Blasting Human & Chimp DNA

1. Go to ncbi.nlm.nih.gov. This site is the NCBI home page through which you can access both the DNA databases and BLAST.
2. At the top of the home page, search All Databases (default) for HBB. HBB is the abbreviation for beta hemoglobin. Click "Go."
3. In the middle of the page on the left, click on "Nucleotide: core subset of nucleotide sequence records."
4. Near the top of the page, click on "HBB (*Homo sapiens*)." You may have to right-click with your mouse to open this link. A new screen will come up that gives you the nucleotide sequence of the gene that codes for the human beta-hemoglobin protein. There is a lot of additional information on this, and you will see links to many other pages. You can explore these after you have finished the basic tutorial.
5. Scroll down a little until you see, in a gray box on the left, "Genomic regions, transcripts and products." Click on "Reference sequence details."
6. On the left side of the screen, find "Genome Reference Consortium Human Build" in a gray box. Under this, find "Genomic," then find "Download." To the right of "Download," click on "FASTA." You will see that the beta hemoglobin gene contains 1606 nucleotides ("letters"), here shown in what we call FASTA format. FASTA is a standard format, used for both genes and proteins, in which the first line begins with a □ symbol, followed by any text. The text is sometimes longer than one line, as it is in this example. This is OK so long as there is no "return" in the text. After this one line of text, there is a return, followed by the nucleotide sequence of the gene or the amino acid sequence of the protein. In cases of proteins, the single-letter abbreviations for each of the 20 amino acids are used.
7. Copy the entire DNA sequence. Do not copy the first line.
8. Go back to the NCBI home page (ncbi.nlm.nih.gov) by clicking on the NCBI logo at the top of the page.
9. At the top of the page, click on "BLAST."
10. Under "BLAST Assembled Genomes" near the top of the page, click on "*Pan troglodytes*." This link will allow you to compare the human beta-hemoglobin gene with the chimpanzee genome, and not with the entire NCBI database. This step saves time, because beta hemoglobins have been sequenced for innumerable organisms. Other BLAST searches could allow you to compare your sequence with the entire NCBI database or with different subsets of the sequence.
11. Paste the human beta-hemoglobin sequence that you have copied into the large box at the top of the page. This is your query. The directions for the box say "enter an accession…."
12. Scroll to the bottom of the page and click on "Begin Search."
13. In the middle of the page, to the right of "Request ID," click on "View Report" in a gray box.
14. Scroll down and look at the Graphic Summary. Notice that there is one match to a gene on the number 11 chromosome of the chimpanzee. Note that the gene for human beta hemoglobin is on human chromosome number 11. At the bottom of the box, you will see one bright red line indicating that there is one match. The fact that the line is bright red indicates that this is a very close match. Mouse over this line, and it will tell you, in technical terms, which chimpanzee gene your query (human beta hemoglobin) matches.
15. Scroll down to "Alignments." The "Query" is human beta hemoglobin. The "Subject" is chimpanzee beta hemoglobin. Notice that whenever the gene is identical, there is a vertical line between the identical base pairs.
16. Now go to the top of the page on the left and click on "Download."
17. Near the top of the page under "Alignment," click on "text."
18. Click on "save to disk."
19. Open the file. This file will give you a comparison of the two beta hemoglobins. You can print this file out if a printout is not included in the packet your teacher has given you (it is presented here as Figure 2). When this is on your computer screen, you will be able to search for the beginnings and ends of the exons in later steps.
20. Now look at the copy of the human-beta hemoglobin gene included in the packet your teacher has given you (Figure 1). This copy of the gene is from 1980 and uses the three-letter abbreviations for the 20 amino acids. There are three exons and two introns in this gene. You can identify the three exons because the amino acids they code for are given above the base sequences. Look at the sequences at the beginning and end of each exon. Use a highlighter to highlight the nucleotide sequences of each of the three exons in the human beta-hemoglobin gene on this copy of
the human beta-hemoglobin gene.
21. Take out your printout of the human–chimpanzee beta-hemoglobin comparison from the BLAST search (Figure 2).
22. Use the find options of the computer (Edit/Find) to identify the beginning and end of each exon of the human beta-hemoglobin gene. Highlight the three exons on your printout of the human– chimpanzee comparison. It is easiest to highlight the exons in the human gene. Hint: search for six bases at the beginning and end of each exon, which you have located in Figure 1. Six bases are enough to identify these positions.

23. You can now answer the following questions :
a. Are the exons identical?. Circle or highlight any differences you find. How many differences are there?
b. Are the introns identical? Circle or highlight any differences you find. How many differences are there?

c. Are there more differences in the exons or in the introns? Can you think of an explanation for this difference?
d. How can there be a difference in the exons if the proteins are identical?

```
Query = Human Beta Hemoglobin
Length = 1606
Score    E
Sequences producing significant alignments: (Bits)  Value

ref|NW_001222274.1|Ptr11_WGA13838_2  Pan troglodytes chromosom...  2867   0.0

Score = 2867 bits (1552),  Expect = 0.0
Identities = 1588/1606 (98%), Gaps = 0/1606 (0%)
Strand = Plus/Minus
```

Query 1	ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATC	60
Sbjct 851953	ACATTTGCTTCTGACACAACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCACC	851894

Query 61	TGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG	120
Sbjct 851893	TGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG	851834

Query 121	TTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCA	180
Sbjct 851833	TTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGCTTAAGGAGACCA	851774

Query 181	ATAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCT	240
Sbjct 851773	GTAGAAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCT	851714

Query 241	CTCTGCCTATTGGTCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAG	300
Sbjct 851713	CTCTGCCTATTGGGCTATTTTCCCACCCTTAGGCTGCTGGTGGTCTACCCTTGGACCCAG	851654

Query 301	AGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAG	360
Sbjct 851653	AGGTTCTTTGAGTCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAG	851594

Query 361	GTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC	420
Sbjct 851593	GTGAAGGCTCATGGCAAGAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGAC	851534

Query 421	AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT	480
Sbjct 851533	AACCTCAAGGGCACCTTTGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGAT	851474

Query 481	CCTGAGAACTTCAGGGTGAGTCTATGGGACGCTTGATGTTTTCTTTCCCCTTCTTTTCTA	540
Sbjct 851473	CCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTTCTTTCCCCTTCTTTTCTA	851414

Query 541	TGGTTAAGTTCATGTCATAGGAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAAC	600
Sbjct 851413	TGGTTAAGTTCATGTCATAGGAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAAC	851354

Query 601	AGACGAATGATTGCATCAGTGTGGAAGTCTCAGGATCGtttttagtttcttttatttgctg	660
Sbjct 851353	AGACAAATGATTGCATCAGTGTGGAACTCTCAGGATCATTTTAGTTTCTTTTATTTGCTG	851294

Query 661	ttcataacaattgttttcttttgtttaattcttgctttcttttttttttcttctccgcaat	720
Sbjct 851293	TTCATAACAATTGTTTTCTTTTGTTTAATTCTTGCTTTCTTTTTTTTTCTTCTCTGCAAT	851234

Query 721	ttttactattatacttaatgccttaacattGTGTATAACAAAAGGAAATATCTCTGAGAT	780
Sbjct 851233	TTTTACTATTATACTTAATGCCTTAACGTTGTGTATAACAAAAGGAAATATCTCTGAGAT	851174

Query 781	ACATTAAGTAACTTaaaaaaaaaaCTTTACACAGTCTGCCTAGTACATTACTATTTGGAAT	840
Sbjct 851173	ACATTAAGTAACTTAAAAAAAATGTTTACGCAGTCTGCCTAGTACATTACTATTTGGAAT	851114

Query 841	ATATGTGTGCTTATTTGCATATTCATAATCTCCCTACtttattttcttttatttttaatt	900
Sbjct 851113	ATATGTGTGCTTATTTGCATATTCATAATCTACCTACTTTATTTTCTTTTATTTTTAATT	851054

Query 901	gatacataatcattatacatatttatgggttaaagtgtaatgttttaatatgtgtacaca	960
Sbjct 851053	GATACATAATCATTATACATATTTATGGGTTAAAGTGTAATGTTTTAATATGTGTACACA	850994

Figure 2. The BLAST printout page, showing the comparison between human beta hemoglobin (Query) and chimpanzee beta hemoglobin (Subject). This figure was printed from NCBI's BLAST site.