Marcantonio Soda Jr
CSE408
Project 3 Report
12 May 2023

Section 1:

Consider a situation where a hypothetical patient were to become infected with a single (arbitrary) strain of HIV. The individual does not later receive any medical treatment for the disease and is never exposed to another different strain of the virus. Logically, it stands to reason that the only subtype of HIV present in the patient should be the only one that he/she was initially exposed to. Interestingly, that assumption is not correct due to the intricacies of HIV reproduction. To fully understand the implications of this complex situation, one must have a moderately more thorough understanding of the HIV virus itself and how it propagates: HIV is a retrovirus, and a characteristic of retroviruses is that they need to convert its RNA into DNA when it enters a given cell for reproductive purposes. This process is carried out by an enzyme called reverse transcriptase, which frequently causes errors. The aforementioned errors result in mutations in HIV's genetic code. This means the virus is constantly changing and evolving within a person's body. So over many years, after the once-unique strain of HIV replicated countless times, it is nearly impossible for it not to have mutated. Therefore, if a bioinformatics researcher were able to sequence the RNA genome present in every HIV capsid in the patient several years after the initial infection, we would likely find a multitude of genetically diverse HIV strains even if the patient were never exposed to a different strain.

Zidovudine, first used in 1987, is an antiretroviral medication that was the first FDA approved drug used to manage and treat HIV-1. It is classified as a nucleoside reverse transcriptase inhibitor, which is a fancy way of saying that it blocks the action of reverse transcriptase, preventing HIV from replicating its RNA to DNA. If the aforementioned patient were to receive a course(s) of Zidovudine after initial exposure, the diversity of the unique HIV strains present in each capsid would be less than if the patient had not received the Zidovudine. This is because the process that would have introduced such diversity is the transcriptase process, which would have been suppressed by Zidovudine. Note that Zidovudine is not 100% effective, meaning that it will not block all transcriptase, thus there will likely still be some diversity over time.

In the case of a patient becoming infected with HIV, not receiving Zidovudine, and never being exposed to another strain of the virus, the phylogenetic tree of HIV genomes would be quite diverse as the virus would have been allowed to reproduce and mutate unchecked (by anything other than the patient’s immune system). In contrast, the phylogenetic tree of the HIV genomes in the hypothetical patient who was treated with Zidovudine would be much less diverse because many strains would be killed off by the

treatment. The surviving HIV strains would share the trait of being resistant to the mechanism Zidovudine employs to inhibit transcriptase. This phenomenon is one of the reasons why phylogenetic trees are such a powerful bioinformatics tool.

HAART stands for Highly Active Antiretroviral Therapy. HAART is a technique in which several antiretroviral drugs are applied in parallel to combat HIV (or any retrovirus, presumably). In this therapy, multiple drugs are used at once in an effort to combat the high mutation rate of HIV. It is important to note that the various prescribed drugs do not all do the same thing. Each attach a different pivotal mechanism employed by HIV for reproduction. If the drugs were applied in series rather than in parallel, the success rate would likely be less due to the fact that, in order to survive, the virus will only have to develop a resistance to one drug at a time rather than all the drugs at once.

HAART requires a strict adherence to a daily medication regimen. If the patient does not follow this strict procedure until the course is complete, HAART is likely to fail. This is an example of medical noncompliance. There are many reasons for medical noncompliance. Some people can't afford the various medications required by HAART. Some people forget to take their medication. Some people have side effects and may decide to discontinue the course. In any of these cases, the virus may rebound and become more resistant to the specific prescribed medication regimen. Interestingly, medical noncompliance during HAART may have the unfortunate effect of transmitting a drug-resistant strain of the virus to other people.

Section 2:

Currently, the evolution of proteins is tracked through the comparative study of their corresponding sequences. These observations primarily rely on sequence alignments, which compare the order of amino acids in different proteins to infer evolutionary relationships. However, focusing on sequence alignment alone can be limiting in certain circumstances, especially when comparing proteins with low sequence identity but similar folds because the similarity of folds may have some biological significance.

Incorporating protein structure into the alignment process is likely to provide a more comprehensive picture of evolutionary history. This is because proteins are not two-dimensional like many state-of-the-art sequence alignment softwares infer (this inferences is necessary for simplicity). The three-dimensional shape, or fold, of a protein is fundamentally tied to its function in the cell. Even if the sequences of two or more proteins differ significantly, a similar fold can suggest a common evolutionary origin. This is due to the fact that the forces of natural selection tend to conserve protein function. A protein's function is certainly attributed to its nucleotide sequence, but also to its shape. It is therefore limiting to only focus on sequence alignment when performing these sorts of analysis. The fact that function is tied to the sequence and fold of a protein may lead to the conservation of overall protein fold even when the underlying sequences diverge significantly.

By considering protein shape in sequence alignment, we can identify and track these functional and evolutionary relationships that might otherwise be missed. For example, structural alignment can highlight conserved patterns that are critical for maintaining the protein's shape and function, even if those patterns are not in the same sequence position. It can also reveal instances of convergent evolution, where different sequences have evolved to form the same fold due to similar functional demands. Moreover, structural information can help to resolve ambiguities in sequence-only alignments, providing a more reliable basis for inferring evolutionary relationships. For instance, in cases where sequence similarity is low, structural alignment can provide evidence of common ancestry that would be hard to detect from sequences alone. Very interesting.

Section 3:

In the context of a phylogenetic tree, the conservation of an amino acid within the members of a subtree, but not within the whole tree can be of great functional significance. This type of conservation indicates that the particular amino acid has a specific role that is crucial for the subset of organisms within that subtree, but not necessarily important for all organisms represented in the entire phylogenetic tree. This could be attributed to a specific evolutionary adaptation or functional requirement unique to that subgroup. For example, the conserved amino acid might be involved in a unique interaction between proteins, a specific enzymatic activity, or it might be responsible for a fold critical to the function of the protein within the organisms of that subtree. Additionally, it could also signify a particular adaptive response to a unique environmental pressure experienced by the organisms within the subtree. This type of variation and conservation highlights the concept of divergent evolution, where different species evolve different traits due to different selective pressures, leading to the creation of biodiversity.

The most conserved amino acids are those that are conserved throughout the entire tree or vary little between branches. These amino acids have been observed to cluster spatially near active sites, i.e. regions on the molecular structure that are essential for some chemical function. The reason for this conservation is twofold. Firstly, these amino acids are likely to be directly involved in the protein's function, such as catalyzing a chemical reaction or binding to a specific molecule. Changing these amino acids could significantly affect the protein's function, potentially rendering it nonfunctional, which would likely be detrimental to the organism's survival. Secondly, these amino acids might play a crucial role in maintaining the structural integrity of the protein. Alterations in these key structural components could lead to a misfolded protein, which could again be detrimental to the organism's survival.