Norman Zvenyika
CSE 308
Project 3 Implementation Report
Spring 2023

1. **Recidivism was a critical problem in the early treatment of HIV, until the development of Highly Active Anti-Retroviral Therapy (HAART), and remains a problem due to the prevalence of medicinal noncompliance. (40%)**

    a) **Imagine a hypothetical patient infected by one strain of genetically identical HIV capsids. The patient does not seek treatment for several years after the initial infection, and is never exposed to another strain of HIV. Following these years, if we could sequence the RNA genome in every HIV capsid in the patient, would we find fewer, more, or the same number of genetically identical HIV strains in this patient?**

        1. In the hypothetical scenario where a patient is infected by one strain of genetically identical HIV capsids and does not seek treatment for several years, we would find more genetically diverse HIV strains in this patient. This is because HIV has a high mutation rate, which contributes to the generation of genetic diversity within the viral population. The reverse transcriptase enzyme, responsible for converting the virus's RNA genome into DNA, is error-prone and lacks proofreading ability, leading to the accumulation of mutations over time.

    b) **Suppose it is the late 80s, and the patient is one of the fortunate few to receive a course of Zidovudine. What is Zidovudine? If we if we could sequence the RNA genome in every HIV capsid in the patient, how would the number of strains afflicting this patient differ from situation (a), above?**

        1. Zidovudine, also known as AZT, is an antiretroviral medication used to treat HIV infection by inhibiting the reverse transcriptase enzyme, which is necessary for viral replication. In the late 80s, if a patient received a course of Zidovudine, the number of HIV strains in the patient would likely be reduced but not eliminated. Zidovudine might suppress the replication of some strains, but the high mutation rate of HIV could lead to the emergence of drug-resistant variants. As a result, the viral population in the patient would still be diverse, albeit with a potentially lower overall level of diversity compared to the untreated scenario in part (a).

c) Compare and contrast, approximately, a phylogenetic tree (there is no need to illustrate) of HIV genomes within this patient before and immediately after the treatment. What property do surviving virii have in common?

1. A phylogenetic tree of HIV genomes within the patient before treatment would likely show a wide range of genetic diversity due to the accumulation of mutations over time. Immediately after Zidovudine treatment, the phylogenetic tree might show a reduction in overall diversity, with surviving viruses sharing a common property: resistance to Zidovudine. The resistant viruses would have specific mutations in their reverse transcriptase enzyme that allow them to replicate despite the presence of the drug.

d) Why does HAART apply several drugs at once? How would the phylogenetic tree of RNA genomes differ if the drugs were applied sequentially, rather than in parallel?

1. HAART applies several drugs at once to combat the development of drug resistance in HIV. By targeting multiple steps in the viral replication process simultaneously, HAART makes it more difficult for the virus to develop resistance to all the drugs in the regimen. If the drugs were applied sequentially rather than in parallel, the phylogenetic tree of RNA genomes would likely show a more stepwise pattern of resistance development. As the virus develops resistance to one drug, treatment with the next drug would select for another set of resistant viruses, potentially leading to the emergence of multi-drug resistant strains over time.

e) If HAART is so effective, why is medicinal noncompliance a problem?

1. Despite the effectiveness of HAART, medicinal noncompliance remains a problem for several reasons. First, the complex dosing schedules and potential side effects of HAART regimens can make it challenging for patients to adhere to their treatment plans. Second, the social stigma associated with HIV can cause patients to avoid taking medications in public or to conceal their diagnosis from friends and family, leading to missed doses. Finally, access to care, financial constraints, and mental health issues can also contribute to noncompliance. Medicinal noncompliance can lead to treatment failure, viral rebound, and the development of drug-resistant strains, further complicating the management of HIV infection.

2. **Proteins with very different sequences (e.g. 25% sequence identity) have been known to exhibit very similar "folds", or overall shape, while proteins with very different folds never exhibit very similar amino acid sequences. Abstractly, how might incorporating protein shape into the alignment of protein sequences, especially for very different sequences, yield additional information for reconstructing evolutionary history? (20%)**

    a) Incorporating protein shape or structure into the alignment of protein sequences can provide valuable insights for reconstructing evolutionary history, particularly when dealing with proteins that have low sequence identity but similar folds. This approach is essential because the three-dimensional structure of a protein often plays a more significant role in its function than its primary sequence. Even when proteins have very different amino acid sequences, convergent evolution, or the independent evolution of similar structures and functions in distantly related organisms, can lead to the conservation of similar folds.

    When comparing protein sequences with low identity, relying solely on sequence alignment can result in limited information about the evolutionary relationship between these proteins. By incorporating structural information, researchers can identify conserved structural features that may indicate similar functions, shared evolutionary origins, or common constraints that led to the conservation of specific folds. In some cases, these conserved structural features can provide additional evidence for homology, or common ancestry, between seemingly unrelated proteins.

    Integrating protein shape into sequence alignment can also help to identify functionally relevant residues, such as those involved in catalysis, ligand binding, or protein-protein interactions. These residues may be conserved across distantly related proteins with similar functions, even when their overall sequences are quite divergent. By identifying these functionally conserved residues, researchers can gain a better understanding of the evolutionary pressures that shaped the proteins' structures and functions.

3. **Given a phylogenetic tree (arranged vertically, so that leaves are at the bottom and the root is at the top) that relates a representative set of homologs that evolve at a similar rate, you can draw a horizontal line across the tree at any point in its height. Subtrees below a point of intersection conserve properties within the subtree that differ from those of other subtrees. If the line is drawn just below the root, fewer, though more complex, subtrees are found, while if the line is drawn just above the**

leaves, perhaps the only subtrees found are individual leaves. Wherever the line is drawn, a set of amino acids is conserved among the members of each subtree. (40%)

a) What might be the functional significance of an amino acid that is conserved within the members of a subtree, but not conserved in the whole tree?

1. The functional significance of an amino acid that is conserved within the members of a subtree but not conserved throughout the whole phylogenetic tree can be attributed to the unique properties and evolutionary history of the proteins within that subtree. Several factors might explain the conservation of such an amino acid.

   Functional specialization is one possibility. The conserved amino acid may enable functional specialization among the proteins within the subtree. This amino acid could be critical for a specific function, such as catalytic activity or substrate binding, unique to the proteins within that subtree. The conservation of this amino acid could result from evolutionary adaptations to specific environmental conditions or biological roles, leading to its retention within that subtree.

   Another explanation for the conservation of an amino acid within a subtree is its contribution to the structural stability of the proteins in the subtree. Proteins sharing similar folds or structures might require particular amino acids at specific positions to maintain their overall conformation and stability. In this context, the conserved amino acid may not be directly involved in the protein's function but is essential for preserving the protein's structure and stability under physiological conditions.

   In some cases, the conserved amino acid may play a role in protein-protein interactions, signaling pathways, or other regulatory mechanisms specific to the members of the subtree. Proteins within the subtree might have evolved to interact with particular partners or respond to specific signals in a way that requires the conservation of that amino acid. This conservation would ensure the proper functioning of these interaction networks and maintain the biological role of the proteins within the subtree.

   Lastly, the conservation of an amino acid within a subtree might be the result of neutral evolution or genetic drift. While this scenario is less likely for conserved amino acids with a functional or structural role, it is possible

that some amino acids are conserved within a subtree due to the random processes of mutation and selection that occur during evolution.

b) The most conserved amino acids (e.g. those conserved throughout the tree, or those that vary very little between branches) have been observed to cluster spatially near active sites: regions on the molecular structure that are essential for some chemical function. Why might this be the case?

1. The observation that the most conserved amino acids cluster spatially near active sites in proteins can be attributed to several reasons, primarily the critical role these amino acids play in protein function and structure.

   Firstly, conserved amino acids near active sites are often directly involved in the protein's function. Active sites are regions on the molecular structure essential for chemical functions such as catalytic activity, ligand binding, or protein-protein interactions. The high conservation of amino acids in these regions is crucial because mutations at these positions could lead to a significant loss of function or even complete inactivation of the protein. Therefore, strong evolutionary pressure exists to maintain these amino acids to ensure the protein's proper function.

   Secondly, conserved amino acids near active sites contribute to the structural stability of the protein. These amino acids interact with other amino acids and the surrounding environment, maintaining the overall fold and conformation of the protein. In some cases, conserved amino acids form essential hydrogen bonds, salt bridges, or hydrophobic interactions that stabilize the active site's structure. This stability ensures that the active site remains in the proper conformation to carry out its function.

   Additionally, the conservation of amino acids near active sites can also be related to their role in allosteric regulation. These amino acids might be involved in transmitting conformational changes from one region of the protein to the active site, affecting its activity. The conservation of such amino acids ensures the proper functioning of regulatory mechanisms and maintains the protein's ability to respond to various signals or conditions.

   Lastly, conserved amino acids near active sites may facilitate the formation of protein complexes or multi-subunit assemblies. In this context, conserved amino acids can be essential for the proper interaction between protein subunits or for the formation of functional protein complexes.