

Comparing Influenza Protein Sequences Using BLAST (Basic Local Alignment Search Tool)

Aim:

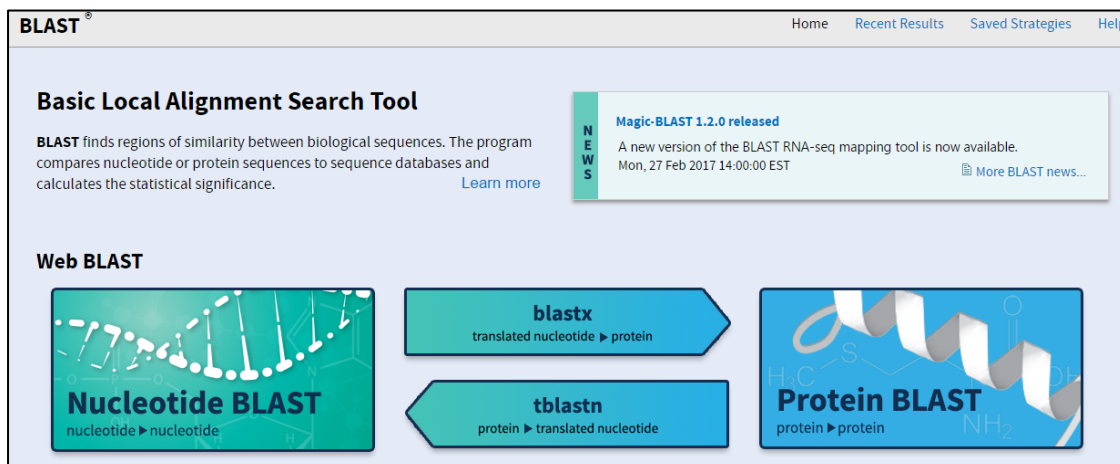
To understand how genetic tests are performed by comparing the DNA and protein sequences of a family to determine who is carrying the *BRCA1* mutation.

Background on *BRCA1*:

For cells to function properly, they need to be able to repair errors in their DNA. These errors can arise when DNA is being copied, or when DNA somehow becomes damaged when exposed to chemicals or radiation. The breast cancer susceptibility gene (*BRCA1*) encodes a protein that is involved in DNA repair. When a DNA strand is broken, the *BRCA1* protein works with other proteins to help repair the break. If these breaks are not repaired, the DNA damage can ultimately lead to cancer. Therefore, *BRCA1* is known as a tumor suppressor, because it helps prevent the formation of tumors (which can arise when DNA errors go unrepaired). Mutations to the *BRCA1* gene can interfere with or abolish the *BRCA1* protein's normal function, thus allowing cancer to develop.

Background in BLAST:

One tool in the bioinformatics is called **BLAST – Basic Local Alignment Search Tool**. BLAST can be used to look at differences in the sequences of two or more proteins (**Protein BLAST**) or nucleic acid molecules (**Nucleotide BLAST**). BLAST, begins with a **Query Sequence**. This is also called the **reference sequence** when performing medical tests such as *BRCA1* genetic testing. The Query Sequence is the sequence that you are going to use to relate to or compare to other sequences. You can also compare a single Query Sequence to a collection of sequences in a database at the National Center for Biotechnology Information (NCBI). All of the *results* from your BLAST are called **Subject Sequences**. You use this when you are trying to find or identify a sequence, such as when doing DNA barcoding or finding contamination in a DNA sequencing reaction. The results of BLAST are in the form of an **alignment** to find regions that are the same between your Query Sequence and your Subject Sequence or sequences.



In this experiment, you will do a **Nucleotide BLAST** and a **Protein BLAST** with the BRCA1 sequences from the Lawler family members from the Case Study that you read, as well as a **Protein BLAST** using the BRCA1 protein sequence(s) from the patient(s) that you are studying.

You will use the **BRCA1 Reference Sequence**, which is known to have NO cancer-causing mutations, as your **Query Sequence**. The other sequences from Lawler family members and patients will be your **Subject Sequences**.

Research Questions:

1. Based on your BLAST results, which member or members of the Lawler family have an increased risk of breast cancer due to mutations in the *BRCA1* gene?
2. Based on your BLAST results, which member or members of the Lawler family have an increased risk of breast cancer due to mutations in the *BRCA1* gene?

Part I Procedure: Analyzing Lawler Family *BRCA1* DNA Sequences

1. Download the “**BRCA1 Sequences**” file provided by your teacher.
2. Go to BLAST, either using your search engine or the URL: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. Because we are comparing DNA sequences, click “**Nucleotide BLAST**” on the BLAST homepage, shown in the picture on page 1.
4. Note: The default view in BLAST has only one sequence box, in which you enter your Query Sequence. This is because BLAST is most frequently used to compare a Query Sequence to all of the sequences in the NCBI databases for sequence identification.
5. To make the “**Subject Sequences**” box display, click on the box “**Align two or more sequences.**”

The screenshot shows the NCBI BLAST homepage. The 'Standard Nucleotide BLAST' tab is selected. In the 'Enter Query Sequence' section, there is a checkbox labeled 'Align two or more sequences' which is highlighted with a red box. A blue arrow points from this checkbox to the 'Job Title' field. Below the 'Enter Query Sequence' section is the 'Choose Search Set' section, which includes options for 'Database' (Standard databases, rRNA/ITS databases, Genomic + transcript databases, Betacoronavirus) and 'Organism' (Optional).

6. Copy and paste your Query Sequence, “>BRCA1_Reference_DNA_Sequence” from the sequence Word document into the “**Enter Query Sequence**” box. Be sure to include the top line of text in the sequence, “>BRCA1_Reference_DNA_Sequence” when you paste the text into the “**Enter Query Sequence**” text box.

- Copy and paste the DNA sequence from the Lawler family member that you have been assigned into the “**Enter Subject Sequence**” box, making sure to include the “>” caret and the sequence names when you copy and paste.
- At the bottom of the page, click the “**Show Results in a new window**” and then click “**BLAST**.”

- When your results appear in a new window, you will see a summary of your work, called the “Descriptions” tab in the lower, menu, on the left side (the first tab of that menu). The sequence that is your Query Sequence is listed as your “**Job Title**” in the top, left list. There are also additional details that you don’t need to worry about for this analysis, but one useful thing below the term “**Job Title**,” you will see next to the term “**Subject Descr**” the name of the Subject Sequence that you entered on the previous page. If you enter more than one subject Sequence, as you will below, you will see a drop-down menu that you can click to see the names of *all* of the names of the Subject Sequences that you included in your BLAST analysis [just in case you want to make sure that you included everything that you need to include].

- If you scroll down further, you will see a summary of the **BLAST Scores** for each of your sequences. We’ll discuss BLAST scores later, but in general these are ways to *quantitatively* measure how “good” the match is between your Query Sequence and each Subject Sequence. See the example below.

Subject Descr [See details](#) ▼

Subject 2807

Length

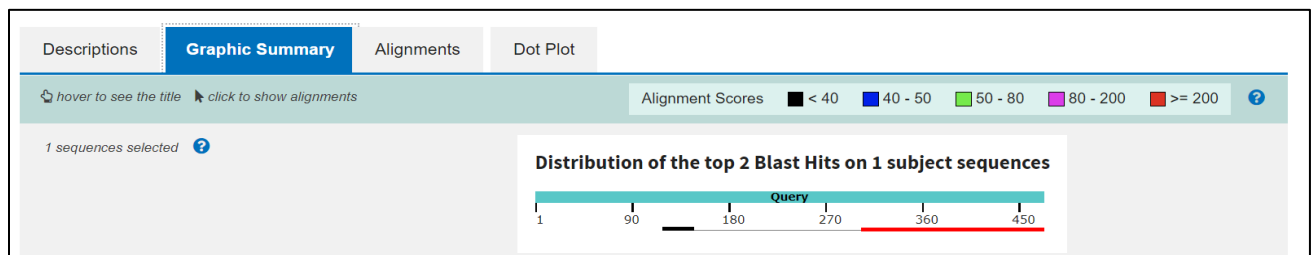
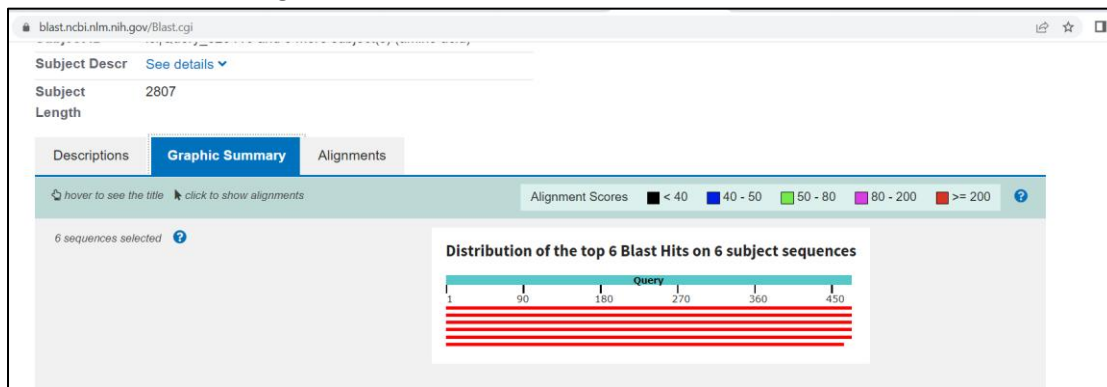
Descriptions Graphic Summary Alignments

Sequences producing significant alignments Download ▼ Select columns ▼ Show 100 ?

☒ select all 6 sequences selected [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Patient A17		956	956	100%	0.0	99.79%	469	Query_325421
<input checked="" type="checkbox"/> Patient A13		956	956	100%	0.0	99.79%	469	Query_325420
<input checked="" type="checkbox"/> Patient A12		956	956	100%	0.0	99.79%	469	Query_325419
<input checked="" type="checkbox"/> Patient A11		956	956	100%	0.0	99.79%	469	Query_325418
<input checked="" type="checkbox"/> A/California/07/2009(H1N1)-2011-2016 Vaccine		941	941	100%	0.0	97.01%	469	Query_325416
<input checked="" type="checkbox"/> A/Brisbane/59/2007(H1N1)-2008-2010 Vaccine		773	773	98%	0.0	80.74%	462	Query_325417

11. Click on the second tab, “Graphic Summary.” You will see an image like the one below. The color refers to how many nucleotides or amino acids of the Query Sequence and each Subject Sequence match (i.e., are the same). In the top image, we see **red lines** – this means that each of these sequences match one another by ≥ 200 nucleotides or amino acids. In the second image below, we see that ≥ 200 amino acids of the Query Sequence and Subject Sequence match near the C-terminus [seen in **red**], and ≤ 40 only of the amino acids match near the N-terminus, as seen in **black**. We also get to see where these matches are, or are not



12. Next, click the “Alignments” tab. In this case, you will see which individual nucleotides (or amino acids) match between the Query Sequence and the Subject Sequence of Sequences. See the image below. The Default settings in BLAST show you each alignment between the Query Sequence and each Subject Sequence, showing all of the one-letter amino acid abbreviations. Some people find this format difficult to analyze: For example, where exactly are the differences? Are there any differences between the various Subject Sequences (if you included more than one Subject Sequence in your analysis, as we did here)?

Descriptions Graphic Summary **Alignments**

Alignment view: Pairwise [Restore defaults](#)

6 sequences selected

[Download](#) [Graphics](#)

Patient A17
Sequence ID: Query_325421 Length: 469 Number of Matches: 1

Range 1: 1 to 469 [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
956 bits(2470)	0.0	Compositional matrix adjust.	469/469(100%)	469/469(100%)	0/469(0%)

Query 1: 1 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000 2100 2200 2300 2400 2500 2600 2700 2800 2900 3000 3100 3200 3300 3400 3500 3600 3700 3800 3900 4000 4100 4200 4300 4400 4500 4600 4700 4800 4900 5000 5100 5200 5300 5400 5500 5600 5700 5800 5900 6000 6100 6200 6300 6400 6500 6600 6700 6800 6900 7000 7100 7200 7300 7400 7500 7600 7700 7800 7900 8000 8100 8200 8300 8400 8500 8600 8700 8800 8900 9000 9100 9200 9300 9400 9500 9600 9700 9800 9900 10000 10100 10200 10300 10400 10500 10600 10700 10800 10900 11000 11100 11200 11300 11400 11500 11600 11700 11800 11900 12000 12100 12200 12300 12400 12500 12600 12700 12800 12900 13000 13100 13200 13300 13400 13500 13600 13700 13800 13900 14000 14100 14200 14300 14400 14500 14600 14700 14800 14900 15000 15100 15200 15300 15400 15500 15600 15700 15800 15900 16000 16100 16200 16300 16400 16500 16600 16700 16800 16900 17000 17100 17200 17300 17400 17500 17600 17700 17800 17900 18000 18100 18200 18300 18400 18500 18600 18700 18800 18900 19000 19100 19200 19300 19400 19500 19600 19700 19800 19900 20000 20100 20200 20300 20400 20500 20600 20700 20800 20900 21000 21100 21200 21300 21400 21500 21600 21700 21800 21900 22000 22100 22200 22300 22400 22500 22600 22700 22800 22900 23000 23100 23200 23300 23400 23500 23600 23700 23800 23900 24000 24100 24200 24300 24400 24500 24600 24700 24800 24900 25000 25100 25200 25300 25400 25500 25600 25700 25800 25900 26000 26100 26200 26300 26400 26500 26600 26700 26800 26900 27000 27100 27200 27300 27400 27500 27600 27700 27800 27900 28000 28100 28200 28300 28400 28500 28600 28700 28800 28900 29000 29100 29200 29300 29400 29500 29600 29700 29800 29900 30000 30100 30200 30300 30400 30500 30600 30700 30800 30900 31000 31100 31200 31300 31400 31500 31600 31700 31800 31900 32000 32100 32200 32300 32400 32500 32600 32700 32800 32900 33000 33100 33200 33300 33400 33500 33600 33700 33800 33900 34000 34100 34200 34300 34400 34500 34600 34700 34800 34900 35000 35100 35200 35300 35400 35500 35600 35700 35800 35900 36000 36100 36200 36300 36400 36500 36600 36700 36800 36900 37000 37100 37200 37300 37400 37500 37600 37700 37800 37900 38000 38100 38200 38300 38400 38500 38600 38700 38800 38900 39000 39100 39200 39300 39400 39500 39600 39700 39800 39900 40000 40100 40200 40300 40400 40500 40600 40700 40800 40900 41000 41100 41200 41300 41400 41500 41600 41700 41800 41900 42000 42100 42200 42300 42400 42500 42600 42700 42800 42900 43000 43100 43200 43300 43400 43500 43600 43700 43800 43900 44000 44100 44200 44300 44400 44500 44600 44700 44800 44900 45000 45100 45200 45300 45400 45500 45600 45700 45800 45900 46000 46100 46200 46300 46400 46500 46600 46700 46800 46900 47000 47100 47200 47300 47400 47500 47600 47700 47800 47900 48000 48100 48200 48300 48400 48500 48600 48700 48800 48900 49000 49100 49200 49300 49400 49500 49600 49700 49800 49900 50000 50100 50200 50300 50400 50500 50600 50700 50800 50900 51000 51100 51200 51300 51400 51500 51600 51700 51800 51900 52000 52100 52200 52300 52400 52500 52600 52700 52800 52900 53000 53100 53200 53300 53400 53500 53600 53700 53800 53900 54000 54100 54200 54300 54400 54500 54600 54700 54800 54900 55000 55100 55200 55300 55400 55500 55600 55700 55800 55900 56000 56100 56200 56300 56400 56500 56600 56700 56800 56900 57000 57100 57200 57300 57400 57500 57600 57700 57800 57900 58000 58100 58200 58300 58400 58500 58600 58700 58800 58900 59000 59100 59200 59300 59400 59500 59600 59700 59800 59900 60000 60100 60200 60300 60400 60500 60600 60700 60800 60900 61000 61100 61200 61300 61400 61500 61600 61700 61800 61900 62000 62100 62200 62300 62400 62500 62600 62700 62800 62900 63000 63100 63200 63300 63400 63500 63600 63700 63800 63900 64000 64100 64200 64300 64400 64500 64600 64700 64800 64900 65000 65100 65200 65300 65400 65500 65600 65700 65800 65900 66000 66100 66200 66300 66400 66500 66600 66700 66800 66900 67000 67100 67200 67300 67400 67500 67600 67700 67800 67900 68000 68100 68200 68300 68400 68500 68600 68700 68800 68900 69000 69100 69200 69300 69400 69500 69600 69700 69800 69900 70000 70100 70200 70300 70400 70500 70600 70700 70800 70900 71000 71100 71200 71300 71400 71500 71600 71700 71800 71900 72000 72100 72200 72300 72400 72500 72600 72700 72800 72900 73000 73100 73200 73300 73400 73500 73600 73700 73800 73900 74000 74100 74200 74300 74400 74500 74600 74700 74800 74900 75000 75100 75200 75300 75400 75500 75600 75700 75800 75900 76000 76100 76200 76300 76400 76500 76600 76700 76800 76900 77000 77100 77200 77300 77400 77500 77600 77700 77800 77900 78000 78100 78200 78300 78400 78500 78600 78700 78800 78900 79000 79100 79200 79300 79400 79500 79600 79700 79800 79900 80000 80100 80200 80300 80400 80500 80600 80700 80800 80900 81000 81100 81200 81300 81400 81500 81600 81700 81800 81900 82000 82100 82200 82300 82400 82500 82600 82700 82800 82900 83000 83100 83200 83300 83400 83500 83600 83700 83800 83900 84000 84100 84200 84300 84400 84500 84600 84700 84800 84900 85000 85100 85200 85300 85400 85500 85600 85700 85800 85900 86000 86100 86200 86300 86400 86500 86600 86700 86800 86900 87000 87100 87200 87300 87400 87500 87600 87700 87800 87900 88000 88100 88200 88300 88400 88500 88600 88700 88800 88900 89000 89100 89200 89300 89400 89500 89600 89700 89800 89900 90000 90100 90200 90300 90400 90500 90600 90700 90800 90900 91000 91100 91200 91300 91400 91500 91600 91700 91800 91900 92000 92100 92200 92300 92400 92500 92600 92700 92800 92900 93000 93100 93200 93300 93400 93500 93600 93700 93800 93900 94000 94100 94200 94300 94400 94500 94600 94700 94800 94900 95000 95100 95200 95300 95400 95500 95600 95700 95800 95900 96000 96100 96200 96300 96400 96500 96600 96700 96800 96900 97000 97100 97200 97300 97400 97500 97600 97700 97800 97900 98000 98100 98200 98300 98400 98500 98600 98700 98800 98900 99000 99100 99200 99300 99400 99500 99600 99700 99800 99900 100000 100100 100200 100300 100400 100500 100600 100700 100800 100900 101000 101100 101200 101300 101400 101500 101600 101700 101800 101900 102000 102100 102200 102300 102400 102500 102600 102700 102800 102900 103000 103100 103200 103300 103400 103500 103600 103700 103800 103900 104000 104100 104200 104300 104400 104500 104600 104700 104800 104900 105000 105100 105200 105300 105400 105500 105600 105700 105800 105900 106000 106100 106200 106300 106400 106500 106600 106700 106800 106900 107000 107100 107200 107300 107400 107500 107600 107700 107800 107900 108000 108100 108200 108300 108400 108500 108600 108700 108800 108900 109000 109100 109200 109300 109400 109500 109600 109700 109800 109900 110000 110100 110200 110300 110400 110500 110600 110700 110800 110900 111000 111100 111200 111300 111400 111500 111600 111700 111800 111900 112000 112100 112200 112300 112400 112500 112600 112700 112800 112900 113000 113100 113200 113300 113400 113500 113600 113700 113800 113900 114000 114100 114200 114300 114400 114500 114600 114700 114800 114900 115000 115100 115200 115300 115400 115500 115600 115700 115800 115900 116000 116100 116200 116300 116400 116500 116600 116700 116800 116900 117000 117100 117200 117300 117400 117500 117600 117700 117800 117900 118000 118100 118200 118300 118400 118500 118600 118700 118800 118900 119000 119100 119200 119300 119400 119500 119600 119700 119800 119900 120000 120100 120200 120300 120400 120500 120600 120700 120800 120900 121000 121100 121200 121300 121400 121500 121600 121700 121800 121900 122000 122100 122200 122300 122400 122500 122600 122700 122800 122900 123000 123100 123200 123300 123400 123500 123600 123700 123800 123900 124000 124100 124200 124300 124400 124500 124600 124700 124800 124900 125000 125100 125200 125300 125400 125500 125600 125700 125800 125900 126000 126100 126200 126300 126400 126500 126600 126700 126800 126900 127000 127100 127200 127300 127400 127500 127600 127700 127800 127900 128000 128100 128200 128300 128400 128500 128600 128700 128800 128900 129000 129100 129200 129300 129400 129500 129600 129700 129800 129900 130000 130100 130200 130300 130400 130500 130600 130700 130800 130900 131000 131100 131200 131300 131400 131500 131600 131700 131800 131900 132000 132100 132200 132300 132400 132500 132600 132700 132800 132900 133000 133100 133200 133300 133400 133500 133600 133700 133800 133900 134000 134100 134200 134300 134400 134500 134600 134700 134800 134900 135000 135100 135200 135300 135400 135500 135600 135700 135800 135900 136000 136100 136200 136300 136400 136500 136600 136700 136800 136900 137000 137100 137200 137300 137400 137500 137600 137700 137800 137900 138000 138100 138200 138300 138400 138500 138600 138700 138800 138900 139000 139100 139200 139300 139400 139500 139600 139700 139800 139900 140000 140100 140200 140300 140400 140500 140600 140700 140800 140900 141000 141100 141200 141300 141400 141500 141600 141700 141800 141900 142000 142100 142200 142300 142400 142500 142600 142700 142800 142900 143000 143100 143200 143300 143400 143500 143600 143700 143800 143900 144000 144100 144200 144300 144400 144500 144600 144700 144800 144900 145000 145100 145200 145300 145400 145500 145600 145700 145800 145900 146000 146100 146200 146300 146400 146500 146600 146700 146800 146900 147000 147100 147200 147300 147400 147500 147600 147700 147800 147900 148000 148100 148200 148300 148400 148500 148600 148700 148800 148900 149000 149100 149200 149300 149400 149500 149600 149700 149800 149900 150000 150100 150200 150300 150400 150500 150600 150700 150800 150900 151000 151100 151200 151300 151400 151500 151600 151700 151800 151900 152000 152100 152200 152300 152400 152500 152600 152700 152800 152900 153000 153100 153200 153300 153400 153500 153600 153700 153800 153900 154000 154100 154200 154300 154400 154500 154600 154700 154800 154900 155000 155100 155200 155300 155400 155500 155600 155700 155800 155900 156000 156100 156200 156300 156400 156500 156600 156700 156800 156900 157000 157100 157200 157300 157400 157500 157600 157700 157800 157900 158000 158100 158200 158300 158400 158500 158600 158700 158800 158900 159000 159100 159200 159300 159400 159500 159600 159700 159800 159900 160000 160100 160200 160300 160400 160500 160600 160700 160800 160900 161000 161100 161200 161300 161400 161500 161600 161700 161800 161900 162000 162100 162200 162300 162400 162500 162600 162700 162800 162900 163000 163100 163200 163300 163400 163500 163600 163700 163800 163900 164000 164100 164200 164300 164400 164500 164600 164700 164800 164900 165000 165100 165200 165300 165400 165500 165600 165700 165800 165900 166000 166100 166200 166300 166400 166500 166600 166700 166800 166900 167000 167100 167200 167300 167400 167500 167600 167700 167800 167900 168000 168100 168200 168300 168400 168500 168600 168700 168800 168900 169000 169100 169200 169300 169400 169500 169600 169700 169800 169900 170000 170100 170200 170300 170400 170500 170600 170700 170800 170900 171000 171100 171200 171300 171400 171500 171600 171700 171800 171900 172000 172100 172200 172300 172400 172500 172600 172700 172800 172900 173000 173100 173200 173300 173400 173500 173600 173700 173800 173900 174000 174100 174200 174300 174400 174500 174600 174700 174800 174900 175000 175100 175200 175300 175400 175500 175600 175700 175800 175900 176000 176100 176200 176300 176400 176500 176600 176700 176800 176900 177000 177100 177200 177300 177400 177500 177600 177700 177800 177900 178000 178100 178200 178300 178400 178500 178600 178700 178800 178900 179000 179100 179200 179300 179400 179500 179600 179700 179800 179900 180000 180100 180200 180300 180400 180500 180600 180700 180800 180900 181000 181100 181200 181300 181400 181500 181600 181700 181800 181900 182000 182100 182200 182300 182400 182500 182600 182700 182800 182900 183000 183100 183200 183300 183400 183500 183600 183700 183800 183900 184000 184100 184200 184300 184400 184500 184600 184700

16. Click again on the “**Descriptions**” tab from the menu. You just looked at your sequences to get a general idea of how much they vary from one another, but how can we quantify this variation? That is why we now return to **BLAST scores**.
17. **BLAST scores** help us *quantify* the BLAST results. In the example below, a sequence named yellow fluorescent protein, “mLemon-YFP” has been compared to a Query Sequence (green fluorescent protein, or “GFP”). This comparison has a **Max Score** and **Total Score** of 1275, a **Query Coverage** of 100% and a **Percent Identity** of 99%. The **e-value** is 0.0.

Sequences producing significant alignments:						
Select: All None Selected: 0						
 Alignments	 Download	 Graphics				
	Description	Max score	Total score	Query cover	E value	Ident
<input type="checkbox"/> mLemon-YFP		1275	1275	100%	0.0	99%
						57255

Understanding BLAST Scores:

The **Max score** and **Total score** are related to the length of the sequences that are compared to one another and how well they match one another. Generally, the higher the score, the better the two sequences match each other. These scores are particularly helpful when comparing multiple protein sequences to each other.

Query coverage (abbreviated “**Query cover**”) & **Percent identity** (abbreviated “**Ident**”) quantify how much of the sequences match each other (**Query Coverage**), and how well they match (**Percent Identity**). For example, a small portion of the sequences (25% Query coverage) may match well (100% identity). Alternatively, 100% of the sequences may line up with one another (or “**align**”), but might share only 50% of the same nucleotides or amino acids (50% identity) within that matching region.

The **e-value** or **expect value** is an indication of how likely these results are based purely on chance. For example, if you have performed statistical analyses like a *chi-square test*, you should be familiar with a **p-value**. As with p-values, a low e-value mean you can be more confident that your results are not due to chance. Imagine if you just grabbed two random nucleotide or protein sequences and alignment them: how likely is it that they would match each other? This is your e-value.

EXAMPLES:

30% Query Coverage, 100% Identity

3/10 bases (30%) match perfectly (100%)

```

ATGGATACGT
TGAGATGATC

```

100% Query Coverage, 70% Identity

All 10 bases (100%) align, but only 70% match

```

ATGCCGACAG
AGGCAACAG

```

The formatting option “**Query-anchored with dots for identities**” BLAST alignment would look like this, with a dot in the Subject Sequence at each position where it matches the Query Sequence:

ATGGATACGT

TGA•••GATC

ATGCCGATTG

•G•G•A••••

18. Record in the data table below the Lawler family member that you analyzed, as well as the Query Coverage, the Percent Identity and the e-value. Remember that you are comparing this to the Query Sequence, which is the ‘non-mutated’ *BRCA1* DNA sequence.

Lawler Family Member	Query Coverage	Percent Identity	E-Value

19. Return to the BLAST homepage. You can do this by clicking on the previous window that you entered your sequences into, or you can click the “**blastn**” from the menu at the top left corner of the page.
20. Now let’s use BLAST to answer the following two questions: (a) Do other Lawler family members have any mutations in their *BRCA1* genes? (b) If other family members have a mutation or mutations, are they the **same** mutation or mutations among all family members?
21. **In the space below, write down your predicted answer to the following question: If multiple Lawler family members have a mutation or mutations in their *BRCA1* genes, would you expect each family member to have the same mutation as other family members? Why or why not?**
22. **See the instructions above if you need a reminder about how to complete the following steps.** From the Nucleotide BLAST page to enter sequences, make sure that the *BRCA1* Reference Sequence is entered into the Query Sequence box.
23. Copy and paste **all** of the Lawler family sequences into the Subject Sequence box. You can use BLAST to analyze **dozens** of Subject Sequences at the same time! Remember to include the “>Sequence Name” information for each sequence, so you know later whose sequence belongs to who. You can copy and paste all of the sequences at the same time.
24. Click “**Show Results in a new window** and then click “**BLAST.**”
25. Record your BLAST scores for each Lawler family member in Table 1 below.

Table 1: BLAST Scores for Lawler Family DNA Sequence Analysis

Lawler Family Member	Query Coverage	Percent Identity	E-Value
Deborah			
Lori			
Katherine			
Mother			
Father			
Uncle			

26. Based on the BLAST results in your table above, what can you conclude about *BRCA1* mutations in the Lawler family?
27. Any differences, or changes in the *BRCA1* DNA sequence, represent a **mutation** in the *BRCA1* gene. However, we need more information to determine whether this mutation results in a change in the amino acid found in the *BRCA1* protein. Amino acids are encoded by three bases, called a **codon**. Complete Table 2 below, including the codons and resulting amino acids (as represented by a one-letter abbreviation), or create a similar table in your lab notebook or on your homework paper. *See the codon table as instructed by your teacher.*

Table 2: Analysis and Predictions of Lawler Family DNA Mutations

	Reference Sequence	Mutated Sequence
DNA Coding Strand	ATG	
DNA Template Strand	TAC	
mRNA Codon	AUG	
Amino Acid		

28. What does it mean for the individual if that person has the mutation?
29. What does it mean for the individual if that person is free from the mutation?

Part II Procedure: Analyzing Lawler Family *BRCA1* Protein Sequences

Now, let's test your prediction the amino acid change that resulted from the mutation in the *BRCA1* gene among some of the Lawler family members.

30. To begin the **Protein BLAST**, click on the **"blastp"** tab at the top of the webpage.
31. Once you reach the Protein BLAST sequence comparison page, copy and paste the ***BRCA1 Protein Reference*** sequence in the Query Sequence box. **NOTE:** Your *BRCA1* DNA Reference sequence may be auto-filled in the Query Sequence box, depending upon your web browser. There is a small **"Clear"** button to the top right of the Query Sequence box that will remove this.
32. Click the **"Compare Two or More Sequences"** box.

33. Copy and paste **all** of the Lawler family BRCA1 Protein sequences in the Subject Sequence box. The *BRCA1* Lawler family sequences may also be auto-filled in the Subject Sequence box. There is a “**Clear**” button next to the Subject Sequence as well that will quickly remove these sequences.
34. At the bottom of the page, click “**Show results in a new window.**”
35. Click “**BLAST.**”
36. In the table below, record the BLAST scores that you see in the “**Descriptions**” tab that you see when the BLAST results appear.

Table 3: BLAST Scores for Lawler Family Protein Sequence Analysis

Lawler Family Member	Query Coverage	Percent Identity	E-Value
Deborah			
Lori			
Katherine			
Mother			
Father			
Uncle			

37. Are the Protein BLAST scores similar to the DNA BLAST scores that you obtained in Part I? What can you conclude based on your BLAST scores?
38. Click on the “**Alignments**” tab.
39. Reformat your results by clicking on the “**Alignment view**” and selecting “**Pairwise with dots for identities.**”
40. What mutation or mutations, if any, do some of the Lawler family members have in their *BRCA1* gene? Is this consistent with your findings and predictions from your Nucleotide BLAST analyses?

Part III Procedure: Analyzing Patient BRCA1 Protein Sequences

Now that you have learned how to perform both Nucleotide and Protein BLAST analyses, it’s time to analyze the BRCA1 protein sequence(s) from the patient(s) that you have been responsible for. For this analysis, we will use BRCA1 protein sequences so that we can more easily identify mutations.

41. To begin the next **Protein** BLAST, click on the “**blastp**” tab at the top of the webpage.
42. Once you reach the Protein BLAST sequence comparison page, copy and paste the **BRCA1 Protein Reference** sequence in the Query Sequence box. **NOTE:** Your *BRCA1* DNA Reference

sequence may be auto-filled in the Query Sequence box, depending upon your web browser.

There is a small “Clear” button to the top right of the Query Sequence box that will remove this.

43. Click the “**Compare Two or More Sequences**” box.
44. Copy and paste BRCA1 protein sequence for your patient or patients. The *BRCA1* Lawler family sequences may also be auto-filled in the Subject Sequence box. There is a “Clear” button next to the Subject Sequence as well that will quickly remove these sequences.
45. At the bottom of the page, click “**Show results in a new window.**”
46. Click “**BLAST.**”
47. In the table below, record the BLAST scores that you see in the “**Descriptions**” tab that you see when the BLAST results appear.

Table 4: BLAST Scores for Patient Protein Sequence Analysis

Patient ID	Query Coverage	Percent Identity	E-Value

48. Based on these BLAST scores, do you predict that your patient or patients have a mutation or mutations in their *BRCA1* gene?
49. Click on the “**Alignments**” tab.
50. Reformat your results by clicking on the “**Alignment view**” and selecting “**Pairwise with dots for identities.**”
51. What mutation or mutations, if any, is/are found in your patient or patients? Is this consistent with your prediction or predictions based on the BLAST scores?
52. Based on these findings, as a genetic counselor or other health care provider, how you advise your patient or patients regarding their risk for breast cancer?
53. If your patient or patients do not have any mutations in their *BRCA1* gene, does that mean that they won’t get breast cancer? Why or why not?