International Journal of Trend in Scientific Research and Development, Volume 1(2), ISSN: 2456-6470
www.ijtsrd.com

AIDS Detection using Genomics Signal Processing Techniques on DNA

Sheela M. G.
Lecturer in Electronics, Govt. Polytechnic College, Pala, Kottayam, Kerala, India

ABSTRACT
The HIV virus causes AIDS, which is one of the deadliest diseases among those who have the infection. Human immunodeficiency virus is referred to as HIV. These infections claimed the lives of tens of thousands of people, many of whom died to AIDS. As a result, early detection of these disorders is critical. Biomedical signal processing in genomics is concerned with the most recent developments in genetics. As a result, it is now simpler to predict one of the most dangerous diseases, AIDS, using multiple GSP techniques. Using the fast Fourier technique and the discrete wavelet transform in our approach, we have been able to anticipate anomalies in the coding region of the HIV-infected cell's gene DNA using the raw genomic data.

Keywords*: DNA, AIDS, HIV, Genomics, GSP*

Introduction
Over the past four decades, the HIV/AIDS pandemic has claimed the lives of more than 30 million people due to the human retrovirus HIV-1. Around 40 million people are now living with HIV around the world, making HIV infection a major public health problem (PLWH). Chronic HIV infection is now under control thanks to long-term antiretroviral medication (ART). PLWH can live long and healthy lives with the help of ART when it is available, but as of this writing, there is no vaccine or treatment for the disease. [1]

The human immunodeficiency virus (HIV) attacks the immune system and reduces people's ability to fight against a wide range of diseases and some cancers. Individuals infected with the virus eventually lose their ability to fight infection as the virus depletes their immune system's cells and functions. The CD4 cell count is commonly used to assess immune system health. Acquired immunodeficiency syndrome (AIDS) is the most advanced stage of HIV infection and can take many years to develop if not treated. [2] It is defined by the development of certain malignancies, infections, or other long-term clinical symptoms.

An organism's development, metabolism, and susceptibility to infection are all under the control of the DNA sequence found in its genome. Genomic signal processing scientists are working hard to decipher it. It is becoming increasingly critical for us to decipher the intrinsic sequence features that are being generated at an exponential rate in entire DNA sequences. Extraction of distinctive segments, discovery of some hidden structural features, differentiation of coding from non-coding areas in DNA sequences, and exploration of DNA sequence structural similarities have all been the subject of several studies. [3] In order to do this, signal processing will play a significant role, and numerous computational approaches, such as the artificial neural network (ANN), nonlinear model, spectrogram, and statistical techniques, have already been utilised.

AIDs can be diagnosed using a variety of tests, including:

- ELISA Test ELISA,

enzyme-linked immunosorbent assay is a method used to detect HIV infection To confirm the diagnosis, the Western blot test is often used in conjunction with an ELISA test. It is important to get tested again if an ELISA test comes out negative, but you believe you have HIV. In chronic HIV infection, ELISA is quite sensitive. However, because antibodies are not formed immediately after infection, you may test negative for several weeks to several months after infection. Even if your results come back negative, you could still be infected if you have a high amount of the virus throughout this time period.

- Home Tests

The only U.S. Food and Drug Administration-approved home test is the Home Access Express Test, which may be purchased at pharmacies. [4]

- Saliva Tests

To collect saliva from your cheek, a cotton pad is utilised. Testing on the pad will take place in a laboratory after it has been placed in a vial. Within three days, the results will be available. A blood test should be performed if the results are positive.

- Viral Load Test

You can find out if you have HIV in your system with this test. Aside from monitoring treatment progress or detecting early HIV infection, it is also utilised for other purposes. The reverse transcription polymerase chain reaction (RT-PCR), branched DNA (bDNA), and nucleic acid sequence-based amplification assay all detect the HIV virus load in the blood (NASBA). These tests follow the same basic ideas. Detection of HIV is based on DNA sequences that bind to those found in the virus. Note that results may vary from test to test. [5]

- Western Blot

A very specific blood test is performed to verify an ELISA positive result.

Review of Literature

The fractal scaling features of DNA sequences were studied by Arneodo et al. [6] using wavelet transform modulus maxima (WTMM). Genes that contain introns and noncoding regions show evidence of long-range association, and the researchers were able to quantify it. DNA walk profiles fluctuated in a way that was consistent with Gaussian statistics, according to the researchers. This finding sheds light on the non-equilibrium dynamic mechanism that generated DNA sequences and the importance of introns and non-coding intergenic regions.

Wavelet analysis was employed in [7] by Trad et al. to identify distinct bands in protein sequences. They were able to compare protein sequences at multiple resolutions using WT's sequence-scale analysis. The previous idea of sequence similarity, which only included local paired amino acid information and ignored information contained in coarser spatial resolution, was enlarged by this "similarity." The complicated sequence alignment processing for sequences was not necessary with this WT-based technique. Consequently, it is possible to compare proteins with varying sequence lengths.

Bioinformatics traditionally uses DNA and amino acid sequence alignment algorithms to compute similarity scores between sequences, whose primary goal is to find parts of successive nucleotide or amino acid sequences that are common in two or more sequences. They are then rearranged to make it easier to see the comparable areas (White et al., 2010). [8]

GSP approaches have previously been proposed by Zhao, Duan & Yau (2011) [9] and Hoang et al. (2015) [10]. (2015). [10]. Because of this, these methods may be limited in their ability to distinguish between two distinct characteristics in a Fourier spectrum, although the raw spectrum may be more powerful in this regard. These studies used a hierarchical clustering algorithm instead of the K-means clustering approach. We can build plots that are distinct from standard dendrograms and simplify the exploration of data thanks to the K-means features.

Objectives

- DNA Sequence Analysis
- Investigate AIDS Detection Methods
- To learn the fundamentals of signal processing in genomics
- To learn about the most basic AIDS diagnostic tests.

Research Methodology

The term "Research Methodology" refers to the study of how a study's methodologies were selected and implemented. Theoretical notions are also included in this discussion, which offers further information concerning the selection and application of procedures. This research relies on data that has already been published in the form of secondary sources. The data gleaned for this research came from a variety of sources, including the Website.

Result and Discussion

Fig. 1 shows a simplified schematic of a DNA molecule that has had its double helix straightened out for clarity's sake. [11]

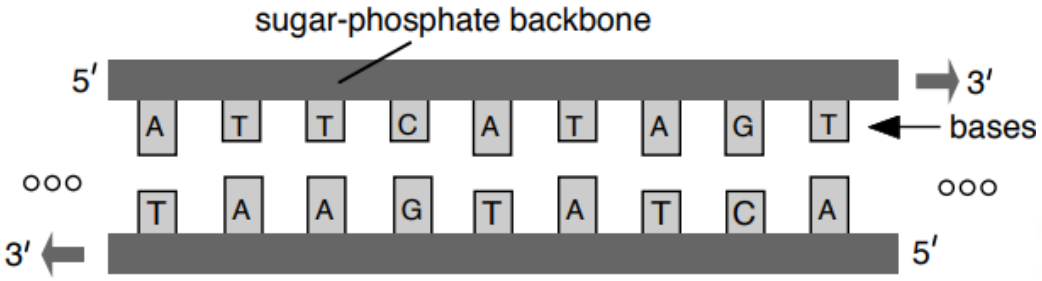
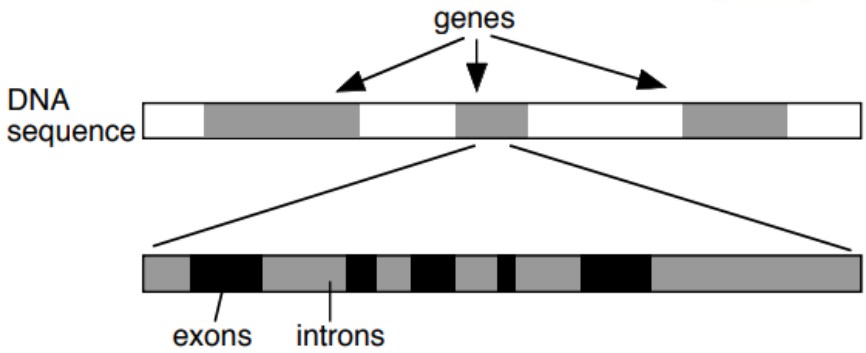


Fig. 1 The double helix of DNA (linearized schematic).

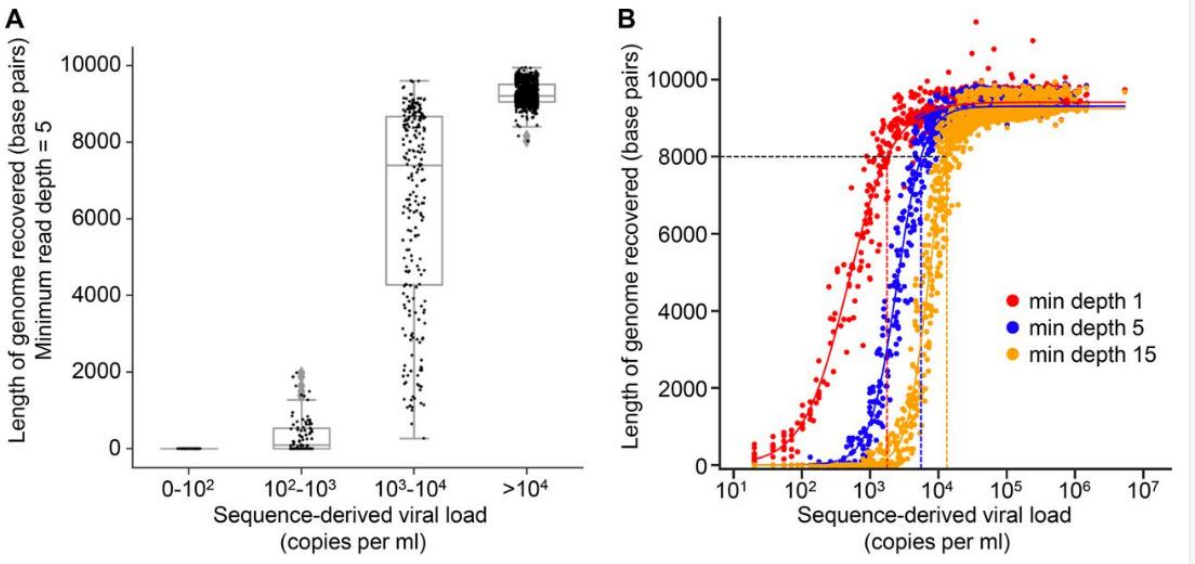
DNA sequences can be broken down into genes and intergenic gaps, as seen in Figure 2. Protein synthesis is carried out by the genes. There is a limited number of genes that are active in each cell family in an organism, even if all cells contain the same set of genes. [12]

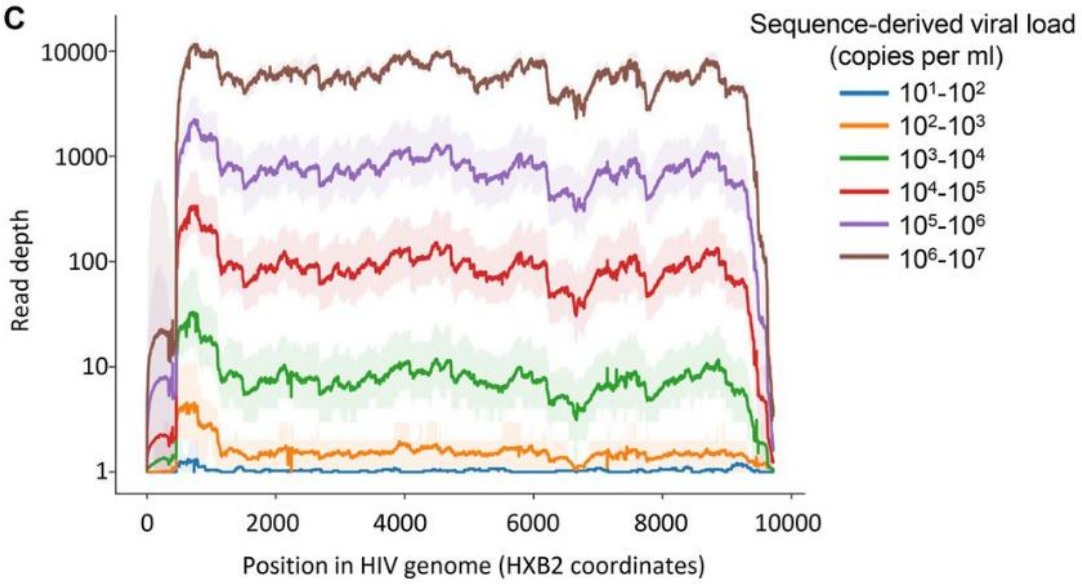


Genes and intergenic areas are shown in Fig. 2. Ecaryotes have exons (protein coding regions) and introns in their genes.

The exons and introns are two subregions of a gene, which for our purposes is a four-base sequence. (Procaryotes, which lack a nucleus, lack introns). Protein coding is limited to exons alone. Groups of three neighbouring bases in the exon region can be visualised. There are three codons in a triplet. [13]

There were samples with viral loads as low as 4,300 copies/ml that did not yield a complete genome, while 97% of those with viral loads greater than or equal to this value did (Fig. 3A). A viral load of more than 1,000 copies/ml is required for most commercially available HIV-genotyping tests. Samples with viral loads between 1,000 and 4,300 copies/ml were found in 6% of this data set; at this range, the average DNA length covered was 4,172 bp (Fig. 3A). [14]





As shown in Figure 3, Viral load has an impact on the success of sequencing. (A) HIV genome lengths reconstructed by Shiver software, using Illumina paired-end reads from samples stratified by log viral load, showed reproducible whole-genome coverage for samples with sequence inferred viral loads of >4 log10 copies/ml and near-complete coverage for the majority of samples with VL between 3 and 4 log10 copies/ml. (B) The intercepts of curves fitted with a sigmoid function illustrate the viral loads at which genome coverage exceeds 8 kb with minimum depth thresholds of 1 read, 5 reads, and 15 reads (after removing PCR duplicates). (C) Samples are categorised by viral load and the median and 95% range of read-depth across the genome are given for all samples.

Greater genome coverage was achieved as a result of increased read depth due to higher viral loads. Figure 4B depicts the relationship between this success rate and sequence-derived viral load more clearly. Between 1,000 and 10,000 copies/ml, sigmoid curves (fitted to data with least-squares) illustrate the viral load thresholds above which at least 8,000-bp genomes tend to be retrieved. Samples with viral loads between 100 and 1,000 copies/ml usually yielded partial genomes (Fig. 3B). [15]

The patterns of read depth were consistent between individuals, with similar patterns of high and low coverage across the genome (Fig. 3C) (Fig. 3C). Importantly, we did not see a drop-off in coverage below five reads to be systematically connected with certain sections of the genome.

Conclusion

Since the catastrophic emergence of HIV, the globe is still suffering from its lethal consequences. As HIV is derived from SIV, so many of the epidemiological, phylogenetic, and genomic properties of HIV are comparable to those of SIV, and this strongly supports the hypothesis of cross-species transmission. ART has profoundly transformed the HIV global epidemiology. Viral genomic sequencing can be used to track the spread of HIV medication resistance, identify optimal antiretroviral regimes, and define transmission dynamics. Despite lowering costs, next-generation sequencing (NGS) is still prohibitively pricey for routine usage in broad HIV epidemics in low- and middle-income countries.

References

[1] Sharma M, Ying R, Tarr G, Barnabas R. Systematic review and meta-analysis of community and facility-based HIV testing to address linkage to care gaps in sub-Saharan Africa. Nature 2015; 528:S77–S85.

[2] Brenner BG, Roger M, Routy JP, Moisi D, Ntemgwa M, Matte C, et al. Quebec Primary HIV Infection Study Group. High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis 2007; 195:951–959.

[3] Powers KA, Kretzschmar ME, Miller WC, Cohen MS. Impact of early-stage HIV transmission on treatment as prevention. Proc Natl Acad Sci U S A 2014; 111:15867–15868.

[4] Peeling, R. W. Diagnostics in a digital age: an opportunity to strengthen health systems and improve health outcomes. Int. Health 7, 384–389 (2015).

[5] Ghani, A. C., Burgess, D. H., Reynolds, A. & Rousseau, C. Expanding the role of diagnostic and prognostic tools for infectious diseases in resource-poor settings. Nature 528, S50–S52 (2015).

[6] A. Arneodo, E. Bacry, P. V. Graves, and J. F. Muzy, "Characterizing long-range correlations in DNA sequences from wavelet analysis," Phys. Rev. Lett., vol. 74, pp. 3293–3296, Apr. 1995.

[7] C. H. Trad, Q. Fang, and I. Cosic, "Protein sequence comparison based on the wavelet transform approach," Protein Eng., vol. 15, pp.193–203, Mar. 2002.

[8] White et al. (2010) White J, Navlakha S, Nagarajan N, Ghodsi M, Kingsford C, Pop M. Alignment and clustering of phylogenetic markers-implications for microbial diversity studies. BMC Bioinformatics. 2010; 11(1):152. doi: 10.1186/1471-2105-11-152.

[9] Zhao, Duan & Yau (2011) Zhao B, Duan V, Yau SS-T. A novel clustering method via nucleotide-based Fourier power spectrum analysis. Journal of Theoretical Biology. 2011; 279(1):83–89. doi: 10.1016/j.jtbi.2011.03.029.

[10] Hoang et al. (2015) Hoang T, Yin C, Zheng H, Yu C, He RL, Yau SS-T. A new method to cluster DNA sequences using Fourier power spectrum. Journal of Theoretical Biology. 2015; 372:135–145. doi: 10.1016/j.jtbi.2015.02.026.

[11] Gray, R. H. et al. Limitations of rapid HIV-1 tests during screening for trials in Uganda: diagnostic test accuracy study. Brit. Med. J. 335, 188 (2007).

[12] Baveewo, S. et al. Potential for false positive HIV test results with the serial rapid HIV testing algorithm. BMC Res. Notes 5, 154 (2012).

[13] J. Zhao, X. W. Yang, J. P. Li, and Y. Y. Tang, "DNA sequences classification based on wavelet packet analysis," in Wavelet Analysis and Its Applications, 2nd Int. Conf., WAA, 2001, pp. 424–429.

[14] D. Dalevi and S. G. E. Andersson, "Discovering the dynamics of microbial genomes," IEEE Eng. Med. Bio. Mag., vol. 20, pp. 55–60, Apr. 2001.