MDSC308 Phylogenetics

0) Human Immunodeficiency Virus (HIV) has two main varieties – HIV1 and HIV2. Both of the types lead to the same symptoms, but their genetic origin is distinct. The virus is closely related to Simian Immunodeficiency Virus (SIV) which infects apes.

In this exercise, you will build a phylogenetic tree with HIV sequences to investigate the origin of HIV. Provided here are gp120 sequences, the envelope glycoprotein important in CD4+ cell binding.

An alignment file is provided with this exercise – hiv.phy. It contains the aligned sequences of 27 samples of HIV1, HIV2 and SIV, collected from various studies. The name of the file identifies the type of the virus and the geographic origin of the sample.

Name	Species	Virus name	Region of origin
HV1W_	Homo sapiens	Human Immunodeficiency Virus type 1	United States, West Coast, mixed ethnic origin
HV1E_	Homo sapiens	Human Immunodeficiency Virus type 1	United States, East Coast, mixed ethnic origin
HV2W_	Homo sapiens	Human Immunodeficiency Virus type 2	United States, West Coast, mixed ethnic origin
HV2E_	Homo sapiens	Human Immunodeficiency Virus type 2	United States, East Coast, mixed ethnic origin
SIVC_	Pan troglodytes (Chimpanzee)	Simian Immunodificiency Virus	Democratic Republic of Congo
SIVM_	Macaca mulatta (Rhesus macaque)	Simian Immunodificiency Virus	India

0.5) In this part, you will build a phylogeny using the sequence data provided. This involves a bit of clicking around the web. We do not mean to test your technical prowess – but we want you to think about phylogenies. So if you have any questions about this part, ask your TAs.

Go to http://www.atgc-montpellier.fr/phyml/, the online interface to PhyML. PhyML is a software package that implements multiple likelihood phylogeny reconstruction by maximum likelihood.

a) Submit the input alignment via the Sequences button (first row - button "Browse...").
b) Select "Amino-Acid" data-type (second row).
c) Leave the rest of the options the same.
d) Name the analysis and type in your email address
e) Hit "Execute & email results"

The phylogeny building might take a few minutes. You will receive an email from **LIRMM bioinformatics**, which will have hiv_phy.phyml.zip attachment. The attachment is a zip archive, you will need to download it and extract it. Ask your TAs for help if you don't know how.
The archive contains 4 files, but we are only interested in hiv_phy_phyml_tree.txt.
Go to http://ivan-kryukov.github.io and upload hiv_phy_phyml_tree.txt to it.
Alternatively, you can open the file in FigTree if you have it on your computer (it is installed on the lab computers)

Study the tree carefully.

1) Click "Download Tree" and insert the picture below. (2 marks)

2) Does the phylogeny distinguish between the types of HIV? In other words, if I gave you an HIV sequence, will you be able to tell me what type it belongs to easily? Does this agree with our understanding of HIV types? (2 marks)

3) Which of the types of HIV is the SIVC1 (Chimp) sample close to? What does that suggest about the evolutionary history of this type of HIV? (2 marks)

4) Are the east coast isolates genetically distinct from the west coast isolates? In other words, is it possible to trace the infection route for the human subjects? What does this suggest about rates of migration within US? (2 marks)