

about this protein. The encoded protein functions as a chloride channel, making it unique among members of this protein family, and controls ion and water secretion and absorption in epithelial tissues.

8. FOR WHAT DISEASE IS A MUTATED FORM OF THIS GENE RESPONSIBLE? You should be able to get this information from the description of the gene. **cystic fibrosis**

9. ON WHAT CHROMOSOME IS THE GENE LOCATED? You should be able to get this information by looking at the “Genomic context” which is just below the description of the gene. **CHROMOSOME 7**

10. Return to the original nucleotide sequence alignment descriptions. CHOOSE A SPECIES (STATE THE SCIENTIFIC NAME) OTHER THAN *HOMO SAPIENS* THAT ALSO HAS A 100% IDENTITY (Per. Ident) FOR THIS SEQUENCE? There will be about 3 species that you could name, so choose one of them. **Pongo abelii**

11. WHAT IS THE COMMON NAME FOR THIS SPECIES? **Sumatran orangutan**

12. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT? **NO BECAUSE STUDIES SUGGEST THAT HUMANS MAY SHARE A COMMON ANCESTOR WITH ORANGUTANS AND THAT ANCESTRY COULD EXPLAIN THE SIMILARITY.**

13. Return to the original nucleotide sequence alignment **DESCRIPTION**. Find the first match that has less than 100% similarity identity. Click on the description to answer this question.

a. WHAT IS THE GENUS AND SPECIES WITH THIS NUCLEOTIDE SEQUENCE? **Genus: Nomascus**
Species: leucogenys

b. WHAT IS THE COMMON NAME? **northern white-cheeked gibbon.**

c. ARE THERE ANY GAPS BETWEEN THE TWO SEQUENCES (THE ONE YOU ORIGINALLY SUBMITTED AND ONE THAT HAS LESS THAN 100% QUERY COVER)? The gap information is just above the sequence alignment. **YES**

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you’ll have to search for on the Internet.) It is a space where one sequence has an insertion or deletion compared to another sequence. It helps align sequences with mutations or evolutionary changes for comparison.

You can also do BLAST searches using an accession number that has been assigned to a particular sequence when was entered into the database. Go back to the Blast home page (www.ncbi.nlm.nih.gov/BLAST.cgi) and again choose “Nucleotide Blast”. Look up the following sequences using the given accession numbers. (Under “Program Selection” near the bottom of the page, choose “somewhat similar sequence (blastn)”. (Again, click on the “BLAST” button at the bottom of the page after you have entered the accession number.)

FOR EACH, STATE WHAT THE GENE IS (#15-18). (Again, give the description of the gene or gene product, not the nucleotide sequence.)

15. NM_145556 - **Mus musculus TAR DNA binding protein (Tardbp), transcript variant 1, mRNA**

16. NM_013444 - **Homo sapiens ubiquitin 2 (UBQLN2), mRNA**

17. NM_001010850 - **Homo sapiens FUS RNA binding protein (FUS), transcript variant 2, non-coding RNA**

18. KJ174530 - **Homo sapiens superoxide dismutase-1 (SOD-1) gene, exon 1 and partial cds**

19. Search Google to answer the following: WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #15-#18? WHAT IS A “COMMON NAME” OF THE DISEASE? (The name of a person; Hint, hint...Baseball season is just beginning...) **Lou Gehrig's disease.**

20. BLAST is possible because of the submission of DNA sequences to GenBank. WHAT IS GENBANK? (You can do an Internet search to find this information.) **GenBank is a public database housing DNA and RNA sequences, maintained by the National Center for Biotechnology Information (NCBI), facilitating global access to genetic information for research and analysis.**

Introduction to Protein BLAST

Assignment Goal: To use the Internet-based site BLAST, Basic Local Alignment Search Tool (NCBI), to 1) translate cDNA and 2) to search for similarities between amino acid sequences.

Assignment: Access the BLAST site at <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
Click on “blastx”. (This is a smaller button next to “Nucleotide BLAST”).