**BLAST That Gene**

**Objective:** In this activity you will use the website for the National Center for Biotechnology Information to identify a sequence of bases from a DNA sample.

Background: The NCBI contains a database of genes sequenced and identified. The work of a number of scientists and a wide variety of areas allows for information to be compiled into this database. The tool used in this activity is BLAST - Basic Logical Alignment Search Tool, which allows a user to submit a sequence of DNA bases (A, T, G, C) into the search engine, BLAST will search the database and find the gene (if any) associated with that sequence. Furthermore, the user can access information about the gene or the disease it causes.

**Instructions**

You will be given nucleotide sequences found in real human DNA that is associated with a genetic disease when mutated. Your job is to compare the sequences you are given with the nucleotide sequence of most known genes, using the BLAST tool to search genetic databases. Open the gene file on the class website to find the gene sequences, this will make it easy to copy and paste the sequences into NCBI.

1. Go to the BLAST homepage for NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)
2. Click on "Nucleotide BLAST".
3. Copy and paste one of the nucleotide sequences from the gene file into the query box. Then click the big blue “BLAST” button. The program will run through all the known sequences and find a match. Depending on how busy the server is, this may take a while.
4. The top search hits will show with a short description. For more information on your sequence, click the link for GENE. This will give you the description of the gene that your sequence matched.

**Analysis**: In your notebook, identify each of the genes. Write a brief summary that describes the gene and its importance to human and biological studies. Include the % query coverage, E-value and % max identity for the top hit returned from each BLAST. See the table below as a reference for how to record this data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Sequence | Gene | % Query coverage | E-value | % Max identity | Description |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |
| 7 |  |  |  |  |  |
| 8 |  |  |  |  |  |

Look at the BLAST tutorial (or look at the glossary section in the BLAST website at <http://www.ncbi.nlm.nih.gov/books/NBK62051/>) and fill in these definitions:

E-value:

 % Max Identity: