Project3 Write-Up

William Marshall[1]

I. Question 1

A. Number of genetically identical strains

The several years that the patient does not seek treatment would allow the single strain of HIV to mutate into several different strains, and as such the patient would contain fewer genetically identical HIV strains. Virii are notorious for their fast mutation rate, which is mainly attributed to two factors. Firstly, the sheer number of virii in an infected patient – on the order of hundreds of billions – mean that mutations occur frequently, even if the mutation rate is low, because of how rapidly the virus copies itself. Secondly, RNA has a higher propensity to mutate than DNA does, since it lacks the error correction that DNA benefits from.

In the end, the single strain of HIV the patient was exposed to would turn into several strains, as shown in the top histogram of figure 1. For readability's sake, we say that the strain only mutates into six strains, when in reality there could potentially be many more.

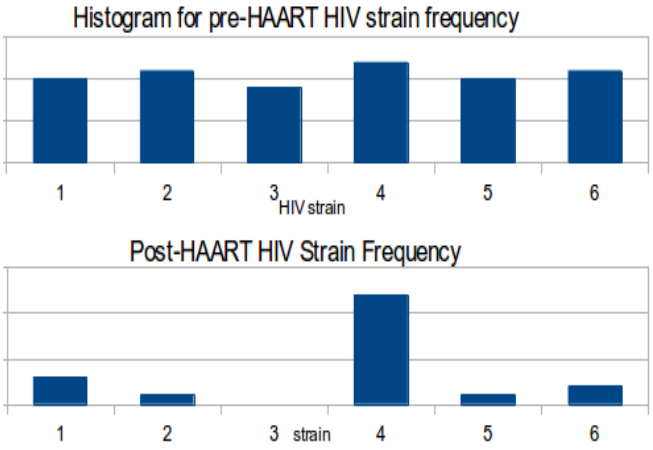

Fig. 1. Histograms that show the difference between the relative frequencies unique HIV strands before and after HAART was applied to the patient. Note that strand 4's reverse transcriptase is not susceptible to AZT as a reverse-transcription inhibitor.

B. Number of strains after Zidovudine

Zidovudine (AZT), is an anti-retroviral drug that can be used to interrupt the HIV lifecycle. Specifically, AZT interferes with the function of HIV's reverse transcriptase enzyme which creates DNA from the virus's RNA. The chemical structure of AZT if very similar to Thymidine, one of the primary nucleosides used in this reverse transcription process, and is accidentally taken in by the reverse transcriptase. When this occurs, reverse transcription stops and the HIV virus cannot insert its DNA into the host cell.

If the patient was fortunate enough to receive a treatment of Zidovudine, then the number of unique HIV strains would likely decrease. Figure 1 represents the differences between viral strain frequency in a patient pre and post Zidovudine treatment. Before treatment, there are six strains, each with similar relative frequencies. After the Zidovudine treatment, the frequency of the strains whose reverse transcriptase was susceptible to AZT decreases significantly, or even disappear completely (as in the case of Fig. 1 strain 3). However, it is possible that the patient's HIV has mutated into a strain that is not inhibited by AZT, so its frequency is unaffected. Yet even if this happens, and the patient isn't cured, Zidovudine is still an effective treatment because it decreases the overall number of HIV viruses, to the point where sometimes HIV does not show up in a blood test.

C. Phylogenetic tree after AZT

Its likely that the HIV strains that are not affected by AZT all occupy a such a subtree within the phylogenetic tree of the strains that are affecting the patient. As such, the phylogenetic tree of the patient's HIV strains after the AZT treatment would likely consist mostly, if not entirely, of that subtree. What the surviving virii all have in common are genes that code for a reverse transcriptase enzyme that does not confuse Zidovudine for Thymidine; a property that could manifest itself as a region of conserved amino acids in the reverse transcriptase gene.

D. Parallel administration of HAART

In practice, Highly Active Anti-Retroviral Therapy (HAART) does not simply apply AZT by itself, it consists of a cocktail of different anti-retroviral treatments. HAART has taken this "shotgun" approach because it has been shown that the treatments are more effective when applied in parallel, rather than serially.

To understand why this is true, consider the phylogenetic tree of strains that are affecting the patient. If the treatments are applied in a one-at-a-time manner, then it is more likely that – at each stage in the treatment – there is a subtree of HIV strains that is immune to the treatment. If the treatments are applied all at once, then the probability that there is a surviving strain is much lower since it would have to be simultaneously immune to all treatments; this would be represented by a significantly smaller subgraph. The result of the latter approach is that the patient would likely display an overall lower concentration of HIV following treatment.

[2]Lehigh University, USA wcm214@lehigh.edu

E. Medicinal noncompliance

HAART's effectiveness depends largely on the consistency of the patient's adherence to the medication routine. If a patient is not consistently taking the HAART medication, then it is more likely that the disease will resurface. The reason for this is similar to the reason that it is important to take antibiotics consistently: if given the chance, it is likely that an immune strain of the virus will take a foothold in the patient. Medical noncompliance in HAART has been attributed to a number of factors, but it seems to be strongly correlated with age and psychological status. The younger a patient, the less likely that they have formed a daily routine that they can use to correctly and consistently take the medication.

II. Question 2

It has been shown previously that evolution within a population of similar organisms tends to conserve regions of DNA that code for the functional regions of proteins. If a mutation occurred that radically altered that shape of the protein it coded for, then it's less likely that the change would be retained. As such, there is a tendency for protein shape to be similar among genetically close organisms, and for structural protein features to be passed down. When performing a sequence alignment to determine evolutionary closeness, it might appear that that two organisms are unrelated; however, that analysis ignores the structural information of the proteins. Two proteins that have a low alignment score, yet a very similar structure, are more similar than two protins that have a low alignment score in addition to a different structure. Therefore, incorporating protein shape into sequence alignment methods would yield more accurate phylogenetic trees.

Abstractly, we can make the following claim about the similarity between two strands using protein structure as a feature:

$$DA \wedge DS \leq DA \wedge SS \leq A \wedge SS$$

Where A implies a similar alignment, DA implies a dissimilar alignment, SS implies a similar structure, and DS implies a dissimilar structure, and the $\leq$ operator indicates that strands with the attributes on the left side are less similar than a strands with the attributes on the right side; i.e., two strands that have dissimilar structure and alignments are less similar than two strands that have dissimilar alignments but similar structure.

As previously stated, if these facts are included in the evaluation of specie similarity, then more accurate phylogenetic trees could be created.

III. Question 3

A. Conserved amino acids in a subtree

An amino acid that is conserved within a subtree, but not the tree as a whole could perform a number of different functions. First, it's important to understand that the conserved sequence exists because the subtree is exposed to an evolutionary pressure for which that sequence is beneficial. This is easy to explain if it can be assumed that the organisms of different subtrees live in different environments.

Specifically, the conserved sequence could code for the hair pigment of a chipmunk within a particular environment, say, the snow-covered Alaskan outback. One could imagine that the subtree's conserved sequence codes for white pigment, allowing the chipmunk to blend in with its surroundings.

B. Conserved amino acids in the entire tree

The amino acids that are conserved throughout an entire phylogenetic tree have a tendency to cluster around the active sites of proteins because even a slight sequence change might prevent an enzyme from binding to its substrate, rendering it useless. If such an enzyme were to play a crucial role in sustaining life, such as Ubiquitin, then its ineffective functioning would be a large evolutionary disadvantage. Therefore, there is a large pressure for the enzyme sequences around active sites to be conserved.