



HIV Progression Depends on Codon and Amino Acid Usage Profile of Envelope Protein and Associated Host-Genetic Influence

Ayan Roy^{1†}, Rachana Banerjee^{2†} and Surajit Basak^{3,4*}

¹ Department of Botany, Bioinformatics Facility, University of North Bengal, Siliguri, India, ² Structural Biology and Bio-Informatics Division, CSIR-Indian Institute of Chemical Biology, Kolkata, India, ³ Department of Molecular Biology and Bioinformatics, Tripura University, Agartala, India, ⁴ Bioinformatics Centre, Tripura University, Agartala, India

OPEN ACCESS

Edited by:

Hirofumi Akari,
Kyoto University, Japan

Reviewed by:

Yorifumi Satou,
Kumamoto University, Japan
Shigeyoshi Harada,
National Institute of Infectious
Diseases, Japan

*Correspondence:

Surajit Basak
basaksurajit@gmail.com

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 17 March 2017

Accepted: 29 May 2017

Published: 15 June 2017

Citation:

Roy A, Banerjee R and Basak S (2017)
HIV Progression Depends on Codon
and Amino Acid Usage Profile of
Envelope Protein and Associated
Host-Genetic Influence.
Front. Microbiol. 8:1083.
doi: 10.3389/fmicb.2017.01083

Acquired immune deficiency syndrome (AIDS) is a spectrum of conditions caused by infection with the human immunodeficiency virus (HIV). Two types of HIV have been characterized: HIV-1 and HIV-2. The present study investigated whether evolutionary selection pressure differs between rapid progressor (RP), slow progressor (SP), and long-term non-progressor (LTNP) of HIV-1 infected individuals. An unexpected association between the evolutionary rate of substitution in envelope (env) gene and disease progression is observed. Our present study suggests that env genes of LTNP are subject to unusually strong functional constraint with respect to RP. We also observed that the three categories of env genes i.e., RP, SP, and LTNP, had their own characteristic pattern of amino acid usage and SP and LTNP sequences shared similar patterns of amino acid usage different from RP sequences and evolutionary rate significantly influenced the amino acid usage pattern of the three different types of env gene sequences. It was also noted that the evolutionary rate for the glycosylation sites of LTNP and SP sequences were even significantly less than the RP sequences. Comparative analysis on the influence of human host on the three categories of env genes are well correlated with the rates of disease progression suggesting the adaptive strategies of the viruses for successful residence and infection. Host associated selective constraints appeared most relaxed on the RP sequences and strongest in LTNP sequences. The present study clearly portrays how evolutionary selection pressure differs between three categories of env genes i.e., RP, SP, and LTNP. The env genes, coding for the env glycoproteins, experience severe selection constraints from the host due to their constant exposure to the host immune system. In this perspective it might be suggested that env gene evolution occurs mainly by negative selection with the occurrence of mutation that might not reach fixation in the viral population. This work also confers a deeper insight into the crucial effects of host factors that govern the overall progression of HIV infection.

Keywords: rapid progressor, slow progressor, long-term non-progressor, evolutionary rate, disease progression

INTRODUCTION

Human immunodeficiency viruses (HIVs) are one of the most crucial members of the retroviral family *Retroviridae*. Among the two main types of HIVs i.e., HIV-1 and HIV-2, HIV-1 has been characterized as the major causative agent of acquired immune deficiency syndrome (AIDS) (Blattner et al., 1988; Weiss, 1993).

Natural variability of HIV-1 is the key underlying the complex biology of the menacing virus. Evolutionary and epidemiological history of HIV can be investigated through phylogenetic analysis of erratic forms of the virus. Interestingly, due to rapid availability of nucleotide sequence data of viral genes, it is nowadays possible to illustrate a detailed phylogenetic relationship of viruses like HIV. Phylogenetic arrangements of HIV-1 are commonly carried out using nucleotide sequences of different sub-genomic regions of the same HIV-1 genome i.e., gag, pol, and envelope (env). This approach has already revealed unique inter-subtype recombinant forms of virus isolates (McCutchan, 2006). Envelope (env) gene encodes known targets for cytotoxic T lymphocytes and neutralizing antibodies (Goudsmit et al., 1988; Chesebro et al., 1991). Despite extensive study of sequence variation of env gene of HIV-1, the exact impact of selection in governing the patterns of variation in the env gene still remains somewhat obscure. In some earlier studies, based on lower frequency of synonymous nucleotide substitutions within the V3 loop in the env genes, it has been circuitously verified that the variation in env gene is maintained by selection for antigenic diversity (Simmonds et al., 1991; Bonhoeffer et al., 1995). Theoretically, it has also been implicated that selection pressure and mutability dynamics are the vital forces that affect viral fitness and robustness (Brown, 1997). It has been hypothesized that negative (purifying) selection constraints, operational in HIV-1, execute a crucial role in devising molecular evolutionary patterns of HIV-1 in contrast to positive (diversifying) selection which has been reported to play a minor role (Seo et al., 2002; Drummond et al., 2003).

Investigations pertaining to viral divergence in HIV-1 patients, associated with varying rates of progression of disease, have often produced interesting and conflicting results (Ganeshan et al., 1997; Arts and Quiñones-Mateu, 2003; Rangel et al., 2003). Some researchers have noted that non-synonymous substitutions in the env gene have been higher in long-term non-progressor (LTNP) whereas, there has been no difference between slow and normal progressors (NPs) regarding the influence of synonymous substitution (Strunnikova et al., 1995; Bagnarelli et al., 1999). However, Markham and co-workers observed opposite results with greater accumulation of non-synonymous substitutions in NPs (Markham et al., 1998). The event of T-cell activation can be effectively employed to estimate the progression to AIDS in HIV-1 infected individuals and might be utility in apt assessment of viral replication rates and evolutionary complexities. Consequently, activation of immune responses might enforce critical constraints on the evolutionary tricks and viral generation rates of HIV. However, such effects yet remain unexplored from the evolutionary perspective. Extensive investigations pertaining to evolutionary signatures might prove handy in elucidating the patterns of molecular adaptation of HIV in human host. Variations in synonymous substitution rates

reflect changes in generation time or mutation rate, while rate of non-synonymous mutations tend to be affected by changes in selective pressure and effective population size (Lemey et al., 2007).

Envelope (env) gene is the fastest evolving one in the HIV-1 genome. Conflicting selective pressures shape the evolutionary dynamics of the virus (Korber et al., 2000; Ross and Rodrigo, 2002; Yoshida et al., 2011). In the present study, investigations pertaining to the evolutionary traits of the env genes representing rapid progressor (RP), slow progressor (SP), and LTNP human patients (available in public domain), have been performed extensively through detailed phylogenetic analysis, followed by subsequent estimation of synonymous and non-synonymous substitution rates. Rapid-progressors are associated with speedy development of AIDS within 3 years of infection whereas; slow-progressors are characterized by a comparatively slower gradual onset of AIDS after seroconversion that might take a span 3–10 years (Kumar, 2013) to develop. Comparative analysis of relative synonymous codon usage (RSCU) patterns of the three categories of env genes marked by varying degrees of disease progression i.e., RP, SP, and LTNP types and their similarity index with human host has also been executed in the present endeavor with a motif to address the adaptive strategies of the viruses for successful residence and infection. Results pertaining to RSCU patterns and evolutionary dynamics might prove utility to unravel the molecular underpinnings of viral adaptation and infection in human host and excavate subtle discrepancies among the viral types associated with varying rates of disease progression.

MATERIALS AND METHODS

Sequence Retrieval

All available env gene sequences of HIV-1 subtype B were retrieved from HIV Database (<http://www.hiv.lanl.gov/>) representing three types of patient categories i.e., RP, SP, and LTNP patients. Redundant and erroneous sequences (sequences with internal stop codons) were removed to avoid stochastic variations and sampling errors (Wright, 1990). A comprehensive set of 264 coding sequences constituted final dataset for our analysis (Supplementary Material, Data Sheet 1).

Phylogenetic Analysis

Phylogenetic analysis provides the ancestral relationship of a set of sequences. It involves the construction of a tree, where the nodes indicate separate evolutionary paths, and the lengths of the branches give an estimate of how distantly related the sequences represented by those branches are. In the present study, all the env genes were aligned using the Clustal Omega program (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). The resultant multiple sequence alignments were subsequently used to construct the neighbor-joining method based phylogenetic tree with 500 bootstrap replicates. MEGA 7 was used for phylogenetic analysis (Kumar et al., 2016). Here, we have constructed an unrooted tree, where, the distances and relationships between the taxa have been plotted without making any assumption concerning their descent.

Estimation of Relative Synonymous Codon Usage

Relative synonymous codon usage (RSCU) (Sharp et al., 1986) is calculated as the ratio of the observed frequency of a codon to the expected frequency if codon usage was uniform within a synonymous codon group.

Many genes display a non-random usage of synonymous codons for specific amino acids. A measure of the extent of this non-randomness is given by the RSCU (Sharp and Li, 1986). It is the ratio of observed frequency of the codons with respect to the expected frequency of the same codon if codon usage was uniform within a synonymous codon group.

RSCU is calculated as:

$$RSCU = \frac{\text{Frequency of codon}}{\text{Expected frequency of codon (if codon usage was uniform)}}$$

Relative synonymous codon usage (RSCU) values > 1 indicate that the observed frequency of synonymous codons is more compared to the expected frequency and lower than one indicates the opposite (dos Reis et al., 2003).

RSCU values of the 59 codons [excluding the single synonymous codons AUG (Met) and UGG (Trp) and the three termination codons] of RP, SP, and LTNP env gene sequences were calculated using CodonW (Ver. 1.4.2) software (<http://www.molbiol.ox.ac.uk/cu>) (Peden, 2000). Codon usage frequencies of human host (*Homo sapiens*) were obtained from the Codon Usage Database (<http://www.kazusa.or.jp/codon/>) (Nakamura et al., 2000).

Assessment of Similarity Index

Viral genomes are relatively smaller in size and largely rely on the host to execute crucial biological activities like replication, protein synthesis and transmission (Nasrullah et al., 2015). In this pretext, it has been suggested that viral robustness, survival and evasion of host's immune signals and responses largely depend on the interplay of codon usage patterns of the concerned virus and its respective host (Shackelton et al., 2006; Moratorio et al., 2013). Here, we have considered the similarity index to understand the influence of host genome on the adaptability of virus genome inside the host. The influence of the overall codon usage pattern of the host on the formation of the overall codon usage of the virus is defined as the similarity index.

Relative synonymous codon usage (RSCU) values of the three different types of env gene sequences i.e., RP, SP, and LTNP were compared with that of human host in order to assess the influence of human host system in shaping the patterns of codon usage among the env types. The parameter similarity index, $D(A,B)$ (Nasrullah et al., 2015) is computed as follows:

$$R(A,B) = \frac{\sum_{i=1}^{i=59} a_i X b_i}{\sqrt{\sum_{i=1}^{i=59} a_i^2 X \sum_{i=1}^{i=59} b_i^2}}$$

$$D(A,B) = \frac{1 - R(A,B)}{2}$$

where $R(A,B)$ refers to the cosine value of an included angle between A and B spatial vectors and represents the similarity between particular env type of HIV and overall codon usage pattern of human host. a_i signifies the RSCU value for a particular codon among the pool of 59 codons for every specific type of env (i.e., RP, SP, and LTNP) protein coding sequence. b_i indicates the RSCU value for the same codon in case of human host. $D(A,B)$ signifies the probable impact of human codon usage patterns on the concerned env types of HIV. The value of similarity index, $D(A,B)$, has been reported to lie between 0 to 1.0 (Zhou et al., 2013).

Multivariate Analyses on Amino Acid Usage

Correspondence analysis (COA) (Peden, 2000; <http://www.molbiol.ox.ac.uk/cu>) was used to investigate the major trend in amino acid usage variation among the env genes. Since amino acid usage by its very nature is multivariate, it is necessary to analyse this data with multivariate statistical techniques i.e., COA. Correspondence analysis (COA) is an ordination technique that identifies the major trends in the variation of the data and distributes genes along continuous axes in accordance with these trends. It has the advantage of not to make any assumption that the data falls into discrete clusters and therefore represent continuous variation accurately. Correspondence analysis (COA) on relative amino acid usage (RAAU) of env gene sequences was executed employing the CodonW program.

Evolutionary Rate Calculation

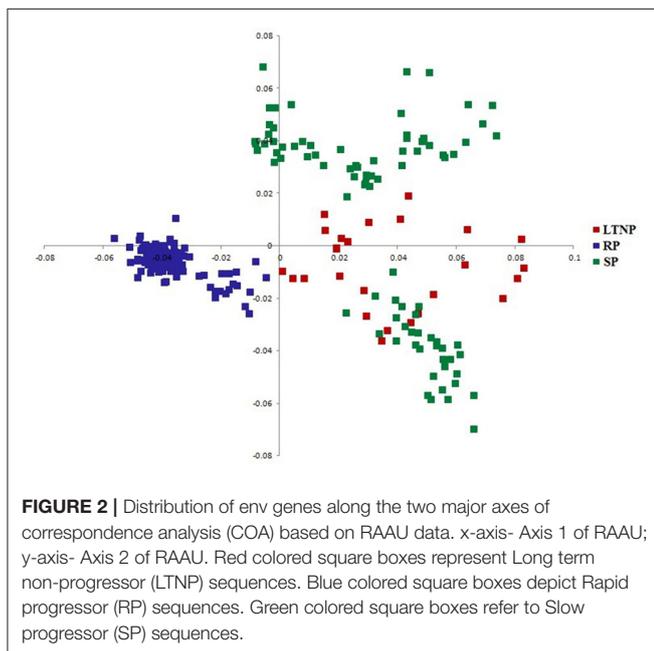
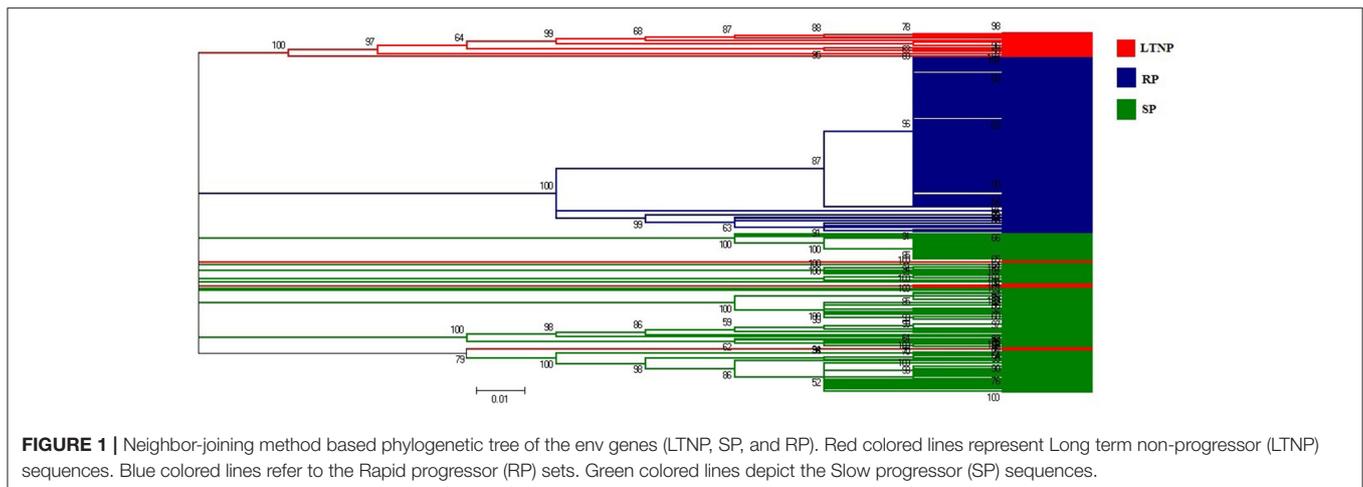
The ratio (ω) of rate of non-synonymous substitutions per non-synonymous site (K_a) to rate of synonymous substitutions per synonymous site (K_s) indicates the impact of evolution on a gene segment. $\omega > 1$ indicates diversifying (positive) selection whereas, $\omega < 1$ signifies purifying (negative) selection (Roy et al., 2015). The evolutionary rates of the orthologous env genes representing all three types under analysis i.e., RP, SP, and LTNP (with reference to progressors) were calculated using Codeml program included in the PAML software package (ver. 4.5) (Nei and Gojobori, 1986; Yang, 2007) (<http://abacus.gene.ucl.ac.uk/software/paml.html>) with runmode = -2 and CodonFreq = 1.

Evolutionary rate of each individual residue for a given env gene sequence was calculated using SWAKK server (<http://ibl.mdanderson.org/swakk/>) (Liang et al., 2006). It estimates the ratio of non-synonymous to synonymous substitution rates (K_a/K_s) between a pair of protein-coding DNA sequences, by a sliding 3D window analysis.

RESULTS

Phylogenetic Profiling of Various Categories of env Genes

Phylogenetic analysis was performed to investigate the evolutionary relationship of 264 env gene sequences of HIV-1 subtype B, representing three different categories of patients (RP, SP, and LTNP). The phylogenetic tree revealed three major lineages where mostly three categories of env gene sequences i.e., RP, SP, and LTNP have been clustered (**Figure 1**).



Trends in Amino Acid Usage of env Genes Representing Varying Rates of Disease Progression

Correlating the observed phylogenetic profile of env genes with RAAU pattern appeared necessary in order to efficiently explore the impact of the various categories of env genes on gene sequence diversity. Accordingly, CoA on amino acid usage was performed using 264 env gene sequences representing three different categories of disease progression genes of HIV-1 subtype B. COA generated two separate clusters (**Figure 2**). **Figure 2** shows the distribution of the genes along two major axes of amino acid usage variation.

It was perceptible from the amino acid usage based COA that SP and LTNP sequences shared similar patterns of amino acid usage different from RP sequences which formed a discrete

cluster along Axis 1 (**Figure 2**). It was apparent from **Figure 2** that the three categories of env genes i.e., RP, SP, and LTNP, had their own characteristic pattern of amino acid usage and the env genes of SP exhibited greater resemblance with env genes of LTNP in terms of amino acid usage pattern. It was also clear that the results of COA were in absolute accordance with phylogenetic profile (**Figure 1**).

Impact of Evolutionary Selection Pressure on Three Categories of env Genes

Earlier, Canducci et al. (2009) observed that evolutionary rate of HIV-1 env gene varies between NP and LTNP patients. The evolutionary rate (ω) for the env sequences of RP, SP, and LTNP type was found to correlate significantly with Axis 1 ($R = -0.31$, $P < 0.01$) of RAAU data. It indicates that evolutionary selection pressure significantly influenced the amino acid usage pattern of the three different types of env gene sequences.

It was evident from COA of amino acid usage that env gene sequences belonging to the category SP and LTNP segregated from RP along Axis 1 of RAAU. Significant negative correlation of ω (Ka/Ks) with Axis 1 of RAAU indicated that the average Ka/Ks values of SP and LTNP sequences would be lesser than the average Ka/Ks value of RP sequences. It was indeed found that the average Ka/Ks value of LTNP sequences (0.48) was significantly less than that of average Ka/Ks value of RP sequences (0.58) ($P < 0.01$). We also observed that the average Ka/Ks value of SP sequences (0.49) was significantly less than that of RP sequences (0.58) ($P < 0.01$). But, the difference between the average Ka/Ks values of LTNP and SP sequences was observed to be statistically insignificant. Considering the fact that evolutionary rate might vary depending on the functional constraints, SWAKK web server (<http://ibl.mdanderson.org/swakk/>) was employed to calculate the evolutionary rate for each single amino acid residue in the concerned protein sequences. Extensive glycosylation of HIV-1 envelope proteins (env) is known to play an important role in evasion of host immune response. Average Ka/Ks value for the glycosylation sites of RP sequences were compared with that of the average Ka/Ks values of the LTNP and SP sequences.

Interestingly, it was noted that the average Ka/Ks values for the glycosylation sites of LTNP (0.37) and SP (0.48) sequences were significantly less than that of the average Ka/Ks value of RP sequences (0.62) ($P < 0.01$). It was also noteworthy that the average Ka/Ks value of glycosylation sites for LTNP sequences was significantly less than that of SP sequences ($P < 0.01$; **Figure 3**). Such an observation clearly indicates the functional importance and vitality of glycosylation sites in case of HIV infection.

Influence of Host Machinery on the Disease Progression of HIV Based on RSCU

In the present study we have considered the similarity index parameter (Materials and Methods) to understand the influence of host genome on the adaptability of virus genome inside the host. Methodical inspection of similarity index of the three concerned categories of env gene sequences with human host revealed that the selection pressure due to human host was more severe on the LTNP sequences in comparison to the SP and RP sequence sets (**Figure 4**). Host associated selective constraints appeared most relaxed on the RP sequences as has been evident from **Figure 4**. Difference between similarity index values of the respective LTNP, SP, and RP sequences appeared to be statistically significant and the pattern of variation among them seemed distinct enough to infer that the LTNP sequences were under a stronger selective impact of human host.

Apart from estimating similarity index, we extended our analysis pertaining to the usage of identical codons among the three diverse types of env gene sequences with human host. Codons were defined as over-represented ($RSCU > 1.6$) and under-represented ($RSCU < 0.6$) as per scheme followed by

Wong et al. (2010). Similar codon usage pattern was inferred when a particular codon was found to display RSCU values < 0.6 or more than 1.6 or found to fall within a range of 0.6 to 1.6 for both human host and the respective env type i.e., RP, SP, and LTNP. Interestingly, it was evident after careful inspection that LTNP sequences shared the highest frequency of similarly selected codons (36 codons out of 59 codons) with human host system in comparison to SP (34 codons out of 59 codons) and RP (32 codons out of 59 codons) types (Supplementary Table 1). Occurrence of higher frequencies of identically shared codons with human host in LTNP might eventually lead to a better adaptability and longer abode of the concerned LTNP in human cellular environment.

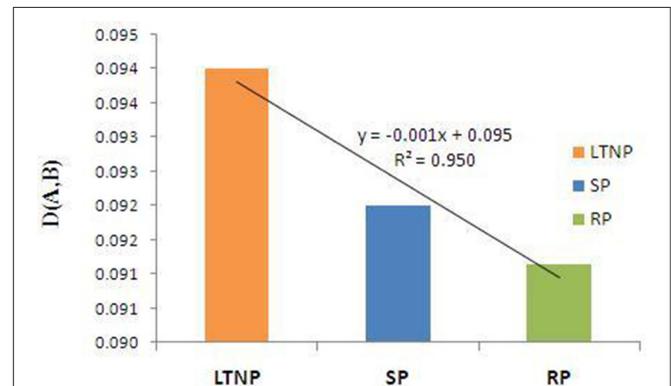


FIGURE 4 | Similarity index [D(A,B)] of env genes (LTNP, SP, and RP) with respect to human host. Orange bar refers to similarity index value [D(A,B)] of Long term non-progressor (LTNP) sequences. Blue bar represents the similarity index value [D(A,B)] of the Slow progressor (SP) sets. Green bar depicts the similarity index value [D(A,B)] of the Rapid progressor (RP) sequences.

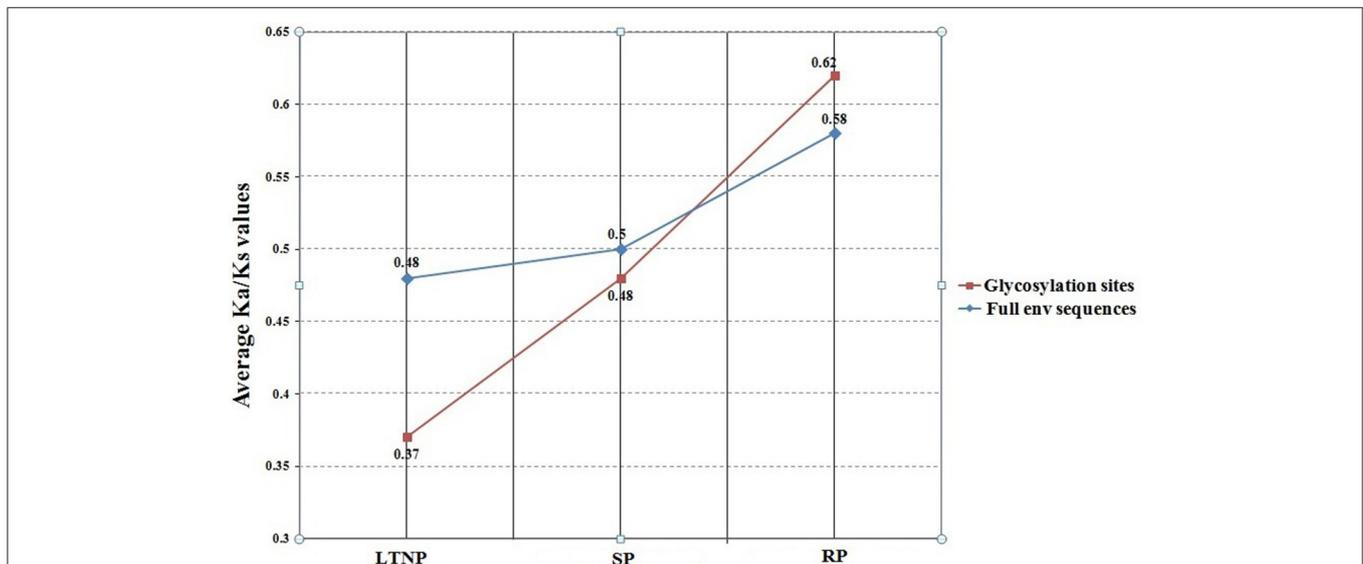


FIGURE 3 | Variation of Ka/Ks of env genes (LTNP, SP and RP) and respective glycosylation sites. LTNP, Long term non-progressor; SP, Slow progressor; RP, Rapid progressor. The blue line refers to the plot of the average Ka/Ks values of the full env sequences of LTNP, SP, and RP. The red line represents the plot of the average Ka/Ks values of the associated glycosylation sites of LTNP, SP, and RP.

DISCUSSION

Intragenomic and intergenomic variations pertaining to codon and amino acid usage patterns can be well explained in the light of multivariate statistical analysis. Correspondence analysis (CoA) is one such type of comprehensive statistical tool which highlights the major variations among codon and amino acid usage data and places them in accordance with such observed variations (Greenacre, 1984). CoA on the basis of RAAU of the three concerned types of env gene sequences revealed their unique amino acid compositional features and separated them according to their diverse compositional traits along the two principle axes of separation of genes i.e., Axis 1 and Axis 2 of RAAU data (Figure 2). SP and LTNP sequences, to certain extent, had identical amino acid usage features as was reflected from COA on RAAU data (Figure 2). However, RP sequences clustered separately along Axis 1 signifying completely diverse amino acid usage patterns from SP and LTNP sequences. Such contrasting features of amino acid composition of RP sequences might have an implication in their rapid rate of disease progression in human host, characterized by a short stay in host cellular environment and speedy proliferation and infection (Kumar, 2013; Jarrin et al., 2015).

Earlier reports suggested that the env proteins of LTNP were subject to purifying selection for their overall mutability and tendencies of variation in LTNP. The impact of stronger purifying selection on env genes of LTNP in contrast to RP might be an outcome of the survival value conferred by adaptive stability of the envelop protein. To further validate our observation, we then fitted linear regression lines between K_a and K_s through the origin assuming that K_a and K_s should both initially be zero at the moment of lineage divergence. The slopes for the lines of whole env genes for LTNP and RP were found to be 0.48 and 0.58, respectively. These results indicated that the increase in K_a with respect to K_s in RP was around 1.2 times faster than in LTNP. Such observations were suggestive of the fact that env proteins of LTNP genes have been subject to more severe evolutionary selection with respect to RP, which eventually has been reflected at a lower level of non-synonymous nucleotide substitution for a given level of synonymous substitution. Similarly, the slope for SP was found to be 0.49, suggesting that the increase in K_a with respect to K_s in SP was around 1.01 times faster than in LTNP. This, in turn, reinforced the hypothesis that unusual functional constraints have been instrumental on the env protein sequences of LTNP. Higher average K_a/K_s value of RP over SP and LTNP sequences (statistically significant) indicated that the evolutionary constraint on the RP sequences has been lesser compared to the impact of evolutionary forces on the SP and LTNP sequences. Such an instance of relaxed selection pressure might confer an added advantage to the RP sequences, associated with rapid infection in a short span of time, in accumulation of non-synonymous mutations and subsequent evasion of host (human) immune response (Khanlou et al., 1996; Jarrin et al., 2015).

HIV-1 shows variation in susceptibility of infection to human (Fellay et al., 2007). Many attempts have been made to

understand the viral genetics associated with disease progression in HIV-1 affected patients. However, molecular mechanism and host genetic influence on effective viral pathogenesis and subsequent progression of disease still remain unclear (Michael, 1999; Carrington et al., 2001).

Viruses mimic the codon usage profile of their respective hosts and rely largely on the chaperon apparatus of their hosts for proficient replication and enhanced robustness (Kunec and Osterrieder, 2016). In this pretext, similarity index of a concerned virus with its respective host has been suggested to be an effective tool to properly assess the impact of host genomic influence on the patterns of associated viral codonic signatures. Several earlier studies have used similarity index to demonstrate the influence of host genome over the viral genome (Nasrullah et al., 2015; Butt et al., 2016). Estimation of similarity index with human host among the concerned types of env sequences associated with varying degrees of disease progression i.e., RP, SP, and LTNP signified the fact that the impact of selection by human host was most intense in LTNP compared to SP and RP. Such observations seemed logical in light of the fact that LTNP sequences reside in the human system for long and are characterized by high CD4+ and CD8+ T-cell counts (Zeller et al., 1996; Kumar, 2013). Thus, higher similarity index with human host in LTNP and usage of a higher proportion of identically selected codons with human host tend to be a consequence of prolonged residence and adaptive efficacy of LTNP in human body. In contrast, the instance of lower similarity index in RP with human host might be well explained from the viewpoint that RP stays in human body for a shorter span and causes rapid infection (Khanlou et al., 1996). However, SP sequences displayed higher similarity index with human host in comparison to RP which might be attributed to a comparatively longer residence and slower disease progression than RP (Hogervorst et al., 1995). Thus, observations pertaining to similarity index among the three types of env sequences with human host seemed well correlated with the rates of disease progression and duration of abode in human host.

CONCLUSION

Decreased HIV-1 strain evolution correlates with weaker viral fitness and the inability to evade the host immune system. Some studies have shown that viral strains from LTNPs are less evolved and thus, less capable of evading the host immunological response when compared with progressor strains (Wang et al., 2003; Sandoñis et al., 2009). In our study, K_a/K_s decreased significantly for LTNP env gene sequences compared to RP gene sequences. Similar trend was observed for glycosylation sites of the respective sequences. Findings of the present study confirm that HIV-1 env gene of LTNP have been under stronger purifying selection compared to the respective RP counterpart. The env genes, coding for the env glycoproteins, experience severe selection constraints from the host due to their constant exposure to the host immune system. The env protein mediates HIV-1 entry into host CD4+ T cells. Reduction in K_a/K_s for env genes of

LTNP might result in significantly slower infection of CD4+ T cells compared with RP. Our hypothesis seems well justified in light of the fact that the env protein plays essential role in receptor binding and cell entry (Wyatt and Sodroski, 1998) and efficiently maintains extensive glycosylation on its surface to avoid antibody neutralization (Reitter et al., 1998). In this perspective, it might be suggested that env gene evolution occurs mainly by negative selection with the occurrence of mutation that might not reach fixation in the viral population.

Present study addresses an evolutionary perspective that recognizes the ongoing dynamic interplay between human host and genetic diversity of HIV. Our analysis also provides an elucidation of the driving force accountable for the selection of favorable host genetic profiles as well as for viral mutations that are best adapted to the concerned host effects. Outcomes of the current research work confer a vivid portrait about the role of host factors in determining the overall progression of HIV infection.

REFERENCES

- Arts, E. J., and Quiñones-Mateu, M. E. (2003). Sorting out the complexities of HIV-1 fitness. *AIDS* 17, 780–781. doi: 10.1097/01.aids.0000050881.72891.32
- Bagnarelli, P., Mazzola, F., Menzo, S., Montroni, M., Butini, L., and Clementi, M. (1999). Host-specific modulation of the selective constraints driving human immunodeficiency virus type 1 env gene evolution. *J. Virol.* 73, 3764–3777.
- Blattner, W., Gallo, R. C., and Temin, H. M. (1988). Hiv causes AIDS. *Science* 241, 515–516.
- Bonhoeffer, S., Holmes, E. C., and Nowak, M. A. (1995). Causes of HIV diversity. *Nature* 376, 125–125.
- Brown, A. J. L. (1997). Analysis of HIV-1 env gene sequences reveals evidence for a low effective number in the viral population. *Proc. Natl. Acad. Sci. U.S.A.* 94, 1862–1865. doi: 10.1073/pnas.94.5.1862
- Butt, A. M., Nasrullah, I., Qamar, R., and Tong, Y. (2016). Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerg. Microbes Infect.* 5:e107. doi: 10.1038/emi.2016.106
- Canducci, F., Marinozzi, M. C., Sampaolo, M., Berrè, S., Bagnarelli, P., Degano, M., et al. (2009). Dynamic features of the selective pressure on the human immunodeficiency virus type 1 (HIV-1) gp120 CD4-binding site in a group of long term non-progressor (LTNP) subjects. *Retrovirology* 6:1. doi: 10.1186/1742-4690-6-4
- Carrington, M., Nelson, G., and O'Brien, S. J. (2001). Considering genetic profiles in functional studies of immune responsiveness to HIV-1. *Immunol. Lett.* 79, 131–140. doi: 10.1016/S0165-2478(01)00275-9
- Chesebro, B., Nishio, J., Perryman, S., Cann, A., O'Brien, W., Chen, I. S., et al. (1991). Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. *J. Virol.* 65, 5782–5789.
- dos Reis, M., Wernisch, L., and Savva, R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* 31, 6976–6985. doi: 10.1093/nar/gkg897
- Drummond, A., Oliver, G., and Rambaut, A. (2003). Inference of viral evolutionary rates from molecular sequences. *Adv. Parasitol.* 54, 331–358. doi: 10.1016/S0065-308X(03)54008-8
- Fellay, J., Shianna, K. V., Ge, D., Colombo, S., Ledergerber, B., Weale, M., et al. (2007). A whole-genome association study of major determinants for host control of HIV-1. *Science* 317, 944–947. doi: 10.1126/science.1143767
- Ganeshan, S., Dickover, R. E., Korber, B. T., Bryson, Y. J., and Wolinsky, S. M. (1997). Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. *J. Virol.* 71, 663–677.

AUTHOR CONTRIBUTIONS

SB planned and designed the work. AR and RB generated the data. SB, AR, and RB analyzed data and prepared the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Authors thankfully acknowledge the Department of Biotechnology, Government of India (BT/BI/12/055/2012), for their computational facility at Bioinformatics Centre, Tripura University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.01083/full#supplementary-material>

- Goudsmit, J., Debouck, C., Meloen, R. H., Smit, L., Bakker, M., Asher, D. M., et al. (1988). Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 85, 4478–4482. doi: 10.1073/pnas.85.12.4478
- Greenacre, M. J. (1984). *Theory and Applications of Correspondence Analysis*. London: Academic Press.
- Hogervorst, E., Jurriaans, S., de Wolf, F., van Wijk, A., Wiersma, A., Valk, M., et al. (1995). Predictors for non- and slow progression in human immunodeficiency virus (HIV) type 1 infection: low viral RNA copy numbers in serum and maintenance of high HIV-1 p24-specific but not V3-specific antibody levels. *J. Infect. Dis.* 171, 811–821. doi: 10.1093/infdis/171.4.811
- Jarrin, I., Pantazis, N., Dalmau, J., Phillips, A. N., Olson, A., Mussini, C., et al. (2015). Does rapid HIV disease progression prior to combination antiretroviral therapy hinder optimal CD4+ T-cell recovery once HIV-1 suppression is achieved? *AIDS* 29, 2323–2333. doi: 10.1097/QAD.0000000000000805
- Khanlou, H., Salmon-Ceron, D., and Sicard, D. (1996). Characteristics of rapid progressors in HIV infection. *Ann. Med. Interne (Paris)*. 148, 163–166.
- Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., et al. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science* 288, 1789–1796. doi: 10.1126/science.288.5472.1789
- Kumar, P. (2013). Long term non-progressor (LTNP) HIV infection. *Indian J. Med. Res.* 138:291.
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Kunec, D., and Osterrieder, N. (2016). Codon pair bias is a direct consequence of dinucleotide bias. *Cell Rep.* 14, 55–67. doi: 10.1016/j.celrep.2015.12.011
- Lemey, P., Pond, S. L. K., Drummond, A. J., Pybus, O. G., Shapiro, B., Barroso, H., et al. (2007). Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. *PLoS Comput. Biol.* 3:e29. doi: 10.1371/journal.pcbi.0030029
- Liang, H., Zhou, W., and Landweber, L. F. (2006). SWAKK: a web server for detecting positive selection in proteins using a sliding window substitution rate analysis. *Nucleic Acids Res.* 34, W382–W384. doi: 10.1093/nar/gkl272
- Markham, R. B., Wang, W.-C., Weisstein, A. E., Wang, Z., Munoz, A., Templeton, A., et al. (1998). Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. *Proc. Natl. Acad. Sci. U.S.A.* 95, 12568–12573. doi: 10.1073/pnas.95.21.12568
- McCutchan, F. E. (2006). Global epidemiology of HIV. *J. Med. Virol.* 78, S7–S12. doi: 10.1002/jmv.20599
- Michael, N. L. (1999). Host genetic influences on HIV-1 pathogenesis. *Curr. Opin. Immunol.* 11, 466–474.

- Moratorio, G., Iriarte, A., Moreno, P., Musto, H., and Cristina, J. (2013). A detailed comparative analysis on the overall codon usage patterns in West Nile virus. *Infect. Genet. Evol.* 14, 396–400. doi: 10.1016/j.meegid.2013.01.001
- Nakamura, Y., Gojobori, T., and Ikemura, T. (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28:292. doi: 10.1093/nar/28.1.292
- Nasrullah, I., Butt, A. M., Tahir, S., Idrees, M., and Tong, Y. (2015). Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evol. Biol.* 15:174. doi: 10.1186/s12862-015-0456-4
- Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426.
- Peden, J. F. (2000). *Analysis of Codon Usage*. Nottingham: University of Nottingham.
- Rangel, H. R., Weber, J., Chakraborty, B., Gutierrez, A., Marotta, M. L., Mirza, M., et al. (2003). Role of the human immunodeficiency virus type 1 envelope gene in viral fitness. *J. Virol.* 77, 9069–9073. doi: 10.1128/JVI.77.16.9069-9073.2003
- Reitter, J. N., Means, R. E., and Desrosiers, R. C. (1998). A role for carbohydrates in immune evasion in AIDS. *Nat. Med.* 4, 679–684. doi: 10.1038/nm0698-679
- Ross, H. A., and Rodrigo, A. G. (2002). Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J. Virol.* 76, 11715–11720. doi: 10.1128/JVI.76.22.11715-11720.2002
- Roy, A., Mukhopadhyay, S., Sarkar, I., and Sen, A. (2015). Comparative investigation of the various determinants that influence the codon and amino acid usage patterns in the genus *Bifidobacterium*. *World J. Microbiol. Biotechnol.* 31, 959–981. doi: 10.1007/s11274-015-1850-1
- Sandonís, V., Casado, C., Alvaro, T., Pernas, M., Olivares, I., García, S., et al. (2009). A combination of defective DNA and protective host factors are found in a set of HIV-1 ancestral LTNP. *Virology* 391, 73–82. doi: 10.1016/j.virol.2009.05.022
- Seo, T.-K., Thorne, J. L., Hasegawa, M., and Kishino, H. (2002). Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics* 160, 1283–1293.
- Shackelton, L. A., Parrish, C. R., and Holmes, E. C. (2006). Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.* 62, 551–563. doi: 10.1007/s00239-005-0221-1
- Sharp, P. M., and Li, W.-H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24, 28–38. doi: 10.1007/BF02099948
- Sharp, P. M., Tuohy, T. M., and Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14, 5125–5143. doi: 10.1093/nar/14.13.5125
- Simmonds, P., Zhang, L. Q., McOmish, F., Balfe, P., Ludlam, C. A., and Brown, A. J. (1991). Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 env sequences in plasma viral and lymphocyte-associated proviral populations *in vivo*: implications for models of HIV pathogenesis. *J. Virol.* 65, 6266–6276.
- Strunnikova, N., Ray, S. C., Livingston, R. A., Rubalcaba, E., and Viscidi, R. P. (1995). Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. *J. Virol.* 69, 7548–7558.
- Wang, B., Mikhail, M., Dyer, W. B., Zaunders, J. J., Kelleher, A. D., and Saksena, N. K. (2003). First demonstration of a lack of viral sequence evolution in a non-progressor, defining replication-incompetent HIV-1 infection. *Virology* 312, 135–150. doi: 10.1016/S0042-6822(03)00159-4
- Weiss, R. (1993). How does HIV cause AIDS? *Science* 260, 1273–1279.
- Wong, E. H. M., Smith, D. K., Rabadan, R., Peiris, M., and Poon, L. L. M. (2010). Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. *BMC Evol. Biol.* 10:253. doi: 10.1186/1471-2148-10-253
- Wright, F. (1990). The effective number of codons used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9
- Wyatt, R., and Sodroski, J. (1998). The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 280, 1884–1888.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yoshida, I., Sugiura, W., Shibata, J., Ren, F., Yang, Z., and Tanaka, H. (2011). Change of positive selection pressure on HIV-1 envelope gene inferred by early and recent samples. *PLoS ONE* 6:e18630. doi: 10.1371/journal.pone.0018630
- Zeller, J. M., McCain, N. L., and Swanson, B. (1996). Immunological and virological markers of HIV-disease progression. *J. Assoc. Nurses AIDS Care* 7, 15–27. doi: 10.1016/S1055-3290(96)80034-3
- Zhou, J. H., Zhang, J., Sun, D. J., Ma, Q., Chen, H. T., Ma, L. N., et al. (2013). The distribution of synonymous codon choice in the translation initiation region of dengue virus. *PLoS ONE* 8:e77239. doi: 10.1371/journal.pone.0077239

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Roy, Banerjee and Basak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.