Middlesex University VAST 2010 Challenge
Genetic Sequences – Tracing the Mutations of a Disease

Authors and Affiliations:

Ian Mitchell, School of Engineering and Information Sciences, Middlesex University, UK, i.mitchell@mdx.ac.uk

Peter Passmore, School of Engineering and Information Sciences, Middlesex University, UK, p.passmore@mdx.ac.uk

Kai Xu, School of Engineering and Information Sciences, Middlesex University, UK, K.Xu@mdx.ac.uk [PRIMARY Contact]

Tool(s):

ClustalW2 (http://www.ebi.ac.uk/Tools/clustalw2/index.html) was used to align sequences and ClustalW2 and TreeView (taxonomy.zoology.gla.ac.uk/rod/treeview.html) were used to draw phylogenetic trees. Analysis was also aided by using Microsoft Excel and some hand coded programs.

Video:

Replace this line by a link to your video.
Middlesex-vast2010-MC3.wmv

ANSWERS:

MC3.1: What is the region or country of origin for the current outbreak? Please provide your answer as the name of the native viral strain along with a brief explanation.

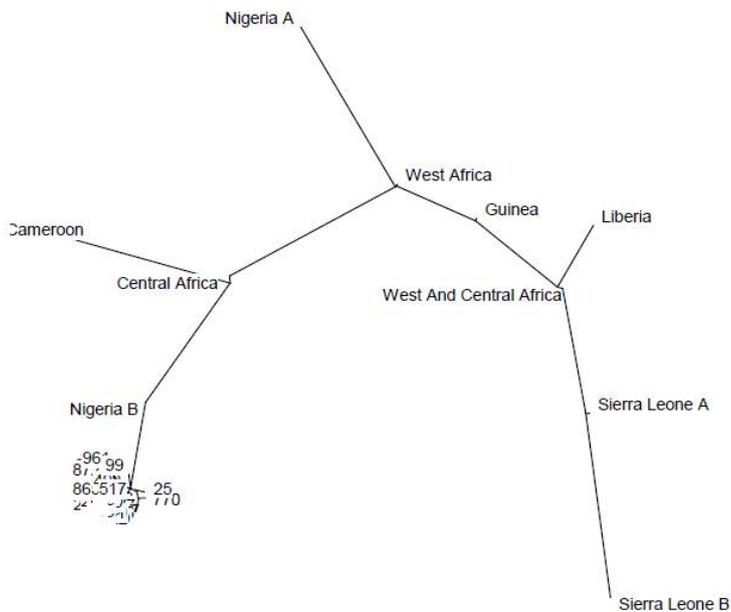Nigeria_B see phylogram in Figure 3.1 [1]

Figure 3.1 Phylogram of the combined native and current sequences. The radial tree shows that Nigeria_B is the origin of the current outbreak sequences.

MC3.2: Over time, the virus spreads and the diversity of the virus increases as it mutates. Two patients infected with the Drafa virus are in the same hospital as Nicolai. Nicolai has a strain identified by sequence 583. One patient has a strain identified by sequence 123 and the other has a strain identified by sequence 51. Assume only a single viral strain is in each patient. Which patient likely contracted the illness from Nicolai and why? Please provide your answer as the sequence number along with a brief explanation.

The phylogram in Figure 3.2 clearly shows that sequence 123 is a direct descendent of sequence 583. Whereas, sequence 51 is neither a direct ascendent or descendent of 583. Closer inspection of the sequences reinforced the above findings – sequence 583 is a single mutation from sequence 123, whereas sequence 583 is three mutations from sequence 51.
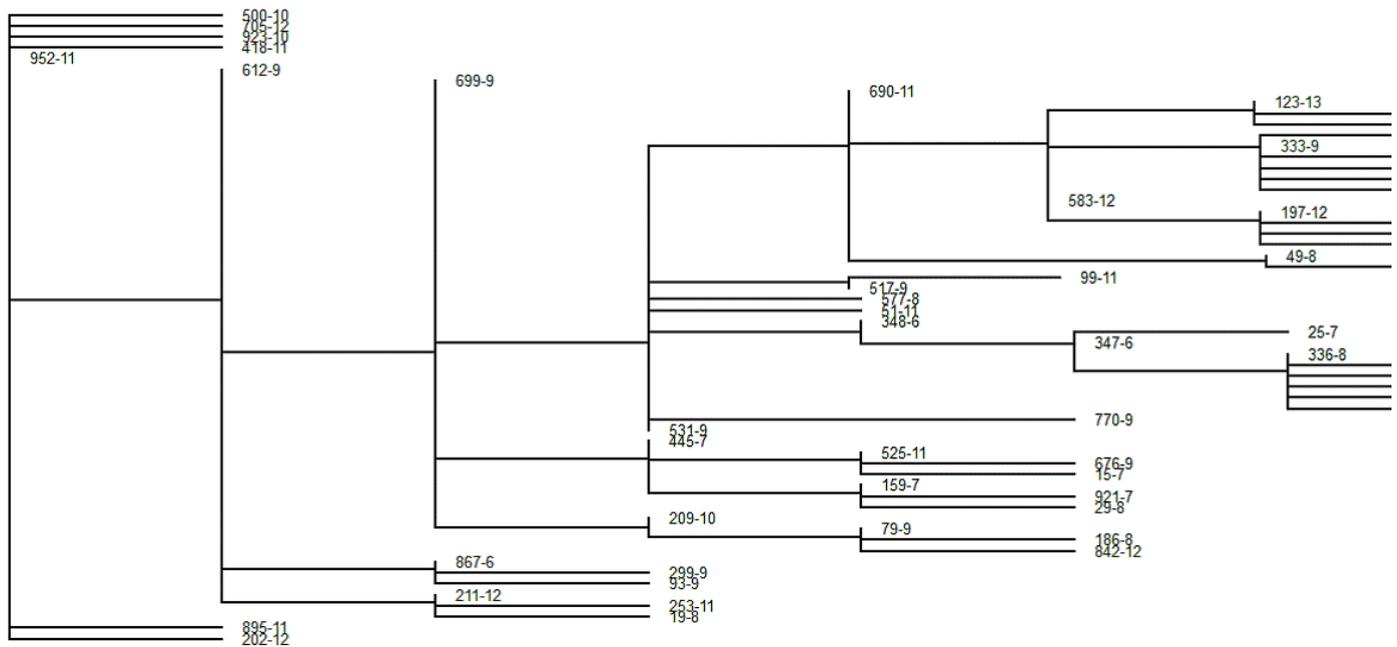
Figure 3.2 Phylogram of outbreak sequences.

MC3.3: Signs and symptoms of the Drafa virus are varied and humans react differently to infection. Some mutant strains from the current outbreak have been reported as being worse than others for the patients that come in contact with them.
Identify the top 3 mutations that lead to an increase in symptom severity (a disease characteristic). The mutations involve one or more base substitutions. For this question, the biological properties of the underlying amino acid sequence patterns are not significant in determining disease characteristics.
For each mutation provide the base substitutions and their position in the sequence (left to right) where the base substitutions occurred. For example,
C → G, 456 (C changed to G at position 456)
G → A, 513 and T → A, 907 (G changed to A at position 513 and T changed to A at position 907)
A → G, 39 (A changed to G at position 39)

A->G, 223 (sequence 612 classified as Mild symptoms changes to sequence 952, classified as severe symptoms)

A->C, 269 and C->A, 494 and A->T, 843 (sequence 51, classified as Mild, changes to sequence 99, classified as Severe)

T->C, 311 and A->T, 946 and A->G, 1087 (sequence 49, classified as Mild, changes to sequence 583, classified as Severe)

There are many other mutations that include combinations of the above, with a majority including mutation at position 223.

MC3.4: Due to the rapid spread of the virus and limited resources, medical personnel would like to focus on treatments and quarantine procedures for the worst of the mutant strains from the current outbreak, not just symptoms as in the previous question. To find the most dangerous viral mutants, experts are monitoring multiple disease characteristics.
Consider each virulence and drug resistance characteristic as equally important. Identify the top 3 mutations that lead to the most dangerous viral strains. The mutations involve one or more base substitutions. In a worst case scenario, a very dangerous strain could cause severe symptoms, have high mortality, cause major complications, exhibit resistance to anti viral drugs, and target high risk groups. For this question, the biological properties of the underlying amino acid sequence patterns are not significant in determining disease characteristics.
For each mutation provide the base substitutions and their position in the sequence (left to right) where the base substitutions occurred. For example,
C → G, 456 (C changed to G at position 456)
G → A, 513 and T → A, 907 (G changed to A at position 513 and T changed to A at position 907)
A → G, 39 (A changed to G at position 39).

Each characteristic of the virus is equal to 1. Add up all the characteristic for each of the strains. Add these numbers to the phylogram – see figure 3.2. Identify lowest and highest scores and search the tree for the shortest paths between them. The top 3 are provided below:

#	Symptom	Mortality	Complication	Drug Resistance	At Risk
867	Mild	Low	Minor	Intermediate	Low
211	Moderate	High	Major	Resistant	Medium

Table 3.4.1 Characteristics of sequences 867 and 211. Two mutations result in five changes in characteristics.
G → T, 720 and A → G, 821 – Mutations transform sequence 867 to 211. From Table 3.4.1 it can be seen that these mutations result in a viral strain with a high mortality rate.

#	Symptom	Mortality	Complication	Drug Resistance	At Risk
867	Mild	Low	Minor	Intermediate	Low
952	Severe	Medium	Major	Resistant	Low

Table 3.4.2 Characteristics of sequences 867 and 952. Two mutations result in five changes in characteristics.
A->G, 223 and G->T, 720. Mutations transform sequence 867 to 952. From Table 3.4.2 it can be seen that these mutations result in a viral strain with severe symptoms and medium mortality rate.

#	Symptom	Mortality	Complication	Drug Resistance	At Risk
333	Moderate	Low	Minor	Resistant	Medium
501	Severe	High	Minor	Resistant	High

Table 3.4.3 Characteristics of sequences 333 and 501. One mutation results in three changes in characteristics.
G-> C, 848. This mutation transforms sequence 333 to 501. From Table 3.4.3 it can be seen that this mutation results in a viral strain with severe symptoms and high mortality rate.

There are other combinations that can transform a sequence as in Table 3.4.3, however these transformations require more mutations, whereas the mutation in Table 3.4.3 requires only a single mutation.

References:
[1] Higgins,D.G. and Sharp,P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5,151-153.