

Lab 4: Bioinformatics

Background

In this lab, we will combine the bioinformatics training from this week with last week's genomic skills to identify and estimate the evolutionary history of an unknown human DNA sequence associated with disease.

Mystery Sequence

```
CCTCACCTGTGGAGCCACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGG
GCTGGGCATGAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAACCTGTGTTCACTAG
CAACCTCAAACAGACACCATGGTGACCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCA
AGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAA
GGAGACCAATAGAACTGGGCATGTGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCT
CTGCCTATTGGTCTATTTTCCACCCCTTAGGCTGCTGGTGGTCTACCCCTGGACCCAGAGGTTCTTTGAG
TCCTTTGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAAGAAAG
TGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTTTGCCACACTGAGTGA
GCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGGTGAGTCTATGGGACCCTTGATGTTTT
CTTTCCCTTCTTTTCTATGGTTAAGTTCATGTCATAGGAAGGGG
```

Instructions

1. What is the above sequence? What gene is it from? Who published the paper that the sequence is from?
2. Identify the mutations in the gene that are associated with disease. There are multiple different methods to the right answer here.
3. Which disease(s) would be associated with the above DNA sequence? Describe the details of the disease(s): What are the symptoms of the disease? What are the biological mechanisms? Is there treatment available for the disease?
4. Using your knowledge from class and internet resources, describe (in your own words, no copy-pasting!) an alignment technique.
5. Using skills from class, find at least 10 other variants of the gene above and download them. Describe how you found these other variants.
6. Using the variants that you downloaded as well as the sequence provided above, align all of the sequences. Describe the method by which you aligned the sequences.
7. Estimate the evolutionary history of the gene by building a phylogeny with the multiple-sequence alignment resulting from step 6. Describe the method by which you built a phylogeny.
8. View your phylogeny and interpret the results. Did the results fit your expectations? What implications, if any, follow, given your estimated evolutionary history? Attach a copy of your phylogeny to the report.

The report

Develop a report (I recommend a Word (or other text editor) document) for your problem set that includes answers to all of the questions posed above, showing plots where appropriate.

Save your report as a PDF file and submit your report through the course 2GW site. Clean up your code and submit it as a supplementary file, along with your main report.

Due date

Day 7, Week 9