The Florida Dentist case: An analysis using Mesquite

In the late 1980’s a cluster of “mystery” cases of HIV infection were discovered in a small community in rural Florida. The mystery infections were a set of individuals (seven in our study) each of whom was HIV positive but lacked in any risk-factors (hemophilia, intravenous drug use, history of unprotected sex). The only commonality found among these individuals was that they all had the same (HIV positive) dentist.

Two hypotheses were proposed: (1) The mystery cases were each infected independently, perhaps via mosquitoes. (2) The mystery cases all acquired HIV from the dentist. We will use molecular data to distinguish these hypotheses.

Look at the data for a sample of viruses

1. Install Mesquite on your computer (www.mesquiteproject.org)
2. Launch Mesquite.
3. Open the file “Viral Backbone” in Mesquite.

The data here are sequences of the DNA of one of the genes in two individual viruses from each person. All individuals in the matrix have known risk factors.

Outgroup = An individual from a different geographical region – helps root the tree
LCa-LCd = Four “local controls” – individuals from this community who are HIV positive
Dentist = The dentist.

The DNA sequences have been “aligned” to correct for some cases where a few nucleotides have been added or lost by some viruses.

Look at the data matrix and discuss some of the patterns you see.

Estimate the “backbone” phylogeny

Look at the default tree in the tab: “Tree Window showing Stored Trees.”

You will see to the right the “treelength.” This is the total number of steps needed to account for the evolution of all the positions in the DNA sequences. A good estimate of the tree for these sequences is the tree that minimizes the treelength. You could try to find this by moving branches around by hand, but it is easier to have the computer search for this tree.

The tool we want is this

After selecting this tool click on the base of the “ingroup” (the branch above the point where the outgroup diverges). Of the various criteria select “Treelength.” Look at the tree
and see how it changed. Are the two viruses from a single patient always “sisters?” Is this surprising?

Add the mystery cases to the data matrix

Open one of the Viral Mystery Files in Mesquite. You will see that it contains the same data as before with the addition of one of viral sequences from one of the mystery cases.

Open a random initial tree. From the “Taxa&Trees” pull-down select “New Tree Window” and “Default Trees.”

Rearrange the tree so the two outgroup sequences are positioned as outgroups. Then click the “Search Within Clade” tool on the branch at the based of the ingroup and do a search based on treelength (using SPR).

Look at the optimal tree. Is it consistent with the mystery patient being infected by the dentist? Can you rule out mosquito mediated infection?