receive 3 tubes labeled:

“Gene 1 - ready to sequence”

“Gene 2 – ready to sequence”

“Gene 3 – ready to sequence”

Not very helpful, except that you know that these samples have been prepared for DNA sequencing. You send off small volumes of each to the Sequencing Core Facility, and then run to your computer to review what you know about sequencing technology. You find the following link and spend some time exploring it:

<http://www.dnalc.org/ddnalc/resources/cycseq.html>

The sequencing lab is quick to process your samples, because it turns out that the DNA fragments are fairly short – only representing small parts of genes. They provide you with an electropherogram and now you have to identify your genes!

- 1. Translate the colored peaks into nucleotides using the key on the electropherogram.** In an electropherogram of a DNA sequencing reaction, each nucleotide of your gene sequence is represented by a colored peak. Create a data table to record the nucleotide sequence of your unknown genes in your lab book.
- 2. Once you have your gene sequence, use a tool called BLAST to identify your genes.** Open your web browser to www.ncbi.nlm.nih.gov. NCBI (National Center for Biotechnology Information) is a website that houses public databases and software tools used to search for and analyze biological data. The **Basic Local Alignment Search Tool (BLAST)** is a sequence similarity search program that can be used to compare your nucleotide sequence of interest to sequences in the NCBI nucleotide databases (filled with DNA sequences contributed by many scientists).
 - Click on the BLAST tab in the menu on the right side of the NCBI homepage.
 - Select the option for “nucleotide blast.”
 - Type one of the gene sequences from left to right in the space under “Enter Query Sequence.” *Be careful – one mistake can drastically change your results.*
 - Check that the Database being queried is “human genomic plus transcript (Human G + T) and the Program is optimized for “Highly similar sequences (megablast).”
 - Click on the BLAST tab and **wait**.
 - Scroll down to get the sequences producing alignments. Click on the accession number (beginning with NM) of your highest hit (this should be the first transcript).
 - Read the information on the resulting page. Create a data table to record:
 - i. the name of the gene
 - ii. the species from which it was sequenced
 - iii. length of the entire gene in base pairs (bp)
 - iv. a disease associated with mutations in the gene
 - Repeat for each of the unknown genes.

