Kevin Johnson

Dr. Rinehart-Kim

BIOL 294 Genetics

April 18, 2024

Genome Assignment

Chromosome Maps

1. WHAT CHROMOSOME DID YOU CHOOSE? Chromosome 22

2 & 3. STATE THE NUMBER OF GENES AND BASE PAIRS ON THE CHROMOSOME YOU CHOSE. Over 800 genes and over 40 million base pairs of which approx. 70% have been determined

4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME. IGL immunoglobulin lambda locus

5. STATE THE FUNCTION OF THE GENE YOU LISTED IN #4. It is important that you state the physiological function of the gene product you select, not the disease listed for the gene. The gene recognizes foreign antigens and initiates immune responses such as phagocytosis and the complement system. Per NIH report.

Nucleotide BLAST

6. WHAT IS THE SECOND SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE? Homo sapiens CFTR (CFTR) gene. Partial cds

7. WHAT DOES THE ENCODED PROTEIN DO IN THE BODY? It encodes a member of the ATP-binding cassette transporter superfamily. The encoded protein functions as a chloride channel, making it unique among the members of this protein family. The protein controls ion and water secretion and absorption in epithelial tissues.

8. FOR WHAT DISEASE IS A MUTATED FORM OF THIS GENE RESPONSIBLE? Cystic fibrosis

9. ON WHAT CHROMOSOME IS THE GENE LOCATED? 7q31.2

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? here will be about 3 species that you could name, so choose one of them. Pan troglodytes

11. WHAT IS THE COMMON NAME FOR THIS SPECIES? chimpanzee

12. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? No it didn't

WHY OR WHY NOT? Based off evolution we came from a common ancestor and this is actually something commonly referenced in different scientific works. I am pretty sure one of the others probably would have surprised me I just recognized the scientific name.

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? Nomascus 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)? No, It is showing 0/120 gaps (0%)

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? It is a maximal sequence of consecutive insertions or deletions(indels) of letters (Barton et al., 2015).

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 ubiquilin 2 (UBQLN2), mRNA

17. NM_001010850 Homo sapiens fusion (involved in t(12;16) in malignant liposarcoma) (FUS), transcript variant 2, mRNA

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? Amyotrophic lateral sclerosis ALS The common name was Lou Gehirg's disease (Amyotrophic Lateral Sclerosis (Als)).

20. BLAST is possible because of the submission of DNA sequences to GenBank. WHAT IS GENBANK? GenBank is the NIH genetic sequence database (Genbank Overview, n.d.).

Protein BLAST

21. First, answer this question: WHAT IS cDNA? Copy DNA It is a synthetic DNA that has been transcribed from a specific mRNA through a reaction using the enzyme reverse transcriptase (*Cdna (Copy Dna)* | *Nhgri*, n.d.).

ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCA
TCTGACTC

CTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTG
GTGGTGAG

GCCCTGGGCAG

22. WHAT IS THE SEQUENCE MATCH? Nucleotide Sequence beta globin

Assignment: While still on BLAST, choose “Protein BLAST”. Check the box which says “Align two or more

sequences”.

Copy and paste the following sequence into the “Enter Query Sequence” box:

Person 1/Sequence 1:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELS
CPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLC
HFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCP
AAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGS
IRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNG
SKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWL
VVLPAEEELVEADEAGSVYAGILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKIS
RFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGK
PLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIG
KHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKD
LVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKK
TTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKE
GHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFE
QYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Copy and paste the following sequence into the “Enter Subject Sequence” box:

Person 2/Sequence 2:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELS
CPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLC
HFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCP
AAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGS
IRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNG
SKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWL
VVLPAEEELVEADEAGSVYAGILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISR

FPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGKP
LGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGK
HKNIINLL

GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQV
ARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPV
KWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPA
NCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQD
TPSSSSSGDDSVFAHDLLPPAPPSSGGSRT

Click on the “BLAST” box.

23. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES? Yes I found several differences they are in red text in the sequences above.

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY? There is a G in the first sequence GILSY**G**VG and an R in the second sequence GILSY**R**VG. Also there was an additional S at the end of the second sequence. First sequence DTP**SSSS**GD Second sequence DTP**SSSSS**GD.

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT? There is a gap between the NLL and GACT in the second sequence.

26. WHAT GENE ENCODES FOR THE POLYPEPTIDE YOU WERE ANALYZING? Amino acid-fibroblast growth factor receptor 3

27. WHAT IS THE FUNCTION OF THIS PROTEIN? It encodes a member of the fibroblast growth factor receptor or FGFR family, with its amino acid sequence being highly conserved between members and among divergent species.

28. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE? Craniosynostosis and several types of skeletal dysplasia

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS? (What do the abbreviations stand for? Who oversees what?) NIH is National Institute of Health, NLM is National Library of Medicine, NCBI is National Center for Biotechnology Information, HHS is U.S. Department of Health and Human Services. They are all part of the HHS and are overseen by the HHS.

30. WHAT WAS ONE POSITIVE THING AND ONE NEGATIVE THING YOU ENCOUNTERED WHILE DOING THIS ASSIGNMENT? The biggest positive thing was enjoying it. I was nervous going into the assignment and once I got started really had fun with it. I am not sure how accurate a couple answers are here I just know it was enjoyable. Especially the 2 sequences and figuring out the differences. The one negative thing is I keep second guessing myself on some of my answers being unsure if I am answering with what is requested.

References

Amyotrophic lateral sclerosis (Als) | national institute of neurological disorders and stroke. (n.d.). Retrieved April 18, 2024, from https://www.ninds.nih.gov/health-information/disorders/amyotrophic-lateral-sclerosis-als

Barton, C., Flouri, T., Iliopoulos, C. S., & Pissis, S. P. (2015). Global and local sequence alignment with a bounded number of gaps. *Theoretical Computer Science*, *582*, 1–16. https://doi.org/10.1016/j.tcs.2015.03.016

Cdna (Copy dna) | nhgri. (n.d.). Retrieved April 18, 2024, from https://www.genome.gov/genetics-glossary/Copy-DNA

Genbank overview. (n.d.). Retrieved April 18, 2024, from https://www.ncbi.nlm.nih.gov/genbank/