Assignment: Evolutionary Tools for Infectious Disease Analysis

A 15-year-old child with no past medical history presents at your hospital with acute febrile illness (38.7C), joint pain, rash, and complaining of headache. He was recently returned from a family trip to Brazil to attend a large sporting tournament. Viral infection was suspected by the attending physician, who sent a blood sample to the lab for RNA extraction and non-specific amplification for next-generation sequencing. The patient's fever and other symptoms have since resolved; however, other members of the family have reported similar symptoms and the public health officer requests your support for an enhanced follow-up. Unfortunately, the laboratory has lost the NGS data, which had been filtered for human mRNA and any ribosomal RNA sequences – the following is the only sequence the technician was able to recover:

GTTCACGTGCTCCAGGGAGATGACTGGGAAGAGCATTCAGCCGGAGAACCTG
GAGTATCGGATAATGCTATCAGTGCATGGCTCTCAGCACAGTGGGATGATTGTTAATGATGAAAACAGAGCAAA
GGTCGAGGTTACGCCCAATTCACCAAGAGC
AGAAGCAACCTTGGGAGGCTTTGGAAGCTTAGGACTTGATTGTGAACCAAGGACAGGCCTTGACTTTTCAGATC
TGTATTACCTAACCATGGATAACAAGCATTG
GTTGGTGCACAAAGAGTGGTTTCATGACATCCCATTGCCCTGGCATGCTGGGGCAGACACTGGAACTCCACATTG
GAACAACAAGGAGGCATTAGTGGAATTCAAGGACGCCCACGCCAGAGGAACCGTCGTGGCTTTGGGGAGCCAG
GAAGGAGCCGTCCACACGGCTCTTGCTGGAGCTCTAGAGGCTGAGGTGGATGGTGCAAAGGGAAGGCTATCCT
CTGGC

1. Using NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), determine a potential cause of the infection based on this sequence.

2. Retrieve all published sequences from the same virus and genomic region using a Genbank query. (Hint: go to http://www.ncbi.nlm.nih.gov/nuccore and enter the pathogen name and gene; e.g., 'measles virus hemagglutinin'). Exclude full length genome sequences from this query set by clicking on 'Custom range. . . ' under the heading 'Sequence Length' in the left column, set this range to 1 to 1000, and click 'Apply' Download the resulting sequences by clicking the 'Send to:' button, choose 'File' as Destination and 'FASTA' as Format.

3. Generate a multiple sequence alignment by pasting or uploading this FASTA file to http://www.ebi.ac.uk/Tools/msa/muscle/. Select 'Pearson/FASTA' as the output format. Download the

result as separate file.

4. Paste the aligned sequences into the input field at https://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny/ and run the Neighbor Joining analysis with default parameters. Save the tree string to another file.

5. View and print your tree. I recommend downloading Fig Tree from: http://tree.bio.ed.ac.uk/software/figtree/ and viewing/colouring your tree as you wish (e.g. by country of sequence origin) before printing. There also online tree viewers which allow you to view and print your tree e.g. http://etetoolkit.org/treeview/

Discussion Paragraph

In one or two paragraphs write a brief discussion of your results and explain how phylogenetics could be useful to discover the source of this pathogen. Considering other related potential causes of infection why might the concept of “Original Antigenic Sin” be important in predicting the outcome for a patient previously infected with the agent you identified in 1. to 5. above?