

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 sequence data associated with disease.

Mystery Sequence

```
MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVVYPDKVFR
SSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTKSNIIIR
GWIFGTTLDSTQSLIVNNATNVVIVKCEVFQFCNDPFLGVYYHKNNKSWMESEFRVY
SSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDL PQ
GFSALEPLVDLPIGINITRFQTLALHRSYLT PGDSSSGWTAGAAAYVGYLQPRTFL
LKYNENGTITDAVDCALDPLSETKCTLKSF TVEKGIYQTSNFRVQP TESIVRFPNITN
LCPFGVEFVNATRFASVYAWNRRKRISNCVADYSVLN SASFSTFKCYGVSPTKLNDLCF
TNVYADSFVIRGDEV RQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYN
YLYRLF RKSNLKP FERDISTEIIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYPY
RVVVSFELLHAPATVCGPKKSTNLVKNKCVNFNFGLT GorTGVLTESNKKFLPFQQFG
RDIADTTDAVRDPQTLEILDITPCSFGGVS VITPGTNTSNQVAVLYQDVNCTEVPVAI
HADQLTPTWRVYSTGSNV FQTRAGCLIGA EHVNNSYECDIPIGAGICASYQTQTNSPR
RARSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSVDCTM
YICGDSTEC SNLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFG
GFNFSQLPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLG DIAARDLICAQKFN
GLTVLPPLL TDEMIAQYTSALLAGTITSGWTFGAGAA LQIPFAMQMAYRFNGIGVTQN
VLYENQKLIANQFNSAIGKIQDSL SSTASALGKLQDVVNQNAQALNTLVKQLSSNFGA
ISSVLNDIL SRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMS
ECVLGQSKRVDFCGKGYHLMSPQ SAPHGVVFLHVTYVPAQEKNFTTAPAICHGKAH
FPREGFVSVNGTHWFTQRNFYEPQIITTDNTFVSGNCDV VIGIVNNTVYDLPQLPELD
SFKEELDKYFKNHTSPD VDLGDISGINASV VNIQKEIDRLNEVAKNLNESLIDLQELG
KYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCC SCLKGCCSCGSCCKFDEDDSE
PVLKGVKLHYT
```

Instructions

1. What is the above sequence? What gene is it from? Who published the paper that the sequence is from?
2. In one sentence, why is this particular gene of interest to medical and public health practice?
3. Using your knowledge from class and internet resources, describe (in your own words, no copy-pasting!) an alignment technique. > Note: steps 4 and 5 may take some time to run!
4. Download the file `all_sequences.fasta` from the Data folder. Align all of the sequences. Describe the method by which you aligned the sequences.
5. Estimate the evolutionary history of these sequences by building a phylogeny with the multiple-sequence alignment resulting from step 4. Describe the method by which you built a phylogeny.
6. 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 (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.

Due date

Day 7, Week 9