**NCBI Genome Database - Comparing Hemoglobin Genes 2019**

**Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Period \_\_\_\_\_\_\_\_\_\_\_\_**

NCBI (National Center for Biotechnology Information) is a huge database where DNA sequences from labs all over the world are deposited. It contains an enormous amount of information about many different organisms, including humans, that is available to anyone at no cost. The database is easily searched using a free program called BLAST (Basic Local Alignment Search Tool). In this activity, you will learn how to use these databases and BLAST to compare a human and chimpanzee hemoglobin gene.

 Find it:

1. What is hemoglobin? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. What is the HBB gene and on what chromosome can it be found? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. What is an exon and an intron? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Describe the relationship between DNA, proteins, and amino acids.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

5. About how long ago did human and chimpanzees share a common ancestor?\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Procedure:**

**Part I – Human hemoglobin exons and introns**

1. Go to **ncbi.nlm.nih.gov** This is the NCBI home page through which you can access both the DNA databases and BLAST.

2. Type **HBB homo sapiens** in the search box at the top.

3. Scroll down and click on **“Nucleotide”** under the subheading “Genomes”.

4. Scroll down to click on the hyperlink entitled: [**Homo sapiens** hemoglobin subunit beta (**HBB**), mRNA](https://www.ncbi.nlm.nih.gov/nuccore/NM_000518.5)

628 bp linear mRNA

It may be #12 in the list????

5. On the right sidebar, under “Related Information”, click on **“Gene”.**

6. Scroll down a little until you see, in a gray box on the left “**Genomic regions, transcripts and products**.” Click on the word **FASTA** under the gray box and to the right. You will see a series of 1606 nitrogen base letters that are in the hemoglobin gene.

7. Copy this entire sequence and paste it into Word – NOT GOOGLE DOCS!!! Make sure you also copy the beginning that includes >NC…..

**NOTE: This will not work properly on any word processing program, including google docs!!! You must use WORD – that’s why I instructed you to download the package at the beginning of the year.**

 Type your names at the top and give it a title when you save (click File, Rename) You can have both partners get a copy by using your USB drive or email. DO NOT PRINT YET – you will do that at step #14.

**Part II – Compare to chimpanzee beta hemoglobin**

9. Go back to the **NCBI home** page by clicking on the NCBI logo at the top left corner of the page. At the top and in the right sidebar under **Popular Resources**, click on **BLAST**. In the Search box titled “**BLAST Genomes**”, type “**Pan troglodytes (taxid:9598)**” and click **Search**. This is the genus and species name for the chimpanzee.

10. In your Word document, copy the first DNA sequence you pasted in step #8 (the HBB homo sapiens).Make sure you include the beginning that starts >NC…. You will paste the sequence in the big, white box of the NCBI BLAST page that says “**Enter accession number(s)**….” . If there are any words in the box, delete them first before pasting. The human sequence is your **“query”**.

11. Scroll to the bottom of the page and click on **BLAST**. It takes some time for the sequences to come up, so be patient.

12. Scroll down to see the Human-Chimp DNA comparison for the beta hemoglobin genes under the heading **“Alignments”.** The top row, “**Query**” ,is human beta hemoglobin and the bottom row, “**Subject**”, is the chimpanzee beta hemoglobin. You will notice that there is a vertical line between each identical nitrogen base and no line if there is a difference in the DNA sequences. This comparison only shows a single strand of Human and the same strand of Chimpanzee (both are the *nonsense strands* – side that is not complementary to the mRNA during transcription).

13. Copy the entire gene up to where you see the # 5086003**.** It will look like this at the top**:**



See the next page to see where you stop copying.



stop copying

 here

14.Then paste it into your Word document. *You must paste as “Keep Source Formatting” or the bases will not line up properly!* Put your names at the top. SAVE IT. Lastly, print one copy for each group member.

**Analysis:**

15. Look at the copy of the HBB gene on the next page of this lab. There are 3 exons and 2 introns. You can identify the exons because I have made the font slightly bigger, bold face type and highlighted. On ***your printout***, identify and **highlight the 3 exons** of the human/query and the matching chimp/sbjct sequence below it. Don’t worry if some of the nitrogenous bases are lower case.

\*\*To make it easier, use the “find” on the Home page of Word to look for a sequence of 6 or so nitrogen bases. Type in the beginning of each exon. It will be highlighted on the Word document. Then type in the ending sequence of the exon. Highlight the ending sequence and then every nucleotide base in between. Highlight BOTH the human and chimpanzee sequence lines.

16. Label each intron and exon on the right side margin of your printout. Example: Exon #1, Intron #1…..

17. Using your codon translator, identify the first 4 amino acids from the first exon for both the chimp and the human. ***Note: the ATG sequence that begins exon #1 is the START codon, AUG,– this is the nonsense (noncoding) strand of DNA. It’s a quick way researchers use to analyze the DNA. All you need to do to find the mRNA sequence is change all Thymines to Uracils. DO NOT DO THIS on your quiz/test!***

 *On the test, I will give you the coding strand to test your understanding of transcription.*

The human mRNA sequence: \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_

 Human AA sequence: *\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_*

The chimpanzee mRNA sequence: \_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_

 Chimpanzee AA sequence: *\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_*

*Why is it easier to translate directly from the nonsense strand rather than the actual coding strand/sense strand? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

18. Despite the differences in the DNA sequences, the chimp amino acid sequence could still be the same as the human amino acid sequence (these are called silent mutations). Explain how this is possible.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

19. Look carefully at the Human and Chimp sequences. Are the exons identical? \_\_\_\_\_\_\_\_\_\_\_ Are the introns identical? \_\_\_\_\_\_\_\_\_\_\_**Circle all the differences in the exons and introns with a pen**.

How many total differences are there in the introns? \_\_\_\_\_\_\_\_\_\_\_\_

How many total differences are there in the exons? \_\_\_\_\_\_\_\_\_\_\_\_

20. Mutations occur over the entire genome at the same rate. Give an explanation for why there are more differences in the introns than in the exons. Hint: it has to do with evolution/natural selection.

21. Turn in this lab paper and your printouts. Make sure your name is on everything. ***For homework, write a 1 paragraph (approx.. ½ page) reflection and attach it to this lab.*** *You could answer questions such as: What did you learn? What did you struggle with? What might a scientist struggle with? How is this database important? What further information could we (or scientists) gain by using an activity similar to this? How else could this activity be used?*

**HBB gene (the exons are highlighted):**

 70501 gccagggctg ggcataaaag tcagggcaga gccatctatt gcttacattt gcttctgaca

 70561 caactgtgtt cactagcaac ctcaaacaga cacc**atggtg catctgactc ctgaggagaa**

 70621 **gtctgccgtt actgccctgt ggggcaaggt gaacgtggat gaagttggtg gtgaggccct**

70681 **gggcaggttg gtatcaaggt tacaagacag gtttaagg**ag accaatagaa actgggcatg

 70741 tggagacaga gaagactctt gggtttctga taggcactga ctctctctgc ctattggtct

 70801 attttcccac ccttagg**ctg ctggtggtct acccttggac ccagaggttc tttgagtcct**

 70861 **ttggggatct gtccactcct gatgctgtta tgggcaaccc taaggtgaag gctcatggca**

 70921 **agaaagtgct cggtgccttt agtgatggcc tggctcacct ggacaacctc aagggcacct**

70981 **ttgccacact gagtgagctg cactgtgaca agctgcacgt ggatcctgag aacttcagg**g

 71041 tgagtctatg ggacgcttga tgttttcttt ccccttcttt tctatggtta agttcatgtc

 71101 ataggaaggg gataagtaac agggtacagt ttagaatggg aaacagacga atgattgcat

 71161 cagtgtggaa gtctcaggat cgttttagtt tcttttattt gctgttcata acaattgttt

 71221 tcttttgttt aattcttgct ttcttttttt ttcttctccg caatttttac tattatactt

 71281 aatgccttaa cattgtgtat aacaaaagga aatatctctg agatacatta agtaacttaa

 71341 aaaaaaactt tacacagtct gcctagtaca ttactatttg gaatatatgt gtgcttattt

 71401 gcatattcat aatctcccta ctttattttc ttttattttt aattgataca taatcattat

 71461 acatatttat gggttaaagt gtaatgtttt aatatgtgta cacatattga ccaaatcagg

 71521 gtaattttgc atttgtaatt ttaaaaaatg ctttcttctt ttaatatact tttttgttta

 71581 tcttatttct aatactttcc ctaatctctt tctttcaggg caataatgat acaatgtatc

 71641 atgcctcttt gcaccattct aaagaataac agtgataatt tctgggttaa ggcaatagca

 71701 atatctctgc atataaatat ttctgcatat aaattgtaac tgatgtaaga ggtttcatat

 71761 tgctaatagc agctacaatc cagctaccat tctgctttta ttttatggtt gggataaggc

 71821 tggattattc tgagtccaag ctaggccctt ttgctaatca tgttcatacc tcttatcttc

 71881 ctcccacag**c** **tcctgggcaa cgtgctggtc tgtgtgctgg cccatcactt tggcaaagaa**

 71941 **ttcaccccac cagtgcaggc tgcctatcag aaagtggtgg ctggtgtggc taatgccctg**

 72001 **gcccacaagt atcactaagc tcgctttctt gctgtccaat ttctattaaa ggttcctttg**

 72061 **ttccctaagt ccaactacta aactggggga tattatgaag ggccttgagc atctggattc**

 72121 **tgcctaataa aaaacattta ttttcattgc** aatgatgtat ttaaattatt tctgaatatt

 72181 ttactaaaaa gggaatgtgg gaggtcagtg catttaaaac ataaagaaat gaagagctag

 72241 ttcaaacctt gggaaaatac actatatctt aaactccatg aaagaaggtg aggctgcaaa

 72301 cagctaatgc acattggcaa cagcccctga tgcatatgcc ttattcatcc ctcagaaaag