Comparative Genome Analysis using Computational Approach in HIV-1 and HIV-2

Dr. DSVGK Kaladhar[1] and A.Krishna Chaitanya[2]

[1,2]Department of Bioinformatics, GIS, GITAM University, Visakhapatnam. Andhra Pradesh,

Corresponding Address
dr.dowluru@gmail.com

***Abstract*:** Comparative genome analysis are playing important role in genomics, where the scientific studies can provide the relationship of two or multiple organisms. Both HIV-1 and HIV-2 predicted nine genes in entire genome. HIV is a nano particle (sometimes living organism) has molecular information which can use host machinery in the formation of proteins such as gag polyprotein, gag-pol polyprotein, vif, vpr, tat, vpu, gp160, envelope glycoprotein and Nef located in various regions of the genome. Vpu is not predicted in HIV-2. The three dimentional molecules can further be used for the identification of drug targets for the control of diseased molecules located in HIV.

***Keywords*: HIV-1, HIV-2, gene prediction, modelling.**

1. Introduction

Human immunodeficiency virus (HIV) is a retrovirus [1] that causes acquired immunodeficiency syndrome (AIDS), a condition in humans in which the immune system [2], [3] begins to fail, leading to life-threatening opportunistic infections.

The human brain often compared to digital computer [4], [5]. There are various logistic methods to explore the nature of life. Earth is a system which contains power of life. Humans are also be compared to earth with complex system dominating other systems [6].Viruses though have simple system, it has the power in control of complex systems such as plants and animals, and even humans. HIV is one of the powerful machine (have both living and non-living features) [7] have capability to control human cellular and immune systems. The relationships between organisms, such as those between prey or predator, host and parasite, and between mating partners, are complex and multidimensional [8].

Bioinformatics is a science increasingly essential to navigate and manage the host of information generated in cellular systems: to improve study design, make candidate gene identification, interpret and manage data, and to explore light on the molecular pathology of disease-causing mutations [9]. In the genome age, after completion of Human genome project (HGP), a major research goal is to find the functions of genes and to define their interactions in a particular organism [10].

Several disciplines such as genomics, proteomics, immunoinformatics and systems biology have recently emerged within bioinformatics which for the first time enable biological systems to be studied on a scale commensurate with there inherent complexity. Most of these studies are consequently assuming a central play in modern drug development, with a wide spectrum of practical applications embracing target discovery, target validation, lead compound selection, investigation of drug modes of action, diagnostics, toxicology and clinical development [11].

In 1983, scientists led by Luc Montagnier at the Pasteur Institute in France first discovered the virus that causes AIDS [12]. The structure of HIV is different from other retroviruses and is about 120 nm in diameter (120 billionths of a meter; around 60 times smaller than a red blood cell) and roughly spherical [13].

HIV-1 and HIV-2 are two species of HIV infect humans. HIV-1 is thought to have originated in southern Cameroon after evolving from wild chimpanzees (*Pan troglodytes*) to humans during the twentieth century. HIV-2 is largely confined to West Africa [14].

2. Methodology

Biological databases creates public databases which are conducting research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information for the better understanding of molecular processes affecting human health and disease. A complete genome of Human immunodeficiency virus 1 ACCESSION NC_001802 (9181bp) and Human immunodeficiency virus 2 ACCESSION NC_001722 (10359 bp) were selected for the present study.

FGENESV0

FGENESV0 is the fastest and most accurate *ab initio* gene prediction program for viruses. Its variants that use similarity information in gene prediction, resulting in fully automatic annotation of quality similar to that of manual annotation.

Protein Molecular Weight

Protein Molecular Weight accepts a protein sequence and calculates the molecular weight of the submitted protein sequence. Protein molecular weight is calculated using tolls provided from ExPASy server.

Protein Isoelectric Point

Protein Isoelectric Point calculates the theoretical pI for the submitted protein sequence. Isoelectric point of a protein is calculated using ExPASy server.

BLASTP
BLASTP accepts protein (AA) sequences and compares them against the protein databases. The BLAST (Basic Local Alignment Search Tool) programs have been designed for speed to find high scoring local alignments. BLAST uses a heuristic algorithm which seeks local as opposed to global alignments and is therefore able to detect relationships among sequences which share only isolated regions of similarity. Because of its design for speed, there may be a minimal loss of sensitivity to distant sequence relationships.

SWISS-MODEL
SWISS-MODEL is a fully automated protein structure homology-modeling server, accessible via the ExPASy web server. The purpose of this server is to make Protein Modelling accessible to all biochemists and molecular biologists World Wide.

3. Results

HIV is a single stranded, linear, RNA containing organism. Nine genes are predicted in HIV-1 by viral gene prediction server and the genes characterized by BLASTP shown as:
Gene 1 Gag Polyprotein
Gene 2 Gag-Polpolyprotein
Gene 3 vif (viral infectivity factor)
Gene 4 Protein Vpr (Viral protein R)
Gene 5 tat protein(p28-tev)
Gene 6 Protein Vpu (Viral protein U)
Gene 7 Envelope surface glycoprotein gp160
Gene 8 envelope glycoprotein
Gene 9 Protein Nef(Negative factor)(F-protein)

In HIV-2 genome also predicted nine genes by FGENEV0 server and are shown as:
Gene 1 nef gene
gene 2 gag-pol fusion protein
gene 3 gag polyprotein
gene 4 gag pol fusion protein
gene 5 vif protein
gene 6 vpr protein
gene 7 tat protein
gene 8 env polyprotein
gene 9 nef protein .

Gag-Pol Polyprotein is having highest molecular weight and tat protein is having lowest molecular weights in both HIV-1 and HIV-2. All the proteins stands between 4 to 11 pH levels which has shown that the proteins of viruses can denature or become inactive, if there is a decrease of salinity below 4 pH or above 11 pH (Table 1 and 2).

The identification and characterization Viral protein U is absent in HIV-2 and is present in HIV-1. This protein may provide highest infectivity to humans and can have the capability in the control of human immune system.

Table 1: HIV-1 gene identification and characterization

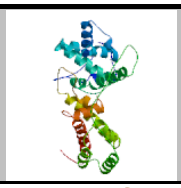
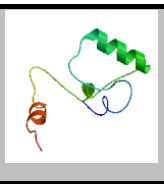
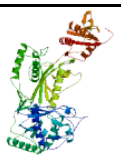
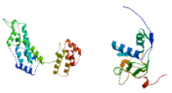
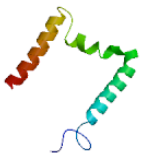
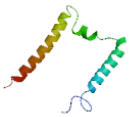
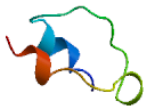
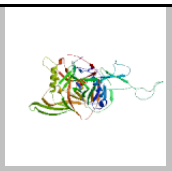
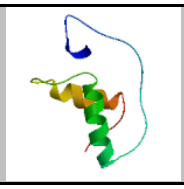
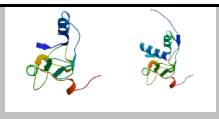
Gene No.	Type	Mol.wt kDs	pI
Gene1	Gag polyprotein (Pr55Gag)	55.94	9.42
gene2	Gag-Pol polyprotein (Pr160Gag-Pol)	103.58	8.93
gene3	vif (viral infectivity factor)	22.52	10.39
Gene4	Protein Vpr (Viral protein R)	9.31	4.76
Gene5	tat protein	8.39	10.48
Gene6	Protein Vpu (Viral protein U)	9.24	4.43
Gene7	Envelope surface glycoprotein gp160	97.23	9.19
Gene8	envelope glycoprotein	20.24	7.12
Gene9	Protein Nef (Negative factor) (F-protein)	13.69	6.96

Table2: HIV-2 gene identification and characterization

Gene No.	Type	Mol.wt kDs	pI
Gene1	Nef (Negative factor)	15.03	5.57
gene2	Gag-Pol polyprotein (Pr160Gag)	45.89	9.56
gene3	Gag polyprotein (Pr55Gag)	11.50	6.80
Gene4	Gag-Pol polyprotein (Pr160Gag	110.09	8.95
Gene5	Virion infectivity factor (Vif)	25.32	10.18
Gene6	Protein Vpr (Viral protein R)	10.07	6.96
Gene7	tat protein	11.24	9.65
Gene8	Envelope glycoprotein gp160	98.95	8.79
Gene9	Protein Nef (Negative factor)	29.92	6.11

Modelled structures of these nine proteins are provided in Table 3.

Table 3: Protein modelling using Swissmodel

Protein	Protein modelling HIV-1	Protein modelling HIV-2
Gag polyprotein (Pr55Gag)		
Gag-Pol polyprotein (Pr160Gag-Pol)		
vif (viral infectivity factor)	**No Templates found**	**No Templates found**
Protein Vpr (Viral protein R)		
tat protein		
Protein Vpu (Viral protein U)		**Gene not predicted**
Envelope surface glycoprotein gp160		
envelope glycoprotein	**No Templates found**	**Gene not predicted**
Protein Nef(Negative factor)(F-protein)		

4. Discussion

Information Technology and Biological sciences are being transformed due to enormous growth of data from laboratories worldwide. Most of the biologists and computer scientists focus to explore innovations of their research in faster rate using developments in Information technology. A complete genome of HPV-92 predicts six genes by gene prediction technique [15]. The comparison of two or more sequences of numbers or letters is common in several fields such as molecular biology, bioinformatics, speech recognition and computer science [16].

Hence by alignment of strings in the genome of organisms can predict the nature and functions of various cellular systems, which can in turn predict the health of the species. A small organism such as HIV has the capacity to kill the complex organisms such as Humans. Hence the scientists are focusing on how the nanoorganisms such as viruses are destroying the complex organisms which are measuring in few feet's of high. There are compounds which are still lesser in sizes than nano and have the capabilities to change the cellular processes of life.

The present studies have provided the string which can have the capacity to process the data. The stored information in HIV can lead the process of 3D compounds and has the capability to control the cellular mechanisms of human life. The studies of data provided the sizes and pI values which are the biological properties of HIV. The nine proteins of HIV-1 and HIV-2 can be controlled, if all the systems mechanisms are predicted using the advanced information technologies.

5. Conclusion

Predictions of molecular structures are highly necessary in the studies of systems biology. The increased data availability can provide answers for the life process from birth to death. Hence the high end processing of data can explore the characterization and functional cellular changes in various species which are measuring from nano to larger sizes.

6. Acknowledgements

The authors would like to thank GITAM University for providing computational facility and access to e-journals to carry out this research.

References

[1] J. F. Roeth, K. L. Collins, "Human Immunodeficiency Virus Type 1 Nef: Adapting to Intracellular Trafficking Pathways", Microbiol. Mol. Biol. Rev., Vol 70, 548-563, 2006.

[2] J. B. Alimonti, T. B. Ball, K. R. Fowke, Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS, J. Gen. Virol., Vol. 84, 1649-1661, 2003.

[3] AS Fauci, **"**The human immunodeficiency virus: infectivity and mechanisms of pathogenesis",Science, Vol. 239, Issue 4840, 617-622, 1988.

[4] Collins, D.L.; Zijdenbos, A.P.; Kollokian, V.; Sled, J.G.; Kabani, N.J.; Holmes, C.J.; Evans, A.C., "Design and construction of a realistic digital brain phantom ", Medical Imaging, V 17, No.3,pp. 463 – 468, 1998.

[5] Hans Moravec, "When will computer hardware match the human brain?",Journal of Evolution and Technology. 1998. Vol. 1, 1998.

[6] Peter M. Vitousek, Harold A. Mooney, Jane Lubchenco, Jerry M. Melillo, Human Domination of Earth's Ecosystems“,Science, Vol. 277. no. 5325, pp. 494 – 499, 1997.

[7] Ralph S. Lillie, “Living Systems and Non-Living Systems”, Philosophy of Science, Vol. 9, No. 4, pp. 307-322, 1942

[8] Vladimir B. Bajić, Tin Wee Tan, “Information Processing and Living Systems”, Imperial College Press, pp. 1, 2005.

[9] Michael R. Barnes, Ian C. Gray, “Bioinformatics for Geneticists”, John Wiley & Sons, pp. 6, 2003.

[10] David W. Mount, “Bioinformatics: Sequence and Genome Analysis”, Cold Spring Harbor Laboratory Press, pp. 3, 2004.

[11] Alexander Hillisch and R. Hilgenfeld, “Modern methods of drug discovery”, Birkhäuser Basel, 1 edition, pp. 19-20, 2003.

[12] Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)", Science, vol. 220, no. 4599, pp. 868-871, 1983.

[13] McGovern SL, Caselli E, Grigorieff N, Shoichet BK, "A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening", J Med Chem, vol. 45, no. 8, pp. 1712-1722, 2002.

[14] Gao, F., Bailes, E., Robertson, D. L., Chen, Y., Rodenburg, C. M., Michael, S. F., Cummins, L. B., Arthur, L. O., Peeters, M., Shaw, G. M., Sharp, P. M., and Hahn, B. H., "Origin of HIV-1 in the Chimpanzee Pan troglodytes troglodytes", Nature, vol. 397, no. 6718, pp. 436-441, 1999.

[15] D.S.V.G.K.Kaladhar, T. Uma Devi, P.V. Nageswara Rao, An *in silico* genome wide identification, characterization and modeling of Human Papilloma Virus strain 92**,** IJEST, Vol. 2, no. 9, pp. 4288-4291, 2010.

[16] Satyasaivani B., Kaladhar DSVGK, Shashi M. and Kesavareddy J., “Prediction of *methanobacterium* using suffixtree”, International Journal of Machine Intelligence, Vol. 1, no. 2, pp-1-4, 2009.

Author Biographies

Dr.DSVGK Kaladhar: Dr. Kaladhar was born in Vijayawada on 1974. He has got his doctoral degree in Biotechnology and M,Sc in Microbiology from ANU, Guntur. He is presently working as Asst. Professor in Bioinformatics, GITAM University, Visakhapatnam. His field of study is neural networks, Image analysis, Bioinformatics and Biology

Mr. A.Krishna Chaitanya: Mr. Chaitanya was born in Rajahmundry on 1984. He has acquired Master of Science in Biochemistry from AU, Visakhapatnam. He is presently working as Asst.Professor in GITAM University, Visakhapatnam. His specializations are Molecular modeling, Bioinformatics and Biochemistry.