Bioinformatics Worksheet

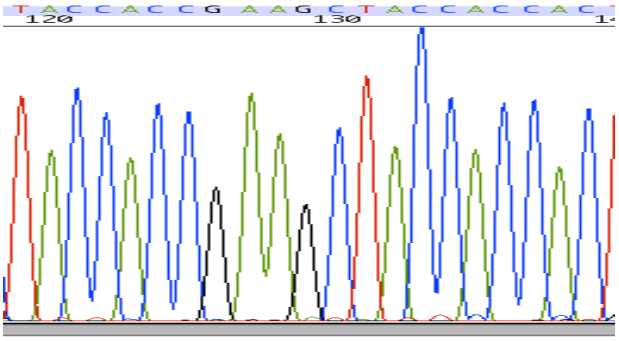
You have received a file of DNA sequence that corresponds to a gene. Today, you will follow this worksheet to identify the gene and the organism that the gene comes from. You will determine if other organisms have genes similar to the gene you identify.

1. Read your sequence

When you open the sequence file, you will see a series of different colored peaks. Each peak corresponds to the color of fluorescence attached to a DNA nucleotide -- Adenine (A) reads green, Cytosine (C) reads blue, Guanine (G) reads black, and thymine (T) reads red. You can read the DNA sequence by the color of the peak.

A. Read and write down the first 50 base pairs of the gene you received. Follow the example below:

Your sequence data:



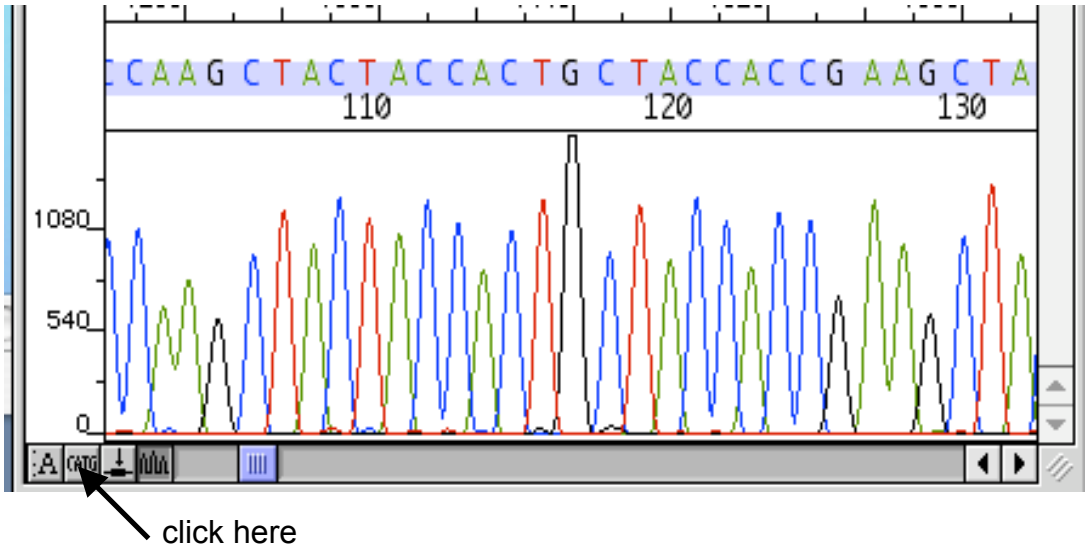
You write: T A CC A C CG A A G C TA C CAC C A C

Enter the first 50 nucleotides from the sequence of your gene here:

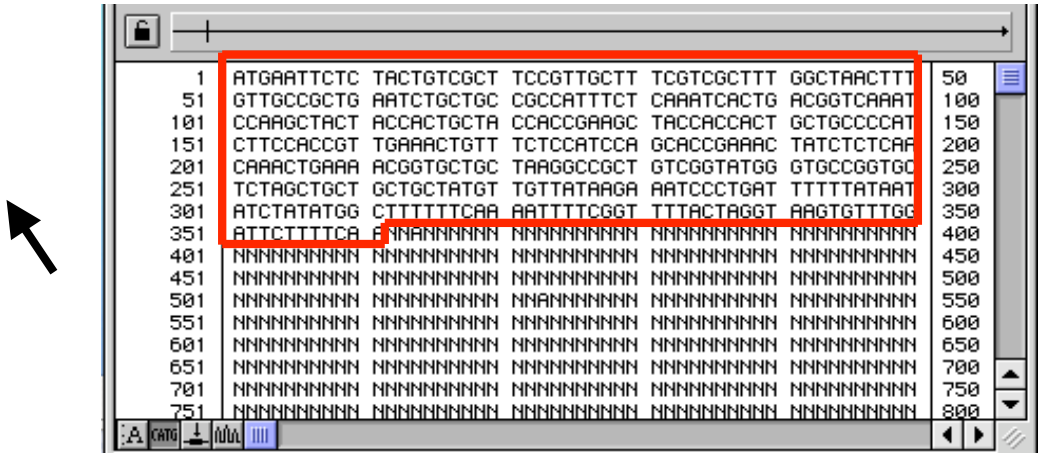
B. Above each band, the machine determines the band's identity. Did you find any places where you think the machine made a mistake? If so, list them below.

2. Retrieve your sequence

A. Click on the "CATG" box in the bottom left corner of your sequence file. Now you see the gene sequence as the sequencing software has determined it.



B. Notice the N's at the end of the sequence. This is sequence that the machine could not read. Copy the good sequence (see red box below). Leave the N's behind.



3. Search the database for your gene

Now, you will use a computer program called **BLAST** to compare your gene to all sequenced genes in all organisms.

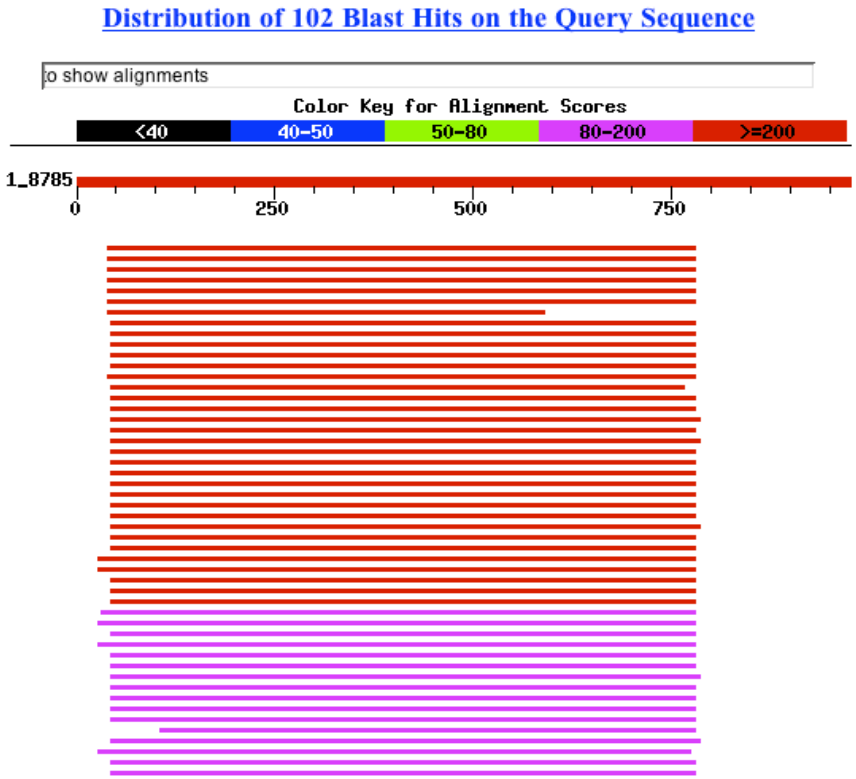
A. Go to the website:
http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semiauto&PAGE=Translations&NCBI_GI=yes&FILTER=L&HITLIST_SIZE=100&SHOW_OVERVIEW=yes&AUTO_FORMAT=yes&SHOW_LINKOUT=yes

You will see an empty box with the words Search to the left. Paste your sequence into the box. Hit the button 

This will take you to a new page. Hit the 

button. DO NOT hit it again. If you do, you will back up the system. You may nave to wait a few minutes while the search is being performed.

When the search is finished, you will get a page that looks similar to this:



This is a pictorial representation of your sequence aligned with other sequences. Not very useful. The useful stuff is below this.

B. You will now learn about your gene from the sequence comparison output. The output is complicated, but you can read it. Below is the output from another gene. Use this example to answer these questions.

Sequences producing significant alignments:	Score (bits)	E Value	
gi|6325264|ref|NP_015332.1| Meoisis-specific component of s...	1168	0.0	G
gi|45190314|ref|NP_984568.1| AEL292Wp [Eremothecium gossypi...	164	8e-39	G
gi|50302391|ref|XP_451130.1| unnamed protein product [Kluyv...	164	8e-39	G
gi|50293537|ref|XP_449180.1| unnamed protein product [Candi...	162	3e-38	G
gi|50420469|ref|XP_458771.1| unnamed protein product [Debar...	72	7e-11	G
gi|42548594|gb|EAA71437.1| hypothetical protein FG08576.1 [...	56	4e-06	G
gi|46435129|gb|EAK94518.1| potential nuclear cohesin comple...	50	2e-04	
gi|46435091|gb|EAK94481.1| potential nuclear cohesin comple...	50	2e-04	
gi|40741374|gb|EAA60564.1| hypothetical protein AN8771.2 [A...	45	0.007	G
gi|6996291|emb|CAB75452.1| putative protein [Arabidopsis th...	44	0.017	G
gi|12006362|gb|AAG44843.1| cohesion family protein SYN3 [Ar...	44	0.017	
gi|7769760|gb|AAF69524.1| meiotic cohesin REC8 [Mus musculus]	42	0.083	G
gi|46441209|gb|EAL00508.1| potential nuclear cohesin comple...	41	0.11	
gi|38100299|gb|EAA47445.1| hypothetical protein MG02688.4 [...	41	0.11	G
gi|39583109|emb|CAE60649.1| Hypothetical protein CBG04295 [...	40	0.24	
gi|10178126|dbj|BAB11538.1| SYN1 splice variant 1 [Arabidop...	40	0.31	G
gi|6453717|gb|AAF08982.1| SYN1 splice variant 2 [Arabidopsi...	40	0.31	G
gi|32420365|ref|XP_330626.1| predicted protein [Neurospora ...	39	0.41	G
gi|28828121|gb|AAO50804.1| hypothetical protein [Dictyostel...	39	0.41	
gi|31982699|ref|NP_064386.2| REC8-like 1 [Mus musculus] >gi...	39	0.54	G
gi|23619064|ref|NP_705026.1| hypothetical protein [Plasmodi...	39	0.70	G
gi|14790110|gb|AAH10887.1| REC8L1 protein [Homo sapiens] >g...	39	0.70	G
gi|13278774|gb|AAH04159.1| REC8-like 1 [Homo sapiens]	39	0.70	G
gi|52545743|emb|CAH56339.1| hypothetical protein [Homo sapi...	39	0.70	G
gi|17532617|ref|NP_494836.1| COHesin, yeast mitotic condens...	38	0.91	G
gi|9845293|ref|NP_005123.1| REC8-like 1 [Homo sapiens] >gi|...	38	1.2	G
gi|19112851|ref|NP_596059.1| cohesin [Schizosaccharomyces p...	37	2.7	G
gi|34875067|ref|XP_224192.2| similar to Rec8p, a meiotic re...	37	2.7	G
gi|8392845|ref|NP_058541.1| proacrosin binding protein [Mus...	37	2.7	G

Each **row** represents a different DNA sequence in the database that BLAST has decided is similar to your gene. They go in order – best matches are at the top of the list, worst matches are at the bottom. The row has several components. The blue link to the left takes you to information about the gene and the organism it comes from. The black text lists information about the gene and/or the organism. The blue number is a "score" that BLAST uses to decide how similar your gene is to the one that BLAST found. The next column, labeled "E Value", is another such score.

QUESTIONS

1. Circle which gene you were given: Gene1 Gene2

2. Look at the "E Value" for the first row on your list.

A. What is it?

B. LOWER "E Values" indicated BETTER matches. Is the first item on your list a good match or a bad match?

3. Based on your answer to the above question, what organism do you think your gene came from?

4. Read down the rows of matches. Are genes similar to yours found in many organisms? Is a similar gene found in humans?

5. What is the name of the protein that your gene codes? (This consists of 3 letters, followed by a number, followed by the letter "p," for protein. It is written in black text).

6. Now you will identify the your gene's function. Go to the website: http://www.yeastgenome.org/. Near the top left of the page, you will see:

Enter the name of your protein in the box. Remove the "p" from the end of the name. Hit submit. What is the function of your protein?

7. Think about the function of your gene. Why is it or why isn't it found in many other organisms?