Primates as natural reservoirs of HIV

I) HIV: an introduction

HIV has caused the death of more than 30 million people, since it first began to spread in 1981. A third of all deaths occur in sub-Saharan Africa, where access to medicine and treatment are reduced.

2) Genetic relationship between HIV and SIV

Two species of HIV infect humans: HIV-1 and HIV-2, which are further divided into subgroups. HIV-1 is more virulent. It is easily transmitted and is the cause of the majority of HIV infections globally. HIV1 is divided into three subgroups, HIV1-M, HIV1-N, and HIV1-O, of which HIV1-M is the most prevalent and is spread around the world.

HIV1 and HIV2 are distinguished by their genome.

Question: Where did the disease originate from?

Data suggests that it was transmitted to humans via other primates. The close genetic relationship between humans and primates makes it likely for viruses to be transmitted across species. This can potentially have devastating consequences.

Non-human primates have been identified as carriers of HIV-like viruses, called SIV (simian immunodeficiency virus). The viruses that non-human primates carry are not known to cause any disease in their hosts. They are called asymptomatic carriers.

Reconstructing the phylogenetic relationships among strains of HIV and related viruses from African primates has made it possible to elucidate the origin of the epidemic in humans. By comparing the sequences of HIV-like viruses in other primate species, one can retrace which species most likely transmitted the virus to humans.

The aim of this exercise is to study the phylogenetic relationship between SIV and HIV-1, HIV-2 sequences. We will then determine which sequences are closely related to one another.

II) Analysis of the data

You have been two input files containing protein sequences from the gp120 and pol proteins, respectively.

Multiple alignment and phylogeny command line

1. Log into the server using your username and password eg:
   ssh user12@192.168.1.129

2. Move to the Phylogeny folder by:
   cd Phylogeny

3. Run the alignment program as below

   emma
   Multiple sequence alignment (ClustalW wrapper)
   Input (gapped) sequence(s): g20.fasta
   (aligned) output sequence set [hv1a2.aln]: g20.aln
   Dendrogram (tree file) from clustalw output file [hv1a2.dnd]: g20.dnd

4. Run the distance matrix program fprotdist

   fprotdist
   Protein distance algorithm
   Input (aligned) protein sequence set(s): g20.aln
   Phylip distance matrix output file [g20.fprotdist]:

5. Run the neighbor joining program fneighbor

   fneighbor
   Phylogenies from distance matrix by N-J or UPGMA method
   Phylip distance matrix file: g20.fprotdist
   Phylip neighbor program output file [g20.fneighbor]:

6. Run the consensus tree using fconsense

   fconsense
   Phylip tree file: g20.treefile
   Phylip consense program output file [g20.fconsense]:

Multiple alignment and phylogeny on wEMBOSS

Multiple alignment

We will use the program ClustalW (named emma in wEMBOSS) to make a multiple alignment of the
virus sequences.

1. Download the sequences g20.fasta from the course website
2. To start of you need to create a new project | click on create new project and give it a name.
3. Upload the file gp20.fasta from your computer
4. Load the sequences into ClustalW (wEMBOSS->emma): Alignment->multiple->emma", and select g20.fasta
5. Run emma with default values and save the resulting alignment to your multalign folder.
6.

Phylogenetic tree:

Bootstrapping

1. Open your emma output alignment from before in "wEMBOSS->PHYLOGENY->MOLECULAR SEQUENCE->fseqboot" and run the program with 100 replicates. Save the bootstrapped file by right click
2. Open the emma output alignment from before in "wEMBOSS->PHYLOGENY->MOECULAR SEQUENCE->fprotdist" and run it. Save the output. This step calculates all distanses between the bootstraped sequences.
3. Open the fprotdist output" in "wEMBOSS->PHYLOGENY->CHARACTERS->Distance Matrix->fneighbor" and run it. Browse throught the pop up window and save the treefile. This output contains 100 trees in Newick format.

4. Open the treefile in "wEMBOSS->PHYLOGENY->CONSENSUS->fconsense" and run it. . Browse throught the pop up window and save the consensus
   treefile.
5. View the bootstrapped tree:
   Load the consensus tree in Figtree. Click the branch labels. This will give you the boostrap values on your tree and therefore the support for each branch.

You can now compare the 2 trees you get.

You also have a third input file with an outgroup sequence, repeat the same exercise with this file and root the tree with the outgroup sequence (HTLV).