博士論文

Molecular evolution of HIV-1 Tat

(HIV-1 Tat 遺伝子の分子進化)

.

Chandra Nath Roy

Contents

I. Summary · · · 3

II. Introduction · · · 6

III. Objectives · · · 14

IV. Materials and Methods · · · 15

V. Results · · · 24

VI. Discussions · · · 35

VII. Conclusion · · · 42

VIII. Acknowledgements · · · 43

IX. References · · · 46

X. Figures · · · 63

XI. Tables · · · 86

XII. Appendices · · · 95

Summary

The globally circulating strains of HIV-1 can be broadly classified into four phylogenetic groups: M (the "major"), O (the "outlier"), N and P. The M subgroup contributes to more than 98% of the globally circulating strains, and it is further subdivided into A, B, C, D, F, G, H, J, K and circulating recombinant forms (CRFs). Because of the error prone nature of the virus genome, HIV-1 is continuously undergoing genetic changes and subsequent transmission. Hence, the clinical management and control of HIV-1 are still facing challenges despite huge efforts to develop vaccines and therapeutics. HIV-1 has three structural and six regulatory proteins. HIV-1 Tat (*trans*-acting activator of transcription), is a 101-amino acid regulatory protein encoded by two exons (exon 1: 1 to 72 residues, and exon 2: 73 to 101 residues). Tat protein has been divided into six different functional domains (5 domains in exon1 and 1 domain in exon 2). It plays essential roles in the replication through viral mRNA and genome transcription from the HIV-1 LTR promoter. Tat is therefore a promising target for developing vaccines and therapeutics while Tat undergoes continuous amino acid substitutions. As a consequence, the virus escapes from host immunity. My study is therefore aimed to investigate the molecular evolution of HIV-1 Tat in globally circulating HIV-1 stains.

For my study, I curated the HIV-1 Tat sequences from the Los Alamos National Laboratory (LANL) HIV sequence database. My study data sets contains the amino acid sequences from a total 45 countries from six different continents; Africa (8 countries), North America (5 countries), South America (11countries), Asia (9 countries), Oceania (1 country), and Europe (11 countries).

I first prepared a dataset of a total 1179 sequences to investigate the intersubtype genetic variation of Tat exon-1 in all major subtypes of HIV-1: A, B, C, D, F, G, H, J, and K from the available globally circulating strains up to 2013. Then, I further investigated the genetic divergence in full-length Tat of HIV-1 subtypes B and C. For that, I prepared a new data set containing full-length HIV-1 Tat sequences of subtypes B (n=493) and C (n=280) collected from globally circulating strains. I performed phylogenetic tree, mutation, and selection pressure analyses to understand the genetic variability and evolution of HIV-1 Tat. Bayesian graphical model analysis (BGM) was done to detect the co-evolving sites from multiple alignments of amino acid sequence data and to identify significant associations among sites. Moreover, I also analyzed the evolutionary dynamics of Tat in subtypes B and C in a subset of dataset by a Bayesian coalescent-based approach using the BEAST package.

The mean nucleotide divergence (%) among the analyzed Tat exon 1 sequences of HIV-1 subtypes A, B, C, D, F, G, H, J and K were 88, 89, 90, 88, 86, 89, 88, 97, and 97%, respectively. The second and fifth domains were found comparatively more variable among all subtypes. Site-by-site analysis of positive selection revealed that several positions in all subtypes were under significant positive selection. Positively selected sites were found in the first domain at positions 3, 4, 19; in the second domain at positions 24, 29, 32, 36; in the third domain at position 40; and in the fifth domain for the rest of the positions for all subtypes. Positions 58 and 68 in the fifth domain were positively selected in subtypes A, B, C and B, C, F, respectively. I also observed high variability within positively selected sites in amino acid positions.

In full length Tat sequences (both exon 1 and 2), positively selected sites, such as positions 68 and 70 in both subtypes, were located in the Tat-transactivation responsive RNA (TAR) interaction site located at fifth domain. Positively selected sites were also found in the sixth domain of exon 2, such as positions 75, 77, 80, 81 and 87 for both subtypes. Mutations of the exon 2 were found particularly at intimately-networked coevolving sites with exon1 in the fourth, fifth, and sixth domains. The mean estimated nucleotide substitution rates for Tat in HIV-1 subtypes B and C were $1.53x10^{-3}$ (95% highest probability density- HPD Interval: 1.09 $x10^{-3}$ to $2.08x10^{-3}$) and $2.14x10^{-3}$ (95% HPD Interval: 1.35 $x10^{-3}$ to $2.91x10^{-3}$) per site per year, respectively, which is relatively low compared to HIV-1 structural proteins. In molecular clock analysis, the median times of the most recent common ancestors (tMRCA) were estimated to be around 1933 (95% HPD, 1907–1952) and 1956 (95% HPD, 1934–1970) for subtypes B and C, respectively.

My study provides comprehensive and updated information on the HIV-1 Tat genetic evolution in recently circulating strains at global level. The study results would be an important foundation for developing effective vaccine and Tat-targeted therapeutics. The study findings also accomplish to a better understanding the evolutionary changes in regulatory protein like Tat in process of HIV-1 evolution. Taken together, my study results contribute to those aforementioned aspects to control the global spread of the HIV/AIDS.

Introduction

The basic facts of HIV

The human immunodeficiency virus (HIV) (genus: *Lentivirus*, family: *Retroviridae)* is a single-stranded, positive-sense, enveloped RNA virus[1,2,3]. The length of the viral genome is about 9.8 kb including two viral long-terminal repeats (LTRs) located at both ends when integrated into the host genome [3]. The virus has two phylogenetically distinct types: HIV-1 and HIV-2. HIV-1 spreads all over world [3], however, HIV-2 confined to only West Africa [4]. The HIV-1genome encodes for three genes: Gag, Pol, and Env for making the structural proteins for new virus particle. In addition, the virus also encodes six genes: Tat, Rev, Vpu, Vpr, Vif, and Nef [3]. The regulatory genes encode proteins for controlling the ability of HIV to infect the host cells, replicate for the production of new viral copies of virus [3].

The consequence of HIV infection in human and epidemiology of HIV-1 infection

HIV mainly infects the human immune system and weakens host-defense systems. As a consequence, the infected person becomes immunodeficient and at risk of getting a wide range of infections and some

types of cancer. Lastly, HIV infection causes *Acquired Immunodeficiency Syndrome* (AIDS), which is regarded as the advanced condition of HIV infection. Chronic or infection for a long duration with ultimate progression to AIDS is the hallmark of HIV infection. HIV/AIDS is still one of leading causes of death all over the world. More than 35 million people across the globe are now living with HIV/AIDS, and in 2013, it caused an estimated 1.5 million deaths globally [5]. The globally circulating strains of HIV-1 can be broadly classified into four phylogenetic groups: M (the "major"), O (the "outlier"), N and P [3]. The M subgroup contributes to more than 98% of the globally circulating strains, and it is further subdivided in to A, B, C, D, F, G, H, J, K and circulating recombinant forms (CRFs) [3]. The estimated global prevalence of subtype C accounted for nearly half (48%) of all global infections; subtypes A and B caused 12% and 11% of infections, respectively, followed by subtypes G (5%) and D (2%). Subtypes F, H, J and K together caused less than 1% of infections worldwide [6].

The evolution of HIV-1

Because of the error prone nature of the virus genome, HIV-1 is continuously undergoing genetic changes and subsequent transmission. Hence,

the clinical management and spread of HIV-1 are still facing challenges despite huge efforts to develop vaccines and therapeutics. Since it was first reported in 1981[7], the HIV-1 virus has shown high genetic variability, generating highly divergent strains. The transmission of those HIV-1 variants is a dynamic process because of continuous intermixing among different strains due to increasing human mobility across the globe. From the HIV-1 virus' perspective, diversity in the genome significantly impacts the functionality and immunogenicity of the virus itself, and this diversity is estimated to be about 1% per year [8,9]. The possible reasons for such divergence are high error rate during reverse transcription [10] and increased virus production [11] due to comparatively fast replication kinetics [12]. On the other hand, HIV-1 rapidly evolves during the course of infection in response to selective pressure induced by the host's immune response [13]. The evolutionary characteristics of HIV-1 have mostly studied focusing its structural genes, Gag, Pol and Env in subtypes B and C in different high-risk groups [14-19]. However, the evolutionary changes in the regulatory proteins in the process of HIV-1 evolution have not been studied well.

HIV-1 Tat and its functions

Tat (*trans*-acting activator of transcription) is one of HIV-1 regulatory

proteins having 101-amino acid encoded by two exons (exon 1: 1 to 72 residues, and the exon 2: 73 to 101 residues) in most of the clinical isolates [20]. However, only exon 1 can exert all transcriptional activating properties and replicate the new viruses [3]. Tat plays a pivotal role in viral transcription and replication. It directly enhances the HIV-1 replication through interaction with HIV-1 long terminal repeat (LTR) promoter [21]. Tat can induce the expression numerous of cellular genes; it also acts as a neurotoxic protein [22], leading to dementia in advanced HIV-1 infected cases, known as AIDS dementia complex. As shown in Figure 1, Tat has been categorized into six different functional domains [3, 20, 23]. The first domain (residues 1 to 21) is the N-terminal acidic domain consisting of a Pro-rich tract and a conserved Trp residue at the position 20 (Trp20) [3, 20]. The second domain (residues 22 to 37) contains a highly conserved seven Cys tract at positions 22, 25, 27, 30, 31, 34, and 37 [3, 23]. The third domain (residues 38 to 48) contains a hydrophobic core sequence: 43LeuGlyIleSerTryGly48 [23]. The fourth domain (residues 49 to 57) is a positively charged region composed of a well-conserved arginine-rich motif, 49-ArgLysLysArgArgGlnArgArgArg-57, and acts as a transactivation response element (TAR) binding domain [24]. This domain has an extraordinary property

for nuclear localization [25, 26] and protein transduction, thus it has also been used to deliver various molecules inside the cells *in vitro* [27-32]. The fifth domain (residues 58 to 72) is a Gln rich region [4, 8], and the fourth and fifth domains (residues 49 to 72) are known as basic domains for transactivation [3, 23, 33]. The sixth domain (residues 73 to 101) encoded by the exon 2 is known as RDG domain; this domain contains the highly conserved GluSerLysLysLysValGlu motif, which is related to optimal HIV-1 replication *in vivo* [34], thereby, the region may contribute to viral infectivity and binding to cell-surface integrins [35, 36, 37]. Taken together, Tat plays crucial role in HIV-1 transcription and replication as well as in cell surface binding [38].

Tat-based vaccines and antivirals

Developing vaccines by blocking virus replication and disease progression is one of strategies for anti-HIV vaccine development [39]. Regulatory proteins like HIV-1 Tat is one of the promising target in this perspective as Tat plays a versatile role in viral gene expression, transcription, replication and consequently in HIV-1 disease progression. Therefore, the possibilities of Tat targeted vaccines have been discussing since long time [40]. Several vaccines were also designed with an aim to inhibit circulating Tat protein,

thereby, halting the disease progression [41-44]. It has been recently observed that anti-Tat antibodies can exert protective role in the HIV-1 disease progression [45]. In clinical trials, HIV-1 Tat B cell epitope **v**accine showed promising results in reducing HIV-1 viral loads. Notably, a higher frequency of anti-Tat antibodies has been detected in asymptomatic HIV-infected individuals compared to progressed patients [46]. However, those vaccines sometimes exert ambiguous results in different patient groups, and mutations in the antigenic sites have been suggested an important cause of vaccine ineffectiveness for those cases [47, 48]. Examining the genetic variation at amino acid level is therefore useful to define the epitope presentation on Major Histocompatibility Complex (MHC) class I antigen for afferent immune response [48, 49]. Investigating the genetic variation of Tat in globally circulating strains would be an important foundation for making a consensus sequence for effective vaccine development.

On the hand, Tat binds to hairpin structure, transactivation-responsive (TAR) RNA for transactivation and recruits a pair of host proteins, cyclinT/ cdk9 to the RNA [3]. Blocking that Tat-dependent HIV transcription is one the clever anti HIV-1 therapeutic strategy. Therefore, several antivirals have been designed

to inhibit that Tat-TAR RNA interaction [50, 51]. For example-peptide-based inhibitors such as peptoids [52], Tat-mimicking substances, cyclic b-hairpin peptide causing conformational change have been developed and tested for this purpose [51, 53-54].

Genetic variation of HIV-1 Tat and its impact

Like other HIV-1 protein, Tat also exerts subtype and region specific genetic divergence which has a huge impact on its functionality and pathogenesis of diseases progression. It has been previously reported that the genetic variations observed in Tat in major subtypes like B and C have impact on the functionality of protein and also its effect on the HIV-1 pathogenesis [55, 56]. It has been previously reported that Tat protein in subtype B in exhibits higher ability in viral replication and induction of TNF- *α* and and IL-6 expression compared to subtype C [57]. Furthermore, important structural and functional differences have also been reported in HIV-1 subtypes B and C. On the other hand, in virto experiments it has been found that HIV-1 Tat in subtype C showed greater transcriptional activity in the Jurkat CD4+ T cell line compared to subtypes B and E because of variations in the C-Tat sequence [58]. However, the genetic variation in other subtypes than B and C are not well-studied. The

genetic variation has been examined only in some regionally circulating strains. The Tat functional domain specific genetic variation has not been comprehensively studied yet. Vaccine effectiveness are greatly depends on designing the epitope in conserved sites based on the globally circulating strains. Therefore, any changes in antigenic sites greatly impair the efficacy of developed vaccines and anti-Tat antibodies [48, 59]. On the other hand, any mutations occurred in any TAT-TAR binding site dramatically reduce or impair the effectiveness of those therapeutics [3]. Thus, considering the importance of Tat in HIV pathogenesis and disease progression and its implication in developing vaccines, there should have a complete and compressive overview of its genetic variation at global level in recently circulating strains. It is therefore worthwhile to investigate the genetic variation of HIV-1 Tat for better understanding of its diversity in commonly circulating subtypes, which has an impact on clinical outcome of HIV disease as well as success of Tat-based vaccination and Tat-targeted antagonists.

Objectives

1. To investigate the genetic variation of Tat exon 1 in all subtypes of HIV-1 M: A, B, C, D, F, G, H, J, and K.
2. To examine the genetic variability of full-length Tat protein at the global level in major subtypes, B and C
3. To investigate the evolutionary changes in HIV-1 regulatory protein, Tat in the process of HIV-1 evolution over time

Materials and Methods

Sequence data

For my study, the HIV-1 sequences were obtained from the Los Alamos National Laboratory (LANL) [60] and I prepared a total of three following datasets.

Dataset 1

The HIV-1 Tat exon 1 amino acid sequences of were curated using the following options, "Alignment type: Web (all complete sequence), Year: up to 2013, Organism: HIV-1, Region: TAT (exon 1), Subtype: All subtypes belonged to M (except A-K recombinant), DNA/Protein: DNA, Format: FASTA". According to the LANL website, the curated alignments were published on the year after the sequences became available; therefore, the "up to 2013" alignment contains all sequences published before 2014. On July 3, 2014, 2156 sequences were initially downloaded using 'one sequence per patient' option. The problematic sequences as defined by LANL's pre-built search criteria, were excluded. Sequences likely to have been contaminated with laboratory strains were also excluded through this process. As recommended by LANL, I further checked to ensure that none of the sequences contained ambiguous nucleotides, hypermutation, minor insertions, minor deletions, premature

termination codons, or different sequences from the same strains (verified by the strain name). Through this sorting process, I made a study data set of a total of 1179 sequences for subtypes A to K, which contained 71 amino acids encoded by 213 nucleotides (nt) from the positions 5831 to 6045 nt in exon 1 of the HIV-1_{HXB2} genome (GenBank accession No. K03455). The sequences data was obtained together with the information about the host, subtype, isolation year, and isolated country. The details information of the sequences regarding the continent of origin and other aforementioned information with GenBank accession numbers were mentioned in Appendix Tables 1a and 1b. A multiple-sequence alignment program, ClustalW, was used to align the nucleotide sequences without any gap. Consensus sequences were generated by aligning circulating primary isolates and selecting the most common nucleotide at each position. In this study, the three most common subtypes were B, C, and A, were shown here as 50.7% (n=598), 28.5% (n=336), and 10.9% (n=128) of the sequences obtained, respectively. The following amounts of the total number of sequences were for other subtypes: D (4.4%, n=52), G (2.5%, n=30), F (2.3%, n=27), H (0.3%, n=4), J (0.2%, n=2), and K (0.2%, n=2). It is worthwhile to mention that this distribution of HIV-1 subtypes to some extent

reflects the publicly available and published sequence data, not the global prevalence of subtypes that constitute the HIV pandemic.

Dataset 2

In order to investigate the genetic divergence of full-length Tat in HIV-1 in subtypes B and C, Tat sequences were obtained from the 'Web alignment' in the LANL. The sequences were downloaded on February 2, 2014. A total of 2156 sequences were downloaded initially. Sequence data of the other subtypes than subtype B or C, and of circulating recombinant forms (CRFs) were then excluded. Consequently, totals of 713 and 353 sequences for subtype B and subtype C, respectively, were obtained. The sequence data for full-length coding regions of Tat were used after eliminating the problematic sequences. The reference strains for subtype B and C sequences were also obtained from the Los Alamos HIV-1 sequence data base. The reference sequences were also downloaded on February 2, 2014. As the reference sequences, accession numbers AY423387 (Europe), AY173951 (Asia), and AY331295 (North America) for subtype B and AF067155 (India), U52953 (South America), and AY772699 (Africa) for subtype C were used for alignment. Finally, I prepared a data set of totals of 493 and 280 sequences for subtype B and subtype C, respectively, which contained 100

amino acids encoded by 300 nucleotides (nt) from the positions 5831 to 6045 nt in exon 1 and 8379 to 8463 nt in exon 2, respectively of HXB2 genome (GenBank accession No. K03455). A multiple-sequence alignment of the nucleotide sequences (without any gap) was made using the ClustalW [61]. The divergence of sequences was schematically visualized using Weblogo [62]. The sequence data sets contained sequences from a total 45 countries from six different continents; Africa (8 countries), North America (5 countries), South America (11countries), Asia (9 countries), Oceania (1 country), and Europe (11 countries). Like Dataset 1, the sequences data was obtained together with the information about the host, subtype, isolation year, and isolated country. The details information of the sequences regarding the continent of origin and other aforementioned information with GenBank accession numbers were mentioned in Appendix Tables 2a and 2b.

Data set 3

Next, another subset of data was made from the dataset 2 to investigate the evolutionary dynamics and molecular dating of Tat in HIV-1 subtypes B and C. In order to better determine the relationship between sequences from different

geographic areas, I attempted to include sequences from every country for both subtypes. Through this sorting process, I finally prepared a dataset of 135 and 53 sequences for subtypes B and subtype C, respectively. Like dataset 1 and 2, the sequences data was obtained together with the information about the host, subtype, isolation year, and isolated country. The details information of the sequences regarding the continent of origin and other aforementioned information with GenBank accession numbers were mentioned in Appendix Tables 3a and 3b.

Phylogenetic tree analysis

In case of HIV-1 Tat exon 1 (data set 1), the phylogenetic tree was constructed using the Neighbor-Joining method. Evolutionary distances were computed following the Tamura-Nei [63] method and are in units representing the number of base substitutions per site. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories). The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted using Molecular Evolutionary Genetic Analysis (MEGA) version 6 [63]. The reliability was estimated from 1,000 bootstrap replicates. Simian immunodeficiency virus (SIV) sequence,

CPZ.US.85.US_Marilyn.AF103 was used as the out-group to root the tree. FigTree v1.3.1 (available at: http://tree.bio.ed.ac.uk) was used for the visualization of the tree [65].

In case of full-length Tat (dataset 2), the phylogenetic tree was constructed by the Maximum Likelihood method based on the General Time Reversible model, which is the best model tested by model test in Molecular Evolutionary Genetic Analysis (MEGA) version 6 [63]. A discrete gamma distribution was used to measure the evolutionary rate differences among sites (5 categories) and the analyses were done using 1,000 bootstrap replicates. The tree was rooted by using following reference strain sequence: simian immunodeficiency virus (SIV), CPZ.US.85.US_Marilyn.AF103.

Selection analysis

Global (ω) value of relative rates of non-synonymous (dN) and synonymous (dS) substitutions were calculated to measure the positive selection strength [66]. All analyses were carried out using the online Datamonkey facility [66-68] after identifying the best fit model from every possible time-reversible model. Positive selection pressure analysis was performed at whole gene and site-by-site codon level using three likelihood

methods: single-likelihood ancestor counting (SLAC), fixed effects likelihood (FEL), and interior branches Fixed Likelihood (iFEL). Briefly, in the SLAC method, the mean ratio of non-synonymous changes per non-synonymous site (dN) and the synonymous changes per synonymous site (dS) were measured using SLAC which considered inferred ancestral sequences for each internal node in a phylogeny using a codon model and then, calculated the synonymous and non-synonymous mutations by comparing each codon to its immediate ancestor. The FEL method is based on maximum-likelihood estimates. This method estimates the ratio of non-synonymous to synonymous substitutions on a site-by-site basis for the entire tree. iFEL is principally the same as FEL, except that selection is only tested along the internal branches of the phylogeny. To detect co-evolving sites from multiple alignments of amino acid sequence data and to identify significant associations among sites, Bayesian graphical models (BGM) method was used in Spidermonkey through the Datamonkey web-based interface [69].

Estimation of evolutionary rates and dates by molecular clock analysis

For examining the evolutionary dynamics of HIV-1 full length Tat (data

set 3), a multiple-sequence alignment of the nucleotide sequences (without any gap) was constructed using ClustalW. I first tested sequences saturation (phylogenetic signal) which is useful for the phylogenetic analysis. Then, I performed Xia's test implemented in DAMBE (available at: http://dambe.bio.uottawa.ca/dambe.asp). The substitution rate and timing of the most recent common ancestor (tMRCA) of subtypes C and B were estimated by BEAST v1.8.1 [71] using dataset 3.The GTR+Γ+I substitution model was used in all runs as it was selected by the program, JmodelTest 2.1.6 [70]. Following the procedure described in the previous study [15], BEAST analysis was done using normal distribution with a mean of 2.5×10^{-3} and standard deviation 5×10^{-4} prior using three different molecular clock models: strict relaxed, uncorrelated exponential, and relaxed uncorrelated lognormal [72,73]. The strict clock model was best for subtype C; the exponential relaxed clock model was best for subtype B, as revealed by a Bayes factor analysis performed in Tracer v.1.6. However, in the case of subtype B, this model showed extremely large 95% HPD intervals that caused convergence problems. I therefore considered strict clock, which was the second best model for subtype B (With sub-sampling performed every 2.56105 steps, MCMC simulations were run for 2.56108 chain steps and

analyzed by Tracer. For each tested prior, I considered a value above 250 to be an effective sample size (ESS). The Maximum Clade Credibility (MCC) tree ,with time scale, was created using Tree Annotator v1.8.1 with a burn-in of the first hundred trees and illustrated using Fig tree version 1.4.2 [74].

Demographic dynamics

Phylodynamic analysis of the HIV-1 virus was performed using the Full-length Tat (data set 3) to determine epidemiological and phylogenetic patterns of global diversity. The phylodynamic histories of HIV-1 subtypes B and C were estimated from the Tat gene sequences using Bayesian skyride, which is an extension of the Bayesian Skyline Plot (BSP) implemented in BEAST. In Bayesian skyride analysis, changes in population size between the intervals are smoothed, and thus, I could assume gradual changes in population size using the time-ware prior [73]. Notably, Gaussian Markov random field (GMRF) smoothing prior is used in Bayesian skyride analysis tool. A marginal posterior distribution of the demographic inference was estimated using the Bayesian MCMC method and the effective population size at the most recent time of sampling as well. Those posterior samples were analyzed using the Tracer program.

Results

Phylogenetic tree analysis

To reveal the intersubtype sequence identity of HIV-1 Tat exon1, distant analysis was performed using the Maximum Composite Likelihood method. Sequence identities and mean nucleotide divergence for all subtypes of HIV-1 Tat are summarized in Table 1. In phylogenetic tree analyses (Figure 2), the evolutionary distances of different subtypes of Tat exon 1, estimated by the Neighbor Joining (NJ) method, showed that subtype C was more closely related with CPZ.US.85.US_Marilyn.AF103 strain than other subtypes. From that analysis, it was also revealed that subtypes B and D were more closely related to each other than to other subtypes; notably, subtype D is considered an early clade of subtype B African variant. The genetic distance between the rest of the subtypes and the subtype C was small, especially between subtype A and subtype C. There was no evident monophyletic distribution by year of isolation or by country in any subtype which may indicate that the infections from subtypes were transmitted in a diverse and heterogeneous manner.

In case of full-length Tat (dataset 2), analysis of the phylogenetic relationship was performed using the maximum likelihood (ML) tree based on the nucleotide

sequences of subtypes B and C (Fig 3a and b). The phylogeny of Tat in subtype B featured with well intermixing of sequences among the different continents. The wide genetic diversity and several poorly defined clusters were observed in the sequences particularly in the strains from USA, South America and Europe. However, Asian strains, especially Thai and Korean strains clustered together distinctly, may result from different single introduction. In subtype C, it was hard to define clear clustering (Figure 3b). In the tree, the majority of subtype C strains were found in Africa, and those African stains did not show any monophyletic distribution, and there was a interspersing of Asian, South American and European strains with African stains. Overall, I observed a slow and continuous introduction of new HIV-1 strains in different parts of the world with repeated cross-border transmission as reflected by diffuse distributions and intermixing of the different HIV-1 variants in both subtypes B and C.

Mutation analysis

The following percentages of conserved positions in different subtypes were found: subtype A, 42%; subtype B, 27%; subtype C, 23%; subtype D, 42%; subtype F, 55%; subtype G, 59%; subtype H, 69%; subtype J, 87%; and subtype

K, 89% (Figure 4 and Table 2). Overall, subtypes B and C exhibited more genetic variation compared to other subtypes.

In the first domain (residues 1-21), most of the positions were found conserved in all subtypes. Subtypes B, C, A and F contained more substituted amino acids, especially at positions 7, 12, 19, and 21. At positions 7 and 12, the predominant amino acid was asparagine in all subtypes except subtype B where arginine and lysine were predominant, respectively. Notably, positions 7 and 12 are well-recognized antigen binding sites [48]. At position 19, all subtypes except B and C mostly contained lysine, whereas there are 13 and 10 different types of amino acid substitutions present in subtypes B and C, respectively. Moreover, at position 21, the predominant amino acids were alanine and proline in all subtypes except subtype F, which contained proline, and subtypes J and K contained alanine only.

The second domain (residues 22–37) was the most variable region for all subtypes. At position 24, commonly found amino acids were lysine and asparagine in subtypes A, B, C, D, F, and G, asparagine in subtype H, and interestingly, glutamine in subtypes J and K. However, more variations were

observed in that position in subtypes A, B, and C. At position 29, amino acid changes were commonly observed, and most of the subtypes contained mainly lysine except subtype C, which contained histidine, and in subtype J and K, it was arginine. At position 32, phenylalanine was present in all subtypes except H, J and K, where it was tyrosine. At position 36, a large number of variations were observed, particularly in subtypes B, C and A, in which valine was commonly found; in other subtypes, it was leucine.

In the third domain (residues 38 to 48) and the fourth domain (residues 49 to 57), most of the positions were relatively conserved except 39 and 40. Among the studied subtypes, variation of amino acids was mostly observed in subtype B.

The fifth domain (residues 58 to 72) was found comparatively more variable in all subtypes. All positions except position 65 (mainly in subtypes B, C, D and H) and 66 (in all subtypes) were relatively conserved. Positions 58 to 60, 62 to 64, 67, and 68 were found to be the most variable in subtype B.

In mutational analysis of full-length Tat of HIV-1 (dataset 2), the amino acid sequences divergence in both subtypes B and C were illustrated by

sequence logo (Figure 5a and b). The detailed substitution positions in each domain were described below following HXB2 numbering in both subtypes (the position of amino acid, aa, which had major changes).

In the first domain (residues 1 to 21), positions 1, 11, 14, 15, and 16 were found completely conserved in both subtypes; in addition, Asp5 in subtype C was also completely conserved. Notably, Lys12Asn (90%) and Ala21Pro (53%) were frequently observed only in subtype C.

In the second domain (residues 22 to 37), among the conserved Cys positions, Cys27 in subtype B and Cys22 and Cys34 in subtype C were completely conserved. Lys28 in subtype B and His33 in subtype C were also found as completely conserved in my analysis. Cys31Ser was observed in subtype C (83%), and all sequences of subtype C in my data sets were substituted as Phe32Tyr.

In the third domain (residues 38 to 48), mutations in all positions were found in this domain except Lys41 in subtype C and Gly48 in subtype B. In subtype C, 80% sequences were substituted as Ile39Gln. In both subtypes, most of the sequences contained Ala42Gly (99%).

In the fourth domain (residues 49 to 57), the consensus Arg-rich motif

provided two of the key functions of Tat, nuclear localization and membrane transduction [25-32], a total of 6 Arg at the positions 49, 52, 53, 55, 56, and 57. There was no conserved Arg position found in subtype B. Only positions 49 and 55 were found conserved in subtype C accompanied with Arg57Ser substitution predominantly (88%).

In the fifth domain (residues 58 to 72), only Gln66 was conserved in both subtypes; besides, Gln72 was almost completely conserved in subtype B and subtype C. For other sites, His59Pro substitution was apparent in both subtypes (81% and 95% in subtype B and C, respectively). In addition, Asn61Asp (64%) and Asn67Val (44%) in subtype B, and also Gln60Pro (89%), Asn61Ser (92%), Gln63Glu (75%), Thr64Asp (87%), Ala67Asn (73%), and Lys69Ile (61%) substitutions were frequently found in subtype C.

The sixth domain is (residues 73 to 101) also known as the RGD (Arg-Gly-Asp) domain, and RGD is a ligand for several integrins, which play important roles in HIV replication and cell surface binding [35, 75, 76]. Even though representative conserved domain in Tat was well maintained in both subtypes, genetic variation in the sixth domain was observed of our data set. In subtype B, 87Ser (77%) was most frequently found with many different

substitutions in small percentages at other positions of exon 2. In subtype C, high frequencies of substitutions were observed, including Thr74Leu (92%), Pro77Thr (79%), Pro84Ser (80%), Lys85Glu (89%) and 87Ser (95%) indicating that this domain is also variable. In Tat exon 2 of subtype C, the frequently observed substitutions were 93Ser (98%) and 94Lys (97%). Overall, genetic variation was found in the fourth, fifth, and sixth domains, which may have an impact on the functional properties of Tat, such as Tat-TAR interactions, protein transductions as well as cell surface binding and replication.

Selection analysis

The relative ratios (ω) of non-synonymous (dN) to synonymous (dS) substitutions for the nucleotide sequence of Tat exon 1 of all subtypes except subtypes H, J and K are shown in Figure 6. For subtypes H, J and K, selection pressure analysis could not be performed because of an insufficient number of eligible sequences. Interestingly, subtype F had the highest ratio value, followed by subtype B, then subtype G. In contrast, site-by-site tests of positive selection revealed that several positions in all subtypes were under significant positive selection by all three methods, the FEL, iFEL, and SLAC methods (Table 3). All positively selected codons selected by either one or two or three aforementioned

methods with statistical significance calculated are shown in Table 4. Positively selected sites were found in the acidic region at positions 3, 4, 19; in the cysteine-rich region at positions 24, 29, 32, 36; in the core domain at position 40; and in basic domain for the rest of the positions for all subtypes. Positions 58 and 68 in the basic domain were selected in subtypes A, B, C and B, C, F by all three methods, respectively. Position 70 was also commonly selected only in subtypes B and C. In subtype D, positions 19 and 62 were positively selected; these two positions were also selected in subtype A. There was no positively selected position found with the SLAC method in subtype F. However, positions 3 and 68 were selected using the FEL and iFEL methods. For subtype G, position 29 was positively selected by all three methods. Relative amino acid frequencies at positively selected codon position determined using the SLAC, FEL and iFEL methods in different subtypes are shown in Figure 7.

Remarkably, in case of full-length Tat (dataset 2), a total of 23 and 18 amino acid positions in subtypes B and C, respectively, were found under significant positive selection based on the FEL, iFEL and SLAC methods. The detailed positively selected codons with statistical significance were calculated with SLAC, FEL, and iFEL methods as shown in Figure 8 and Table 5. Position Thr40

in the third domain was positively selected in both subtypes by all three methods. I found several positively selected sites in the basic region of Tat, such as Ser68 and Ser70 in both subtypes; His59, Asn61, Ser62, and Thr64, His65 in subtype B; and Ala58, Ala67, and Leu69 in subtype C. In both subtypes of exon 2, Ser75, Pro77, Asp80, Pro81, and Ser87 were positively selected by all 3 methods.

Bayesian graphical model (BGM) analysis

To identify coevolving sites among the substitutions, especially positive selection sites, BGM analysis was further performed. A total of 10 and 6 pairs of interacting sites were identified in Full-length Tat in subtypes B and C, respectively (Figure 9). In general, the coevolving sites appeared to be more intimately networked and primarily involved in exon 2 with different important domains of exon 1 particularly with basic domain as shown in Figure 9a and b; however, the posterior probabilities were not statistically significant.

Substitution rate

The mean estimated substitution rates for subtypes B and C were 1.53×10^{-3} (95% HPD Interval: 1.09×10^{-3} to 2.08×10^{-3}) and 2.14×10^{-3} (95% HPD Interval: 1.35×10^{-3} to 2.91×10^{-3}) per site per year, respectively (using date set 3). The mean substitution rate was lower in subtype B than subtype C according to

analyses using all three models (Figure 10). Notably, evolutionary potential of a population is expressed by evolutionary rate, which indicates the number of mutations that are fixed per unit of time in the virus population. Bayesian MCMC method was used to estimate rates of evolution; the details are shown in Table 6.

Molecular dating analysis

Bayesian MCMC methodology was used to calculate the time of the most recent common ancestors (tMRCAs) of the sequences of subtypes B and C taken from the global sequences available in the Los Alamos database. The median dates of tMRCAs were estimated to be around 1933 (95% HPD, 1907 to 1952) and 1956 (95% HPD, 1934 to1970) for subtypes B and C, respectively, with a Bayesian model. According to my analysis, subtype C appeared in the global population two decades after the introduction of subtype B.

In case of subtype B, South American strains were introduced around the 1940s, whereas most of the North American strains appeared around the middle of 1950s. Asian strains evolved in the middle of the 1940s; strains originating from China, exhibited two peaks: one was from the middle of the 1960s to the middle of the 1970s, and another was from the middle of the 1980s to the middle of the 1990s. African strains mainly evolved around the middle of

1970s. European strains were introduced at various times, ranging from the early 1950s to the early 1980s (Figure 11a). In case of subtype C, African strains were introduced in the middle of the 1960s; however, most of these strains emerged during the 1970s. Asian strains were first introduced in the early 1970s. European strains emerged earlier than Asian strains, and American strains emerged around the late 1970s (Figure 12a).

Phylodynamic analysis

Temporal changes in genetic diversity (population size) were observed using a GMRF skyride coalescent model. This analysis demonstrates that the early expansion of subtype B was quite rapid, but the duration was relatively short with a dispersion occurring during the late 1980s. There was an exponential escalation in the occurrence of infections from 1960–1970 to the early 1990s with an increasing trend up to the 2010s (Figure 11b). The demographic history of HIV-1 C exhibited a relatively slow occurrence rate until the late 1980s, and then, there was a sharp increase in occurrences up to the end of the 1990s. Thereafter, the rate of occurrence has gradually slowed (Figure 12b). This trend was similar for all three models in both subtypes with a few variations.

Discussion

In my study, the different subtype distributions of Tat exon 1 were broadly similar (phylogenetically) to estimated distributions of other genes like Gag, Pol, and Env; notably, those sequences were obtained using the Los Alamos sequence compendium published in 2012 [77]. I observed a 15 to 22% variation among the Tat sequences of nine HIV-1 subtypes. On the other hand, as reported previously, the genetic variation in Env amino acid sequences was about 25 to 35% among those nine subtypes [78], whereas in Gag and Pol , the genetic variation were approximately 14% [79] and 22 to 23% [80. 81], respectively.

It is apparent in all trees that the branch lengths of the trees is not influenced by sequence length but rather by the density of nucleotide variation in each gene irrespective of phylogenetic method used. In other words, the faster a gene evolves, the longer the branch lengths will be versus more slowly evolving gene over the same time interval. The evolving pace of a gene is governed by several factors, most importantly by mutation rate and selection pressure. The variability in subtype distributions of HIV-1 Tat may increase over time as HIV-1 transmission in different regions of the world continues because of diverse risk

behaviors and high mobility of the infected people [82]. Notably, subtype-specific differences in transmissibility and pathogenesis may not always lead to significant changes in subtype distribution in different populations; rather, those differences might be ecologic [83]. Therefore, the association between specific subtypes and a unique mode of transmission has yet to be established.

The genetic complexity of the global HIV-1 epidemic is increasing due to the continuous molecular evolution of existing subtypes [84]. High variability within positively selected sites was observed possibly due to the fact that those sites are under diversifying selection. Notably, positive selection signifies genetic evolution and changes in antigenic sites for the binding of a virus to a host cell and also for exerting the host immune response against the infection [85]. In our analysis, the majority of selected amino acid residues were found in domains 2 and 5 of HIV-1. Domain 2 is known as the highly conserved cysteine rich domain containing cysteine at seven positions. As revealed by previous studies, any amino acid change in this domain affects the phenotypic characteristics of Tat since this region is responsible for making disulfide bonds [86, 87] and regulating CCR5 expression in monocytes [88]. Domain 5 plays an important role in cell surface binding and internalization, and any changes in this domain also

influence the functionality of Tat [35, 36]. Relatively few positively selected variants were detected at the N terminus of Tat exon 1 between codons 1 to 21, and TAR binding site, a functionally important region. In addition to being conserved and negatively selected, those regions contain several experimentally defined cytotoxic T-lymphocyte (CTL) epitopes [48]. Taken together, the amino acid changes of those two domains may have functional consequences as shown in a previous study [89]. However, further *in vitro* experiments are needed to prove those hypotheses.

I found that positions encoded in exon 2 such as Ser75, Pro77, Asp80, Pro81, and Ser87 were positively selected. As reported previously, exon 2 plays a role in the kappa-light-chain enhancer of activated B cell-(NF-*κ*B) dependent control of HIV-1 transcription in T cells [23, 90]. It has been previously reported that unlike laboratory-passaged strains, such as HIV-1_{HXB2} with premature stop codon at position 87, majority of HIV-1 strains encode 101 amino acids without any truncation beyond the position 86 [20]. I found Ser87 in subtypes B and C which was positively selected. As previously noted, the existence of two exons is essential to maintain stability of Tat *in vivo* [91]; therefore, this position may be crucial to maintain the functional stability of Tat. Again, mutations of the exon 2

were found particularly at intimately networked coevolving sites with exon 1 in the fourth, fifth, and sixth domains. This may also have an impact on HIV mRNA transcription through Tat-TAR interaction and initiation of reverse transcription, which were previously reported as influenced by genetic variation of Tat [23, 91].

Here I also studied the evolutionary dynamics and molecular dating of HIV-1 from the perspective of one of its regulatory proteins, Tat. Previously, HIV-1 evolutionary analyses were mainly performed on the virus' structural proteins, Env, Gag and Pol; however, updated genetic information targeting Tat is limited. My study results estimated that nucleotide substitution rates for Tat in HIV-1 subtype B and C were relatively low. Previous studies have shown that the estimated HIV-1 evolutionary rate was 1 to 17 $x10^{-3}$ substitutions per site per year [16, 92]. To the best of my knowledge, this is the first report demonstrating substitution rates for the Tat gene in major HIV-1 subtypes. As reported previously, nucleotide substitution rates for the HIV-1 surface glycoprotein, Env, were within the range of 7.3–8.7 $x10^{-3}$substitutions per site per year [92,93]. Notably, HIV-1 subtype B strains from both North America and Europe exhibited similar estimated evolutionary rates. The estimated nucleotide substitution rates of Pol in HIV-1 subtypes B and C strains from Central America were 2.8–4.7 $x10^{-3}$

and 1.5-2.3 x10^{-3}substitutions per site per year, respectively. Taken together, I found that the Tat gene exhibited a lower substitution rate than structural genes like Env and Pol. However, the reason for this variation is yet to be revealed.

The current study results of tMRCA for HIV-1 subtype B Tat in the United States was 1954– 1955 with a variable rate of evolution obtained using the Bayesian approach. However, the tRMCA was found around 1967 using the strict molecular clock model. For external genes like Env, the tMRCA for HIV-1 subtype B in the United States and Western Europe was likely to be around the middle of the 1950s to the late 1960s [16,93,94]. According to the Pol gene-based analysis, the spread of subtype B in America occurred through a single introduction event in the Caribbean region around 1964 (1950–1967). It has also been speculated that subtype B-driven epidemics in those regions might have started during that time [17]. On the other hand, subtype C was introduced around the same time in Africa [18]. As shown by tMRCA, the HIV-1 Env gene in the Thai-B strains was estimated to have emerged in the mid-1980s [95]. The tMRCA for the HIV-1 subtype C Pol gene was estimated to be in the early 1970s; notably, these strains were introduced into the general population in several instances through heterosexual or mother-to-child transmission [15].

Overall, the tMRCAs of HIV-1 Tat in subtypes B and C found in my study are in line with other genes of HIV-1.

In the current study, I observed that epidemiological trends proceeded relatively slowly. The time span of HIV-1 epidemic often encompasses months or even years owing to slow inter-host transmission kinetics. Moreover, the chronic nature of the disease leads to the persistence of epidemics through diffusing the infection among vulnerable groups[96]. As described previously, the extent of transmissibility is influenced by the time period of infection dynamics as well as by the changing vulnerability of hosts after epidemics [97]. Infection from multiple genotypes in an individual host provides a low level of cross-immunity [98], which may lead to slow evolutionary kinetics of HIV-1 proteins. It is noteworthy to mention that using the appropriate phylodynamic approach is imperative when correlating an epidemic event to its underlying genetic changes. In this regard, a coalescent tree analysis was performed to determine the nucleotide diversity for each time unit within a given phylodynamic framework. The viral dynamics was measured using GMRF skyride analysis, which enabled me to measure the evolutionary dynamics of HIV-1 based on Tat.

It is noteworthy to mention that in my study the analyses and interpretation were based on the available sequences in the LANL database. In my dataset, the total number of Tat exon 1 sequences under each subtype varies, and the number of sequences for HIV-1 subtypes B and C was greater than for other subtypes. Particularly, there were fewer available sequences for subtypes H, J, and K than other subtypes. However, as I mentioned earlier, this variation derives from the global distribution of HIV-1 subtypes; as a result, it influences the available deposited sequences in LANL. For example, there are more deposited sequences for subtypes B and C than for other subtypes as they account for about 60% of the total number of HIV-1 infections worldwide. Since the LANL sequence database is a widely recognized data source for HIV evolutionary studies, my analyzed data sets are representative and adequate enough for investigating the intersubtype genetic variation of HIV-1 Tat. However, the genetic variability of Tat in some subtypes like H, J, and K needs to be further clarified in a larger sample size. In evolutionary analysis, the current study results might not be completely compared with other published [16,19,93-95] results on structural genes due to a lack of uniformity in the analyzed variables, including insufficient information about the substitution model and limited

information about the protocol used. It is worthwhile to mention that the estimated time of origin or evolutionary rate often varies among the genes within a single virus strain. Moreover, homogeneity of an analyzed sequence dataset, methods of multiple sequence alignment, and different substitution models often generate some variations. Despite those variations and shortcomings, to the best of my knowledge, my study, for the first time, reports the evolutionary dynamics of HIV-1 Tat. Future study can focus on aforementioned issues to get an overview of the evolutionary dynamics of HIV-1 regulatory proteins.

Conclusion

Updating the genetic variation of Tat would be an important foundation for developing effective vaccine and intracellular single chain anti Tat-antibodies. Similarly, for developing Tat-targeted therapeutics, monitoring the target region at the amino acid sequence level should be prioritized as any major or minor genetic changes on those targets may lead to therapeutic failure. Genetic monitoring of HIV-1 Tat needs to be continued to understand the adaptive evolution of the HIV-1. My study results therefore accomplish in those above-mentioned aspects with a broader aim to control the global spread of the HIV/AIDS.

Acknowledgements

I am grateful to many people who supported my academic endeavor throughout my Graduate School of Medicine at Tohoku University, and those who provided me all kinds of help in my personal life in Japan. Without their help, this work would not be possible.

First of all, I am deeply grateful to my Academic advisor Professor Hitoshi Oshitani. To me, he is not only my supervisor, but his contribution in my whole career is enormous. Without his active support and continuous encouragement, I could not even reach at door my graduation. He has introduced me into the fascinating research field of virus evolution and its impact in human. I am greatly inspired by his scientific works and contribution in the field of emerging infectious diseases. He is my guide, and I always follow his philosophy of doing science with a flexible and open mind. He will remain my inspiration throughout my scientific career. I appreciate everything that he has taught me. Finally, just simply a few words are just tip of an ice burg compared to his total contribution in my study and personal life as well.

Next, I would like to thank Dr. Yasuhiro Suzuki who has introduced me in the field of HIV-1 molecular biology. He is my first teacher in Tohoku University,

and I still follow many of his research thinking process. He taught me many basic and sophisticated laboratory techniques for doing researches on HIV-1 Tat. We had a lot of research arguments, sweet and bitter interactions regarding science and global research culture etc. I will remember those learnings throughout my whole life. Next, I would like to thank Professor Toshio Hattori for his continuous inspiration for completing my PhD study. I could not even work with him directly, even though he always pretty kind on me. He is the first person to whom I contacted to study in Tohoku University near about 10 years ago, and he introduced me with others.

I would also like to thank Dr. Shizu Hidema, Assistant Professor of Department of Molecular and Cell Biology, Tohoku University Graduate School of Agriculture. She taught me the all the techniques of protein science including making of HIV-1 Tat protein and its application in rodent models etc. She always extended her helpful hands whenever I faced any troubles during my experiments. I would also like to thank Dr. Eichii N. Kodama for his valuable advices in my journey of graduation. I would also like to thank Drs. Mayuko Saito without discussing with her, I cannot conceptualize the idea of writing my thesis. I would also like to acknowledge the contribution of Dr. Yuki Furuse for his

contribution in manuscript writing and teaching me to prepare the figures according to the journal's instructions, and also to Drs Akira Suzuki and Michiko Okamoto for their continuous friendly support to complete the study. I would also like thanks Mrs. Karen Lewis for her valuable advices for preparing my manuscripts. I learned many useful tips of scientific writing from her. I also thank to our all Office staff especially Ms. Tokiko Nishina for her constant support professional support for preparing all the official documents and PhD thesis. I also thank the all the other past and current members of the Department of Virology for all the scientific discussions, exchange of ideas in journal clubs and progress report are place where I generated my ideas of doing this study.

I am also thankful to Japan Foundation for AIDS Research for their financial support for my first three years and, then, Japan Initiative for Global Research Network on Infectious Diseases (J-GRID) and SATREPS (Science and Technology Research Partnership for Sustainable Development) for providing me financial support.

Finally, I would like express my gratitude to my best friend and beloved wife Dr. Irona Khandaker and my little daughter Chandrima Irina Roy for their continuous indubitable support otherwise it will impossible for me to complete my

dream. At last, I would like to dedicate my PhD thesis to my father Late Professor Chaitanya Chandra Roy and my mother Mrs. Chandana Roy for their enormous support in my life and career.

References

1. International Committee on Taxonomy of Viruses (2002). “61.0.6. Lentivirus". National Institutes of Health.
2. International Committee on Taxonomy of Viruses (2002). “61. Retroviridae". National Institutes of Health. Retrieved February 28, 2006.
3. Eric OF, Martin MA. Human Immunodeficiency Viruses: Replication. In: Knipe D, Howley P, eds. Fields Virology. 6th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013:1532–1533.
4. Reeves J, Doms R (2002). Human Immunodeficiency Virus Type 2. Journal of General Virology 83 (Pt 6): 1253–65.
5. HIV/AIDS Fact sheet N°360. November 2014. Accessed January 10, 2014.
6. Hemelaar J, Gouws E, Ghys PD, Osmanov S, WHO-UNAIDS Network for HIV Isolation and Characterisation (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. 25:679-689.

7. Centers for Disease Control and Prevention (CDC) (2001). First report of AIDS.MMWR Morb Mortal Wkly Rep. 50(21):429.

8. Hoth DF, Bolognesi DP, Corey L, Vermund SH (1994) NIH conference. HIV vaccine development: a progress report. Ann Intern Med 121: 603-611.

9. Haynes BF (1993). Scientific and social issues of human immunodeficiency virus vaccine development. Science 260: 1279-1286.

10. Dougherty JP, Temin HM (1988). Determination of the rate of base-pair substitution and insertion mutations in retrovirus replication. J Virol 62: 2817-2822.

11. Ho DD, Neumann AU, Perelson AS, Chen W, et al. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 373: 123-126.

12. Jetzt AE, Yu H, Klarmann GJ, Ron Y, et al. (2000). High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 74: 1234-1240.

13. Alizon S, Fraser C (2013). Within-host and between-host evolutionary rates across the HIV-1 genome. Retrovirology 10: 49.

14. Hutto C, Zhou Y, He J, Geffin R, et al. (1996). Longitudinal studies of viral

sequence, viral phenotype, and immunologic parameters of human immunodeficiency virus type 1 infection in perinatally infected twins with discordant disease courses. Journal of Virology 70: 3589-3598.

15. Jung M, Leye N, Vidal N, Fargette D, et al. (2012). The origin and evolutionary history of HIV-1 subtype C in Senegal. PLoS One 7: e33579.

16. Korber B, Muldoon M, Theiler J, Gao F, et al. (2000). Timing the ancestor of the HIV-1 pandemic strains. Science 288: 1789-1796.

17. Bello G, Eyer-Silva WA, Couto-Fernandez JC, Guimaraes ML, et al. (2007). Demographic history of HIV-1 subtypes B and F in Brazil. Infect Genet Evol 7: 263-270.

18. Walker PR, Pybus OG, Rambaut A, Holmes EC (2005). Comparative population dynamics of HIV-1 subtypes B and C: subtype-specific differences in patterns of epidemic growth. Infect Genet Evol 5: 199-208.

19. Kim GJ, Yun MR, Koo MJ, Shin BG, et al. (2012). Estimating the origin and evolution characteristics for Korean HIV type 1 subtype B using Bayesian phylogenetic analysis. AIDS Res Hum Retroviruses 28: 880-884.

20. Jeang KT, Xiao H, Rich EA (1999). Multifaceted activities of the HIV-1 transactivator of transcription, Tat. J Biol Chem 274:28837-28840.

21. Das AT, Harwig A, Berkhout B (2011). The HIV-1 Tat protein has a versatile role in activating viral transcription. J Virol 85(18):9506-16.
22. King JE, Eugenin EA, Buckner CM, Berman JW (2006). HIV tat and neurotoxicity. Microbes Infect 8(5):1347-57.
23. Li L, Dahiya S, Kortagere S, Aiamkitsumrit B, et al. (2012). Impact of tat genetic variation on HIV-1 disease. Adv Virol 2012:123605.
24. Rana TM, Jeang KT (1999). Biochemical and functional interactions between HIV-1 Tat protein and TAR RNA. Arch Biochem Biophys 365:175-185.
25. Truant R, Cullen BR (1999). The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin beta-dependent nuclear localization signals. Mol Cell Biol 19:1210-1217.
26. de la Fuente JM, Berry CC (2005). Tat peptide as an efficient molecule to translocate gold nanoparticles into the cell nucleus. Bioconjug Chem 16:1176-1180.
27. Ziegler A, Seelig J (2004). Interaction of the protein transduction domain of HIV-1 Tat with heparan sulfate: Binding mechanism and thermodynamic parameters. Biophys J 86:254-263.
28. Roy S, Delling U, Chen CH, Rosen CA, et al. (1990). A bulge structure in

HIV-1 TAR RNA is required for Tat binding and Tat-mediated trans-activation. Genes Dev 4:1365-1373.

29. Fawell S, Seery J, Daikh Y, Moore C, et al. (1994). Tat-mediated delivery of heterologous proteins into cells. Proc Natl Acad Sci U S A 91:664-668.

30. Hidema S, Tonomura Y, Date S, Nishimori K (2012). Effects of protein transduction with intact myogenic transcription factors tagged with HIV-1 Tat-PTD (T-PTD) on myogenic differentiation of mouse primary cells. J Biosci Bioeng 113:5-11.

31. Suzuki Yasuhiro, Roy Chandra Nath, Promjunyakul W, Hatakeyama H et al.(2013). Single quantum dot tracking reveals that an individual multivalent HIV-1 Tat protein transduction domain can activate machinery for lateral transport and endocytosis. Mol Cell Biol 33(15):3036-49.

32. Imamura J, Suzuki Y, Gonda K, Roy CN et al. (2011). Single particle tracking confirms that multivalent Tat protein transduction domain-induced heparan sulfate proteoglycan cross-linkage activates Rac1 for internalization. J Biol Chem 25;286(12):10581-92.

33. Mason RD, De Rose R, Kent SJ (2009). Differential patterns of immune escape at Tat-specific cytotoxic T cell epitopes in pigtail macaques. Virology

388:315-323.

34. Smith SM, Pentlicky S, Klase Z, Singh M, et al. (2003). An in vivo replication-important function in the second coding exon of tat is constrained against mutation despite cytotoxic t lymphocyte selection. J Biol Chem 278:44816-44825.

35. Brake DA, Debouck C, Biesecker G (1990). Identification of an arg-gly-asp (rgd) cell adhesion site in human immunodeficiency virus type 1 transactivation protein, Tat. J Cell Biol 111:1275-1281.

36. Orsini MJ, Debouck CM, Webb CL, Lysko PG (1996). Extracellular human immunodeficiency virus type 1 Tat protein promotes aggregation and adhesion of cerebellar neurons. J Neurosci 16:2546-2552.

37. Barillari G, Gendelman R, Gallo RC, Ensoli B (1993). The Tat protein of human immunodeficiency virus type 1, a growth factor for AIDS Kaposi sarcoma and cytokine-activated vascular cells, induces adhesion of the same cell types by using integrin receptors recognizing the RGD amino acid sequence. *Proc Natl Acad Sci U S A* 1;90(17):7941-5.

38. A. Pugliese, V. Vidotto, T. Beltramo, S. Petrini et al. (2005). A review of HIV-1 Tat protein biological effects Cell Biochem Funct. 2005; 23: 223–227.

39. Ensoli B, Fiorelli V, Ensoli F, Cafaro A, et al. (2006). Candidate HIV-1 Tat vaccine development: from basic science to clinical trials. AIDS 20(18):2245-61.

40. Goldstein, G. HIV-1 Tat protein as a potential AIDS vaccine (1996). Nature Med. 1, 960−964.

41. Marinaro M, Riccomi A, Rappuoli R, Pizza M, et al. (2003). Mucosal delivery of the human immunodeficiency virus-1 Tat protein in mice elicits systemic neutralizing antibodies, cytotoxic T lymphocytes and mucosal IgA. Vaccine 21:3972–3981.

42. Caputo A, Gavioli R, Altavilla G, Brocca-Cofano E, et al. (2003). Immunization with low doses of HIV-1 tat DNA delivered by novel cationic block copolymers induces CTL responses against Tat. Vaccine 21:1103–1111.

43. Caselli E, Betti M, Grossi MP, Balboni PG, et al. (1999). DNA immunization with HIV-1 tat mutated in the transactivation domain induces humoral and cellular immune responses against wild-type Tat. J Immunol 162: 5631–5638.

44. Cui Z, Patel J, Tuzova M, Ray P, et al. (2004).Strong T cell type-1 immune responses to HIV-1 Tat (1-72) protein-coated nanoparticles. Vaccine

22:2631–2640.

45. Bellino S, Tripiciano A, Picconi O, Francavilla V, et al. (2014). The presence of anti-Tat antibodies in HIV-infected individuals is associated with containment of CD4+ T-cell decay and viral load, and with delay of disease progression: results of a 3-year cohort study. Retrovirology. 24;11:49.
46. Goldstein G, Chicca JJ (2012). Exploratory clinical studies of a synthetic HIV-1 Tat epitope vaccine in asymptomatic treatment-naïve and antiretroviral-controlled HIV-1 infected subjects plus healthy uninfected subjects. Hum Vaccin Immunother 8:479-485.
47. Goldstein G, Damiano E, Donikyan M, Pasha M, et al. (2012). HIV-1 Tat B-cell epitope vaccination was ineffectual in preventing viral rebound after ART cessation: HIV rebound with current ART appears to be due to infection with new endogenous founder virus and not to resurgence of pre-existing Tat-dependent viremia. Hum Vaccin Immunother 8(10):1425-30.
48. Ruckwardt TJ, Tikhonov I, Berg S, Hatfield GS, et al. (2004). Sequence variation within the dominant amino terminus epitope affects antibody binding and neutralization of human immunodeficiency virus type 1 Tat protein. J Virol 78:13190-13196.

49. Kandathil AJ, Kannangai R, Abraham OC, Pulimood SA, et al. (2009). Amino acid sequence divergence of Tat protein (exon1)of subtype B and C HIV-1 strains: Does it have implications for vaccine development? Bioinformation 4:237-241.

50. Davidson A, Leeper T, Athanassiou Z, Patora-Komisarska K, et al. (2009) Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. Proc. Natl. Acad. Sci. U. S. A 106:11931–11936.

51. Richter S.N., Palu G (2006). Inhibitors of hiv-1 tat-mediated transactivation. Curr. Med. Chem 13: 1305-1315.

52. Hamy F, Felder ER, Heizmann G, Lazdins J, et al. (1997). An inhibitor of the tat/tar rna interaction that effectively suppresses hiv-1 replication. Proc. Natl. Acad. Sci. USA 94: 3548-3553.

53. Bai J, Sui J, Zhu RY, Tallarico AS, et al. (2003). Inhibition of tat-mediated transactivation and hiv-1 replication by human anti-hcyclint1 intrabodies. J. Biol.Chem 278, 1433-1442.

54. Sung T.L., Rice A.P (2009) mIR-198 inhibits hiv-1 gene expression and replication in monocytes and its mechanism of action appears to involve repression of cyclin t1. PLoS Pathog. 5: e1000263.

55. Ronsard L, Lata S, Singh J, Ramachandran VG, et al. (2014).Molecular and genetic characterization of natural HIV-1 Tat Exon-1 variants from North India and their functional implications. PLoS One. 23; 9(1):e85452.

56. Kandathil AJ, Kannangai R, Abraham OC, Pulimood SA, et al. (2009). Amino acid sequence divergence of Tat protein (exon1) of subtype B and C HIV-1 strains: Does it have implications for vaccine development? Bioinformation. 4:237-241.

57. Siddappa NB, Venkatramanan M, Venkatesh P, Janki MV, et al. (2006). Transactivation and signaling functions of Tat are not correlated: biological and immunological characterization of HIV-1 subtype-C Tat protein. *Retrovirology*18; 3:53.

58. Kurosu T, Mukai T, Komoto S, Ibrahim MS, et al. (2002). Human immunodeficiency virus type 1 subtype C exhibits higher transactivation activity of Tat than subtypes B and E. Microbiol Immunol 46(11):787-99.

59. Marasco WA, LaVecchio J, Winkler A (1999). Human anti-HIV-1 tat sFv intrabodies for gene therapy of advanced HIV-1-infection and AIDS. *J Immunol Methods* 231(1-2):223-238.

60. Los Alamos National Laboratory (LANL) HIV sequence database http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html. Accessed July 3, 2014.

61. http://www.genome.jp/tools/clustalw/ Acccessed February 8, 2014

62. http://weblogo.berkeley.edu/logo.cgi;2008_ Acccessed February 8, 2014

63. Nei M, Kumar S (2000). *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.

64. Tamura K, Stecher G, Peterson D, Filipski A, et al. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725-2729.

65. Rambaut A (2014). Molecular evolution, phylogenetics and epidemiology. http://tree.bio.ed.ac.uk. Accessed August 28, 2014.

66. Pond SL, Frost SD: Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 2005;21:2531-2533.

67. Kosakovsky Pond SL, Frost SD (2005): Not so different after all: A comparison of methods for detecting amino acid sites under selection. Mol Biol Evol 22:1208-1222.

68. Delport W, Poon AF, Frost SD, Kosakovsky Pond SL (2010): Datamonkey

2010: A suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics 26:2455-2457.

69. Poon AF, Lewis FI, Frost SD, Kosakovsky Pond SL (2008): Spidermonkey: Rapid detection of co-evolving sites using bayesian graphical models. Bioinformatics 24:1949-1950.

70. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing.Nat Methods. 2012 Jul 30;9(8):772. doi: 10.1038/nmeth.2109.

71. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969-1973.

72. Drummond AJ, Ho SY, Phillips MJ, Rambaut A (2006). Relaxed phylogenetics and dating with confidence. PLoS Biol 4: e88.

73. Minin VN, Bloomquist EW, Suchard MA (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol 25: 1459-1471.

74. Rambaut A (2007). Molecular evolution, phylogenetics and epidemiology. UK.

75. El-Sayed A, Futaki S, Harashima H (2009). Delivery of macromolecules

using arginine-rich cell-penetrating peptides: Ways to overcome endosomal entrapment. AAPS J 11:13-22.

76. Sood V, Ranjan R, Banerjea AC (2008). Functional analysis of hiv-1 subtypes B and C HIV-1 Tat exons and RGD/QGD motifs with respect to Tat-mediated transactivation and apoptosis. AIDS 22:1683-1685.

77. Kuiken C, Foley B, Leitner T, Cristian Apetrei (2012) et al., eds. HIV Sequence Compendium 2012. NM,USA: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; No. LA-UR 2-24653.

78. Thomson MM, Pérez-Alvarez L, Nájera R. Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy (2002). *Lancet Infect Dis* 2(8):461-471.

79. Janssens W, Buvé A, Nkengasong JN (1997). The puzzle of HIV-1 subtypes in Africa. *AIDS* 11(6):705-712.

80. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, et al. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of Virology* 73(1):152-160.

81. Purohit P, Dupont S, Stevenson M, Green MR (2001). Sequence-specific

interaction between HIV-1 matrix protein and viral genomic RNA revealed by in vitro genetic selection. *RNA* 7(4):576-584.

82. Buonaguro L, Tornesello ML, Buonaguro FM (2007). Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol* 81:10209-10219.

83. Hu DJ, Buvé A, Baggs J, van der Groen G, et al. (1999). What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS.*13:873-881.

84. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The challenge of HIV-1 subtype diversity. *N Engl J Med.* Apr 2008;358:1590-1602.

85. Bozek K, Lengauer T. Positive selection of HIV host factors and the evolution of lentivirus genes. *BMC Evol Biol.* 2010;10:186.

86. Kalantari P, Narayan V, Natarajan SK, Muralidhar K, et al. (2008). Thioredoxin reductase-1 negatively regulates HIV-1 transactivating protein Tat-dependent transcription in human macrophages. *J Biol Chem* 283:33183-33190.

87. Zheng L, Yang Y, Lu G, Salvato M (2005). Extracellular HIV Tat and Tat cysteine rich peptide increase CCR5 expression in monocytes. *J Zhejiang Univ*

Sci B 6(7):668-672.

88. Zheng L, Yang Y, Lu G, Salvato M (2005). Extracellular HIV Tat and Tat cysteine rich peptide increase CCR5 expression in monocytes. *J Zhejiang Univ Sci B* 6(7):668-672.

89. Rossenkhan R, MacLeod IJ, Sebunya TK, Castro-Nallar E, et al. (2013). tat Exon 1 exhibits functional diversity during HIV-1 subtype C primary infection. *J Virol* 87:5732-5745.

90. Mahlknecht U, Dichamp I, Varin A, Van Lint C, et al. (2008). Nf-kappa b-dependent control of HIV-1 transcription by the second coding exon of Tat in T cells. J Leukoc Biol 83:718-727.

91. Campbell GR, Loret EP (2009). What does the structure-function relationship of the HIV-1 Tat protein teach us about developing an aids vaccine? Retrovirology 6:50.

92. Maljkovic Berry I, Ribeiro R, Kothari M, Athreya G, et al. (2007). Unequal evolutionary rates in the human immunodeficiency virus type 1 (HIV-1) pandemic: the evolutionary rate of HIV-1 slows down when the epidemic rate increases. J Virol 81:10625-10635.

93. Robbins KE, Lemey P, Pybus OG, Jaffe HW, et al. (2003) U.S. Human

immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. J Virol 77: 6359-6366.

94. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, et al. (2007). The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci U S A 104: 18566-18570.

95. Deng X, Liu H, Shao Y, Rayner S, Yang R (2008). The epidemic origin and molecular properties of B': a founder strain of the HIV-1 transmission in Asia. AIDS 22: 1851-1858.

96. Grenfell BT, Pybus OG, Gog JR, Wood JL, et al. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327-332.

97. Gog JR, Grenfell BT (2002). Dynamics and selection of many-strain pathogens. Proc Natl Acad Sci U S A 99: 17209-17214.

98. Heyndrickx L, Janssens W, Zekeng L, Musonda R, et al. (2000). Simplified strategy for detection of recombinant human immunodeficiency virus type 1 group M isolates by gag/env heteroduplex mobility assay. Study Group on Heterogeneity of HIV Epidemics in African Cities. J Virol 74: 363-370.

The results of this PhD research work have been published in three following papers:

1. Chandra Nath Roy, Irona Khandaker, and Hitoshi Oshitani (2015) **Intersubtype genetic variability of HIV-1 Tat.** *AIDS Research and Human Retroviruses* 31(6): 641-648. (PMID: 25748226. Copyright © 2015, Mary Ann Liebert, Inc.)

2. Chandra Nath Roy, Irona Khandaker, Yuki Furuse and Hitoshi Oshitani (2015) **Molecular characterization of full-length Tat in HIV-1 subtypes B and C**. *Bioinformation* 11(3):151-60. doi: 10.6026/97320630011151. eCollection 2015. (PMCID: PMC4403036. Copyright © 2015 Biomedical Informatics)

3. Chandra Nath Roy, Irona Khandaker, and Hitoshi Oshitani (2015) **Evolutionary dynamics of HIV-1 Tat in subtypes B and C.** *PLoS ONE* 10(6): e0129896. doi:10.1371/journal.pone.0129896 (PMCID: PMC4472691. Copyright: © 2015 Roy et al.)

Figure 1: Functional domains of Tat. Schematic presentation of the domains of Tat exon 1 and 2 were highlighted (panel A): domain I (residues 1 to 21), an acidic/Pro-rich region; domain II (residues 22 to 37), a Cys-rich/Zn^2 Finger domain; domain III (residues 38 to 48), containing conserved Phe (F); domain IV (residues 49 to 57, the basic domain); domain V (residues 58 to 72, a Glu rich domain); and domain VI (residues 3 to 101, encoded by the second exon).

Figure: 1

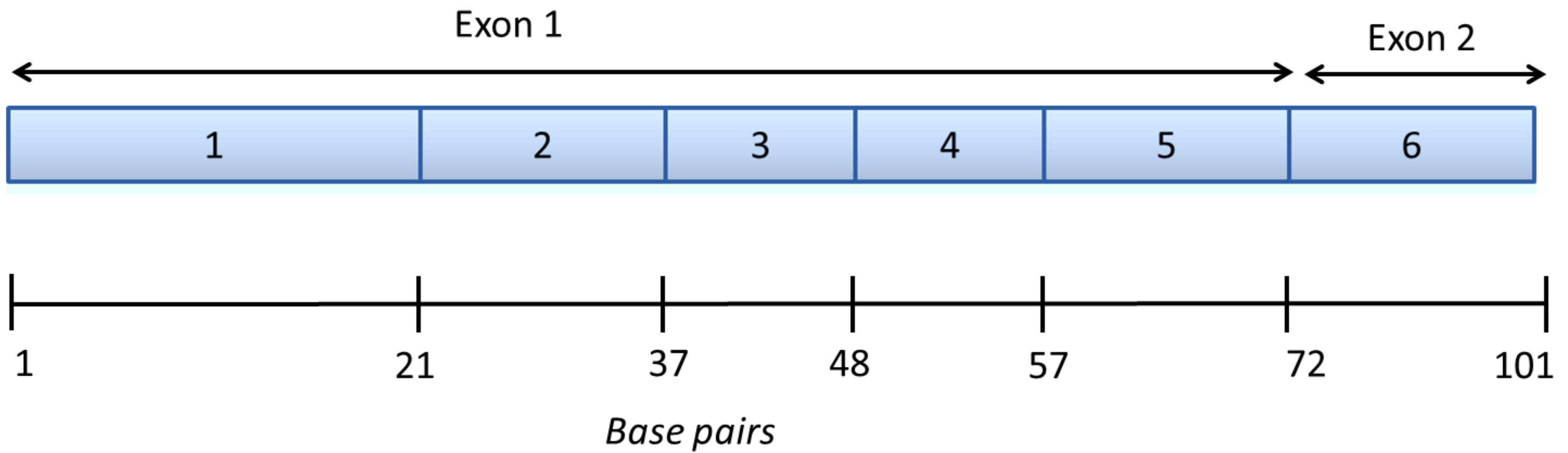
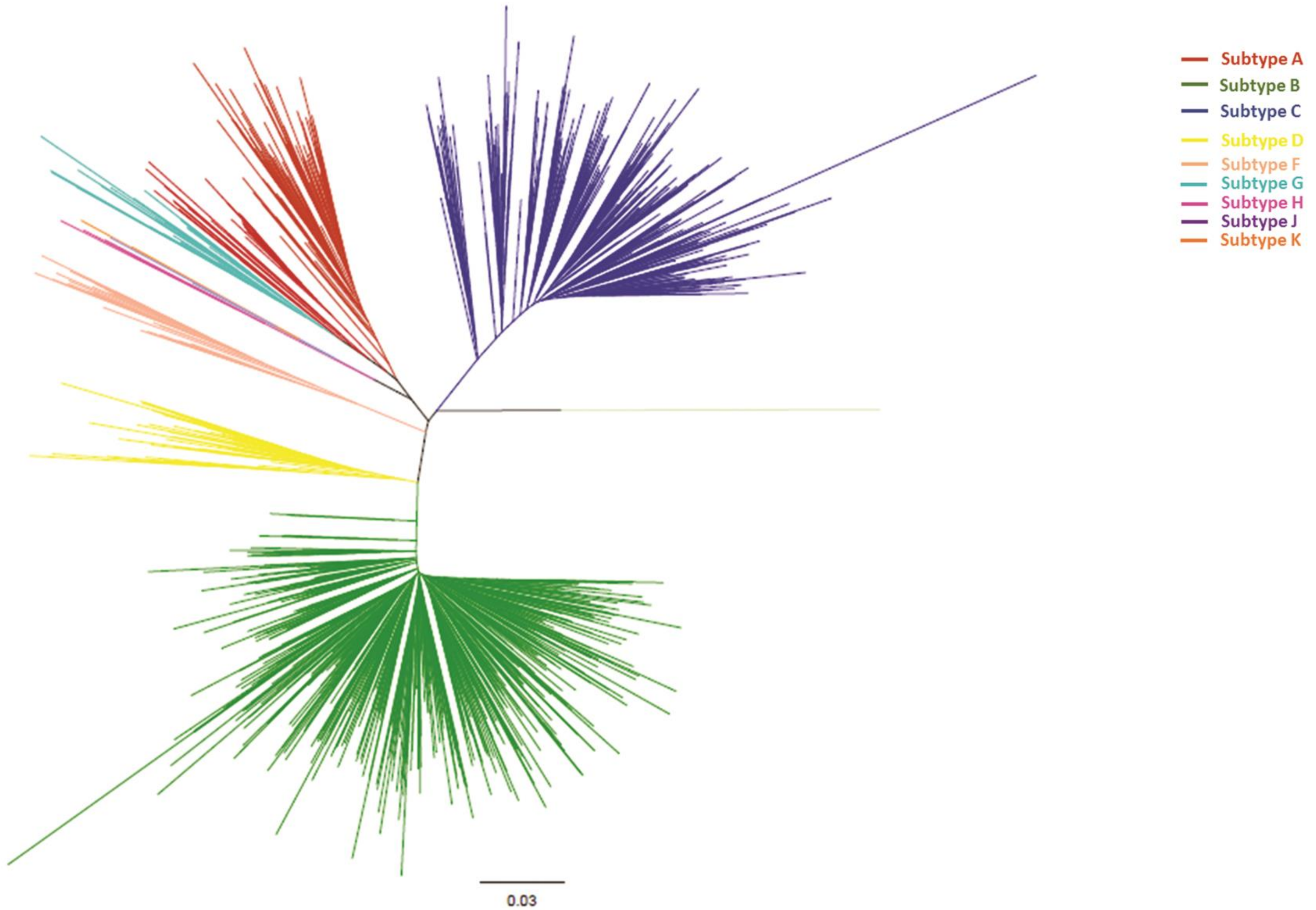


Figure 2: Phylogenetic tree for the HIV-1 Tat exon-1. Figure shows Neighbor-Joining phylogenetic tree of Tat exon 1 sequences in different subtypes of HIV-1. The evolutionary distances were computed using the Tamura-Nei method with GTR +Γ5 nucleotide substitution model in the 1000 bootstrap replicates, using the MEGA 6; the tree was visualized by Fig Tree v1.4.0. Branch colors were added to indicate different subtypes. SIV sequence, CPZ.US.85.US_Marilyn.AF103 was used as the out-group to root the tree (as shown in light green line).

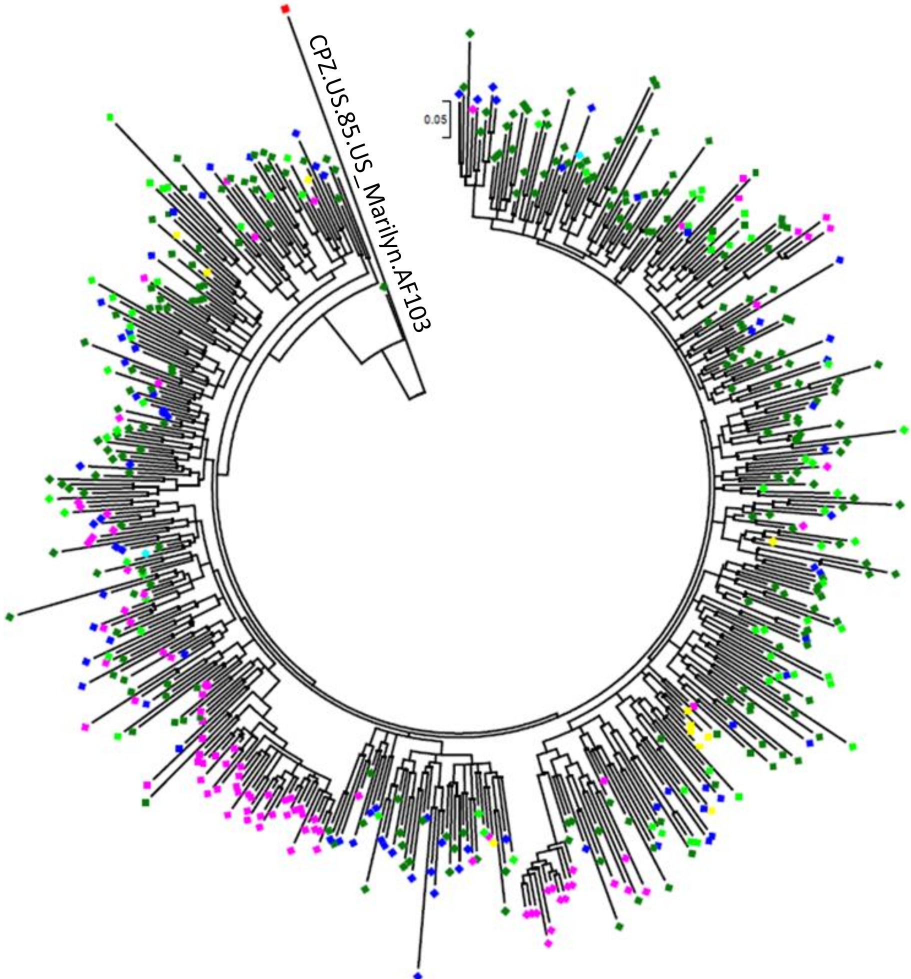
Figure: 2

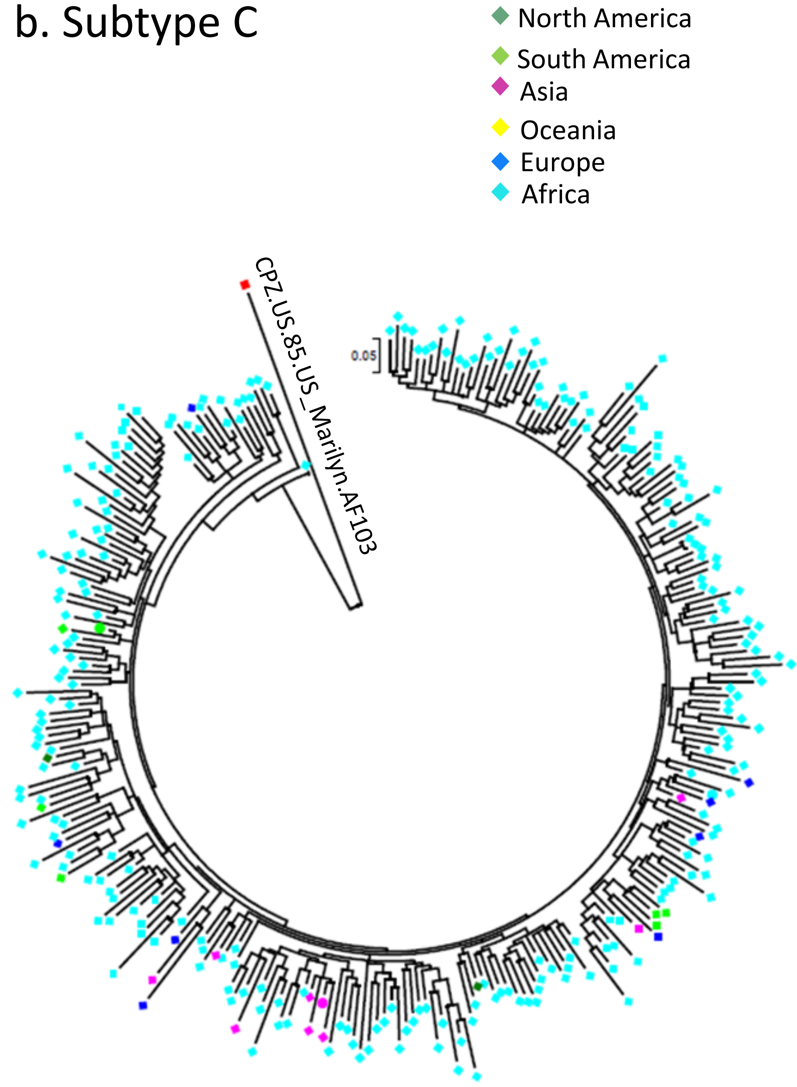


Roy *et al.* *AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Figure 3. Phylogenetic analyses of full-length Tat in HIV-1 subtypes B and C. Maximum likelihood (ML) phylogenetic tree of HIV-1 subtypes B (panel A) and C (panel B) sequences based on 300 nucleotide sites of Tat gene nucleotide sequence. GTR+I+Γ4 nucleotide substitution model was employed with 1000 bootstrapped data sets, using the Molecular Evolutionary Genetics Analysis software ver. 6 (MEGA 6). Reference Tat sequences were also included. SIV sequence (Ref.CPZ.US.85.US_Marilyn.AF103) was used to root the tree showed in red square bullet.

Figure: 3

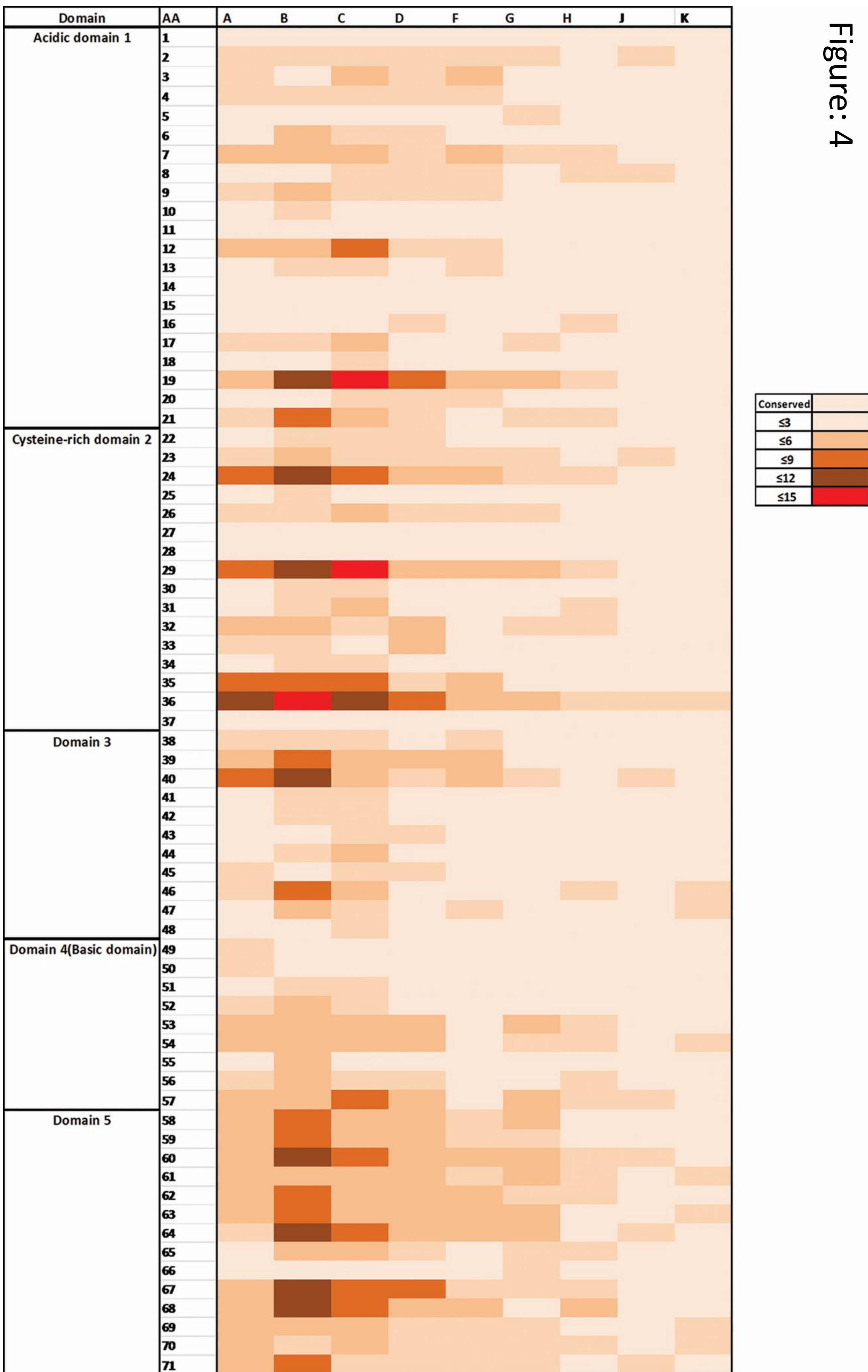




Roy *et al.Bioinformation,* 2015; 11(3):151-60

Figure 4. Amino acid variation in the Tat exon 1 in different subtypes. In the figure, rows indicate amino acid position according to HIV-1 $_{HXB2}$ numbering, and columns indicate different subtypes. The intensity of colors reflects the number of amino acid variations in each position.

Figure: 4

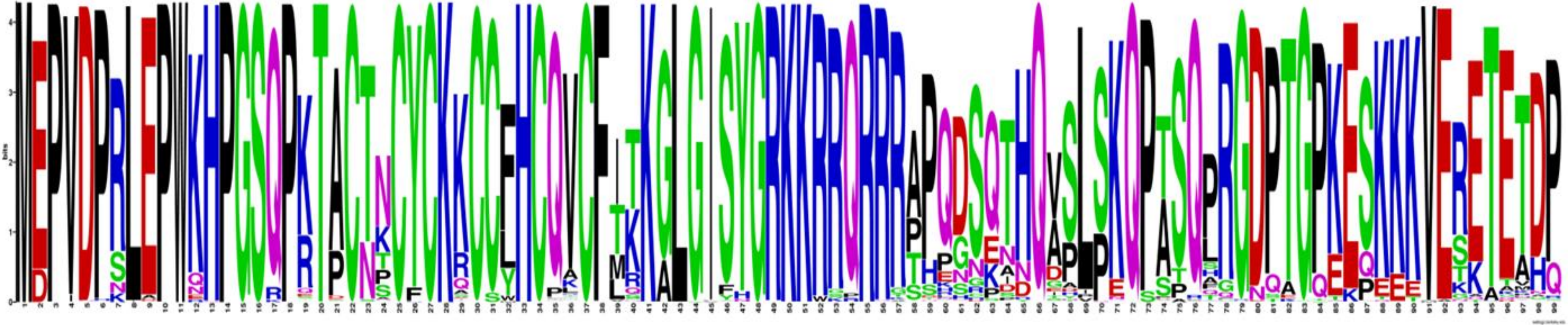


Roy *et al. AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Figure 5. Amino acid variation in Full-length Tat. Sequence logo shows the Full-length Tat amino acid variation observed at positions 1 to 100 in both subtype B (panel a) and subtype C (panel b).

Figure: 5

a



b

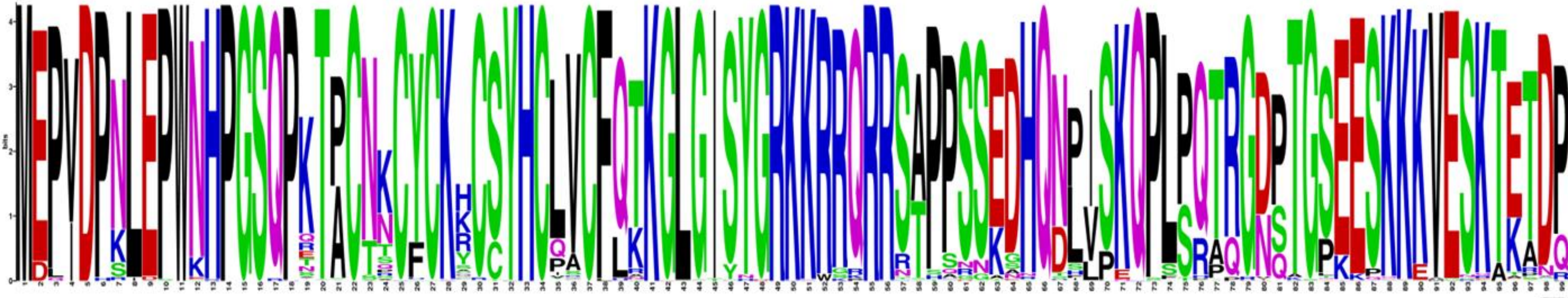
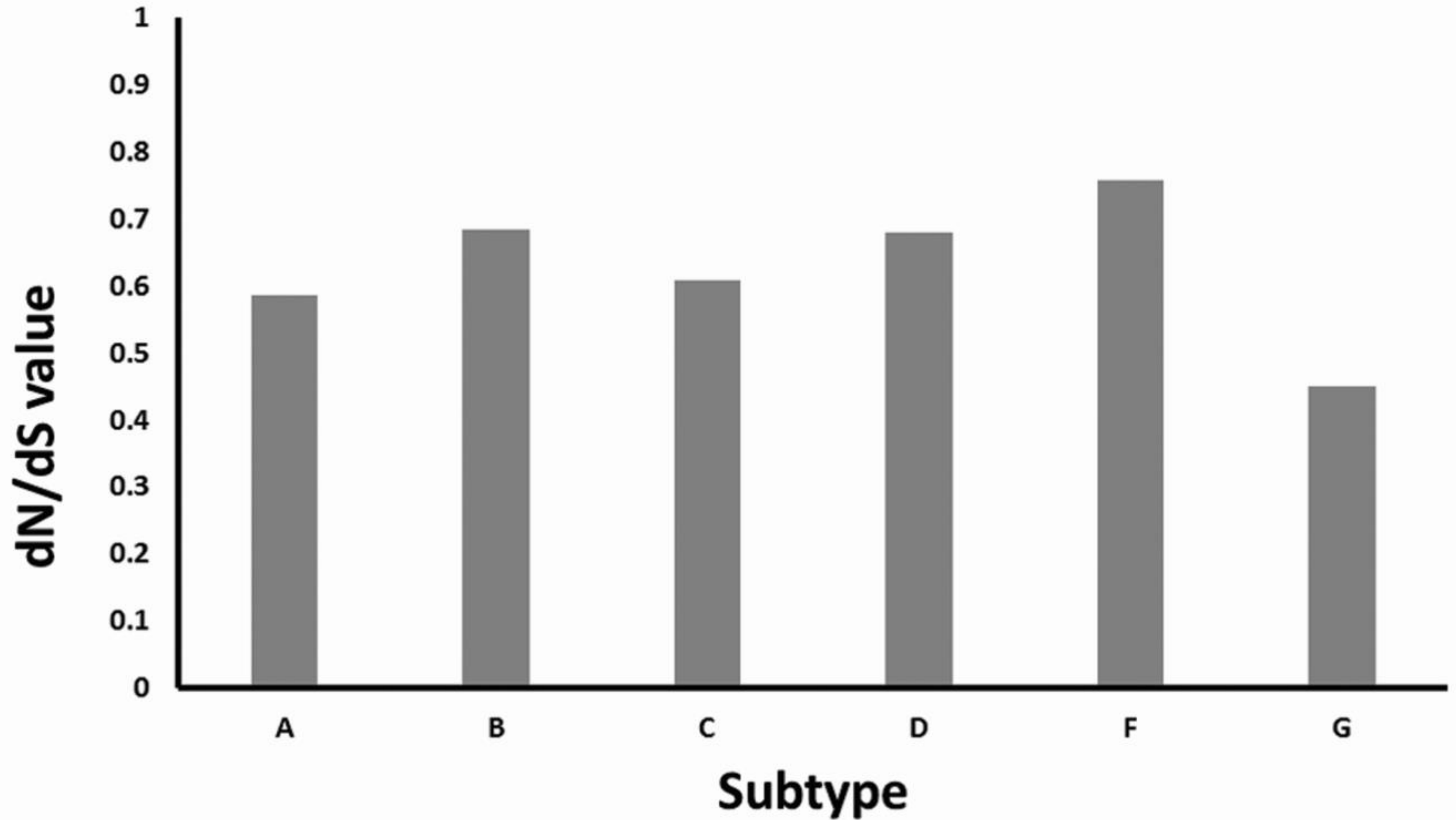


Figure 6. Selection profiles of the Tat exon 1 gene in different subtypes. The dN/dS (ω) values were calculated using the SLAC method. The X-axis shows different subtypes and the Y-axis shows dN/dS (ω) values.

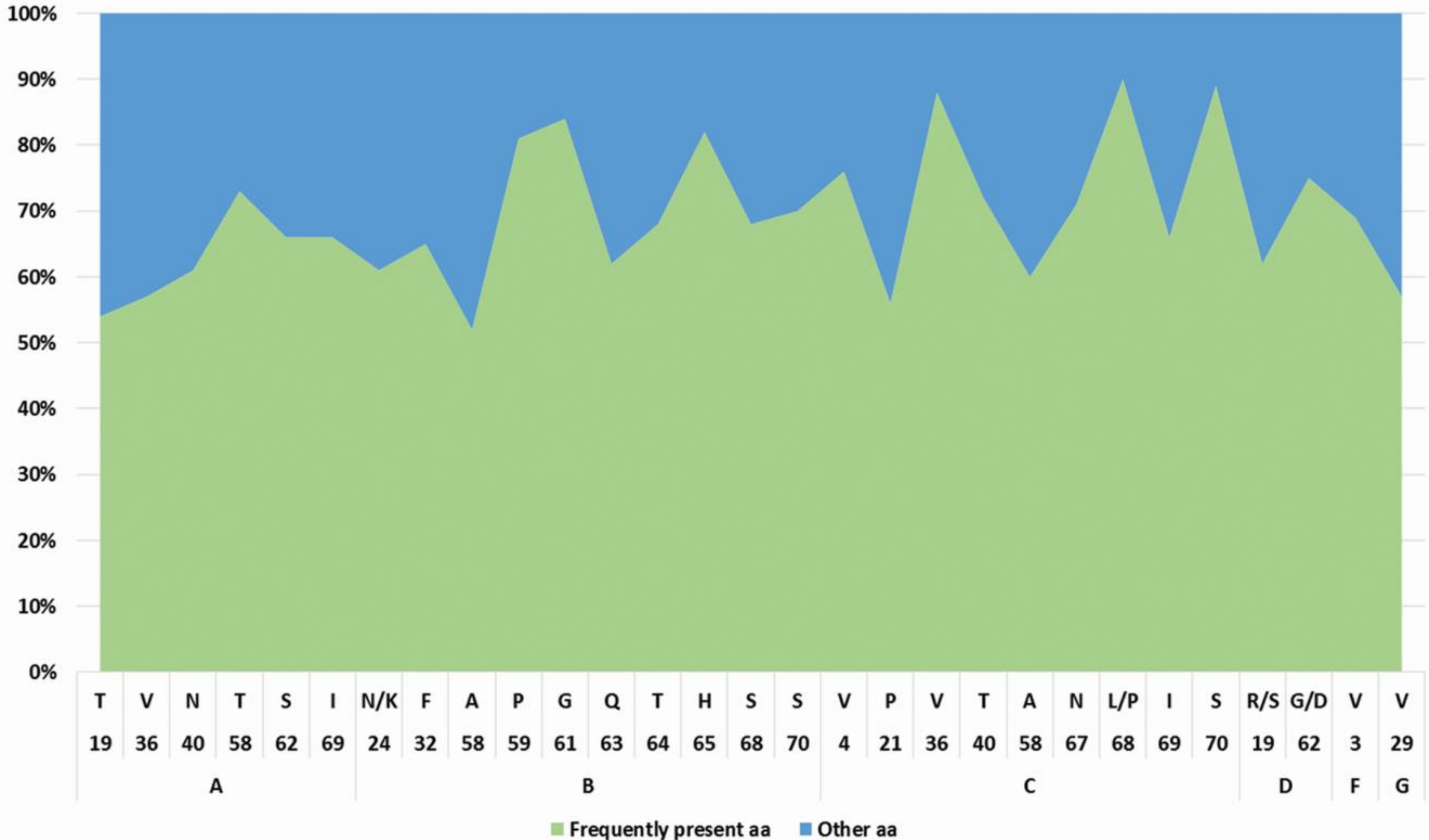
Figure: 6



Roy *et al. AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Figure 7. Amino acid variation in positively selected sites in different subtypes of HIV-1 Tat exon-1. Figure shows the frequency of amino acid variations in each positively selected codon positions determined by SLAC, FEL and iFEL in different subtypes (subtypes A –G).

Figure: 7



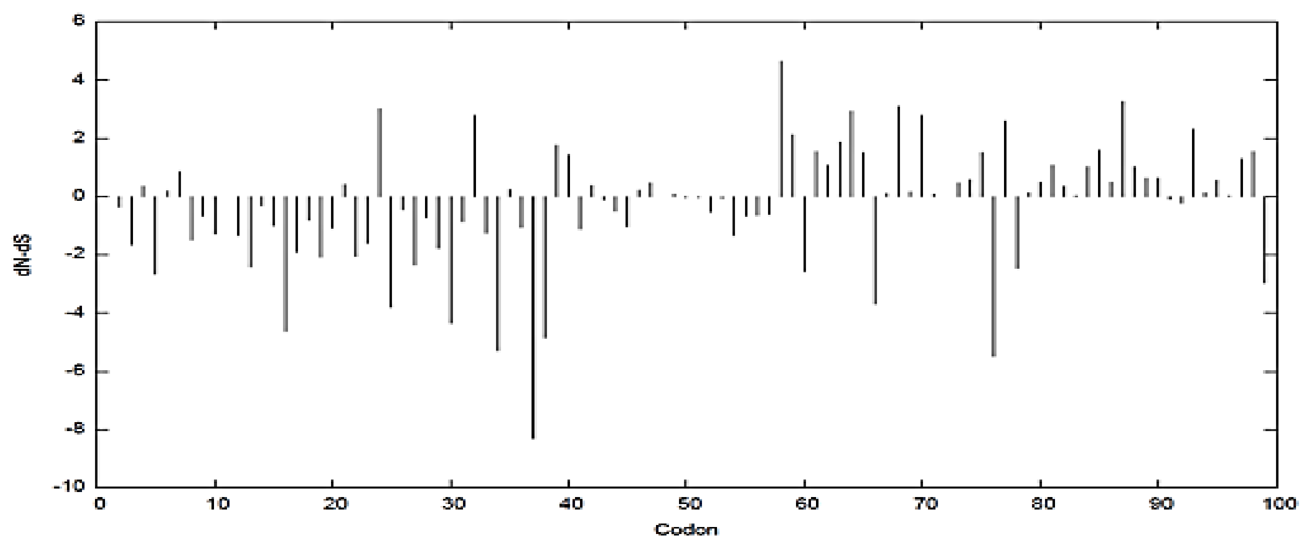
Roy *et al. AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Figure 8. Selection profiles of Tat in subtypes B (panel a) and C (panel b).

dN-dS values were estimated by SLAC and scaled by the total codon tree length. Sites were numbered according to codon Tat position of HIV-1 HXB2 reference strain.

Figure: 8

a



b

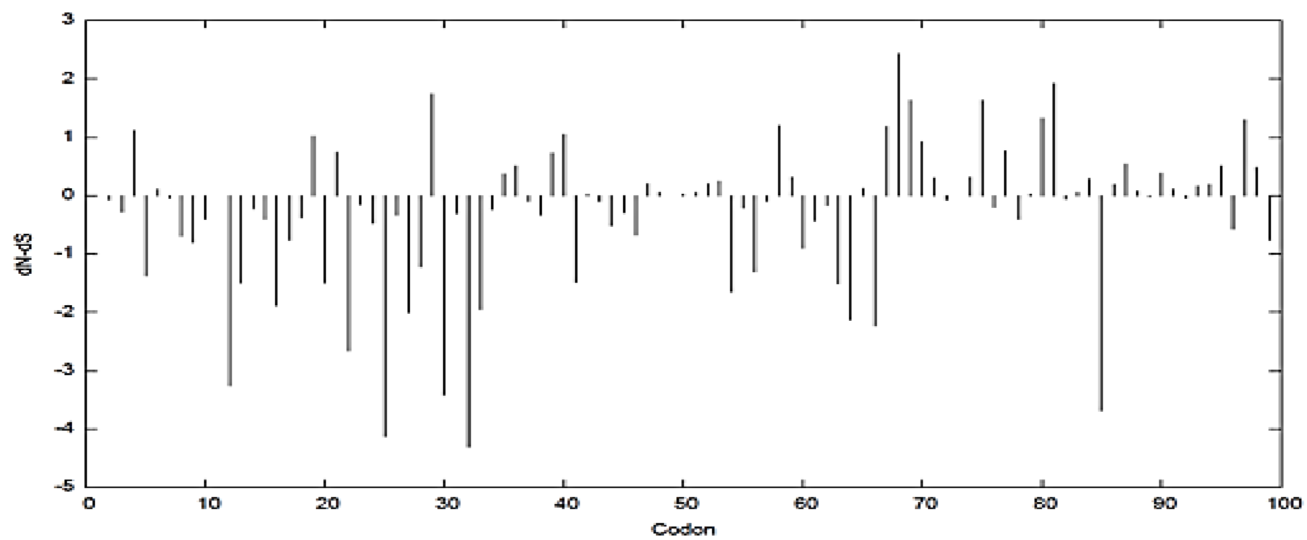
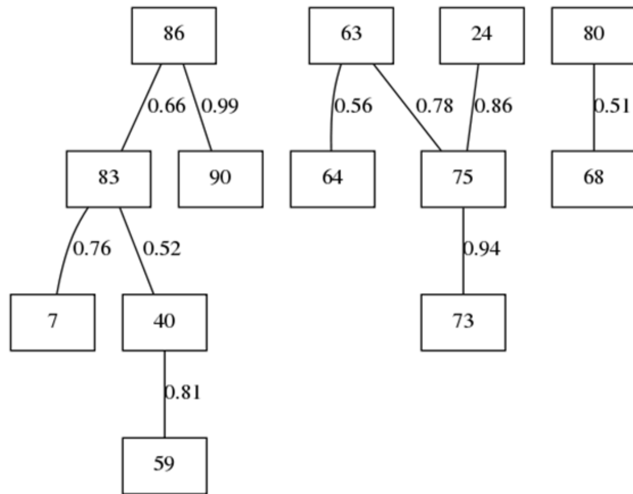


Figure 9.Co-evolving sites identified by Bayesian Graphical Models. Amino acid sites of the Tat gene of subtypes B (panel a) and C (panel b) were depicted as groups showing different sites. Each connection is associated with posterior probabilities for two dependencies. The rectangle contains the HXB2 numbering.

Figure: 9

a



b

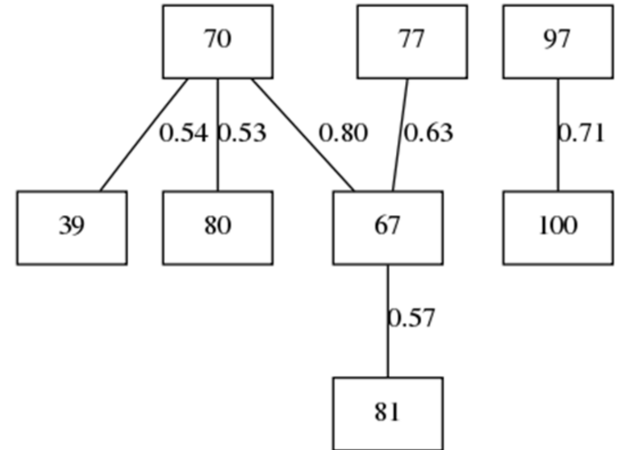


Figure 10: Substitution rates of HIV-1 Tat. The substitution rate was estimated by three molecular based models in both subtypes B and C with normal distribution with a mean of 2.5×10^{-3} and standard deviation 5×10^{-4} prior. Small bars at the top of each column indicate the mean substitution rate per site per year ±SD.

Figure: 10

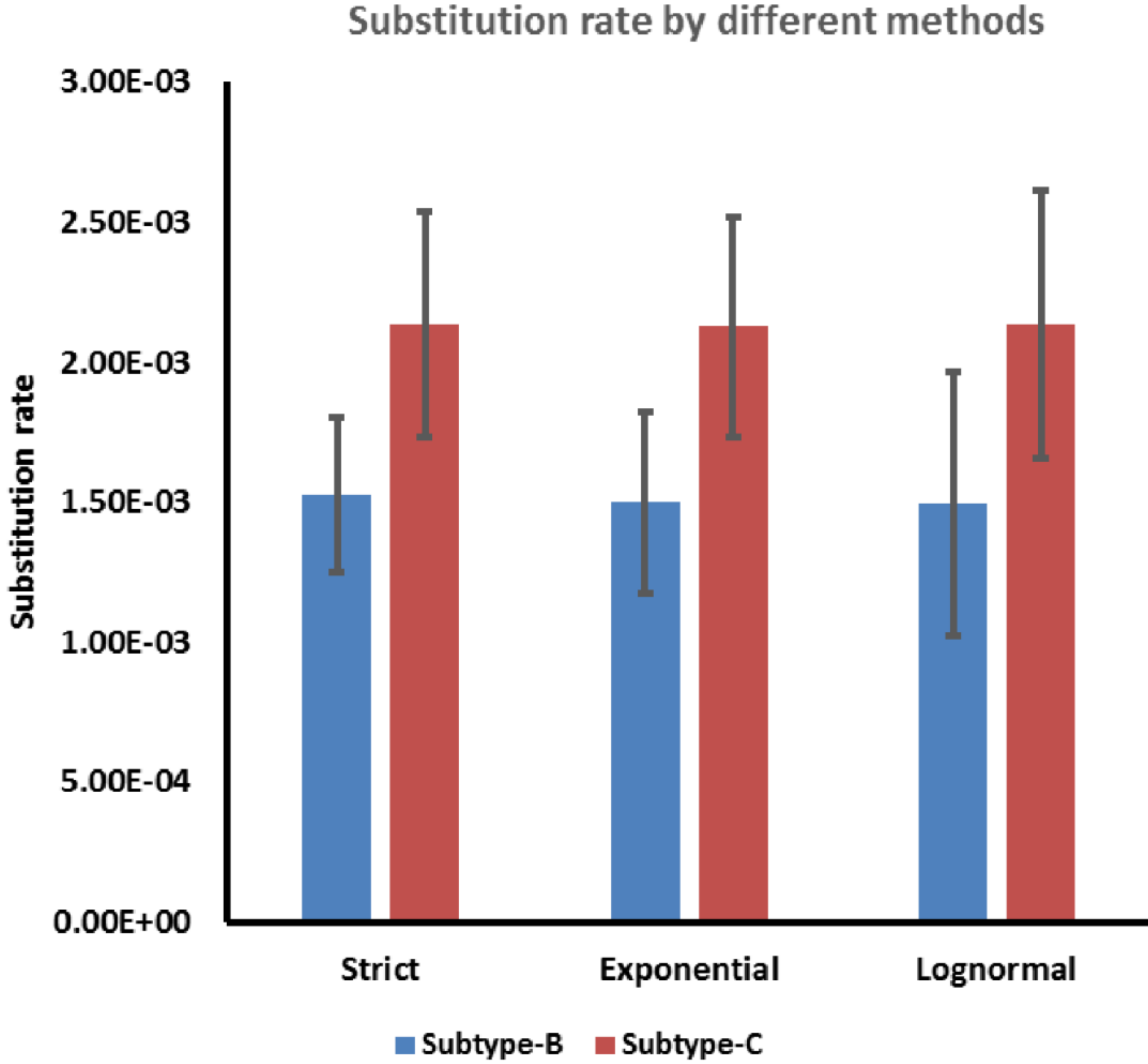
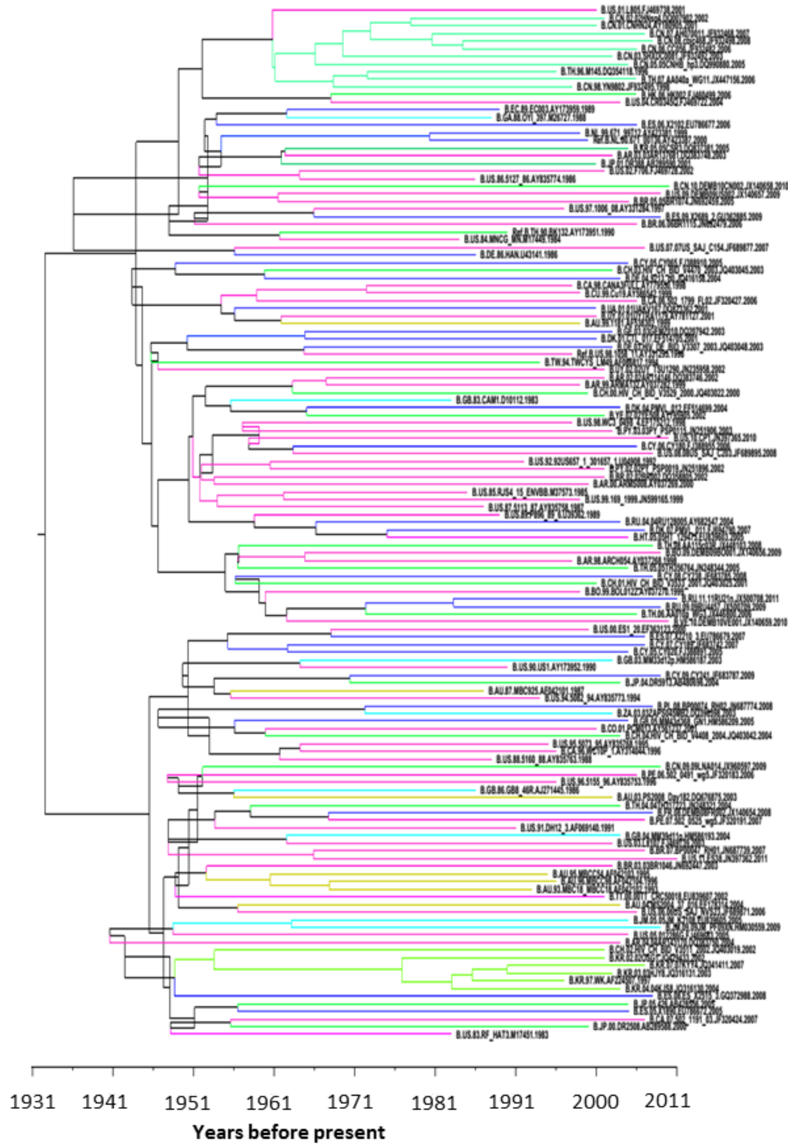
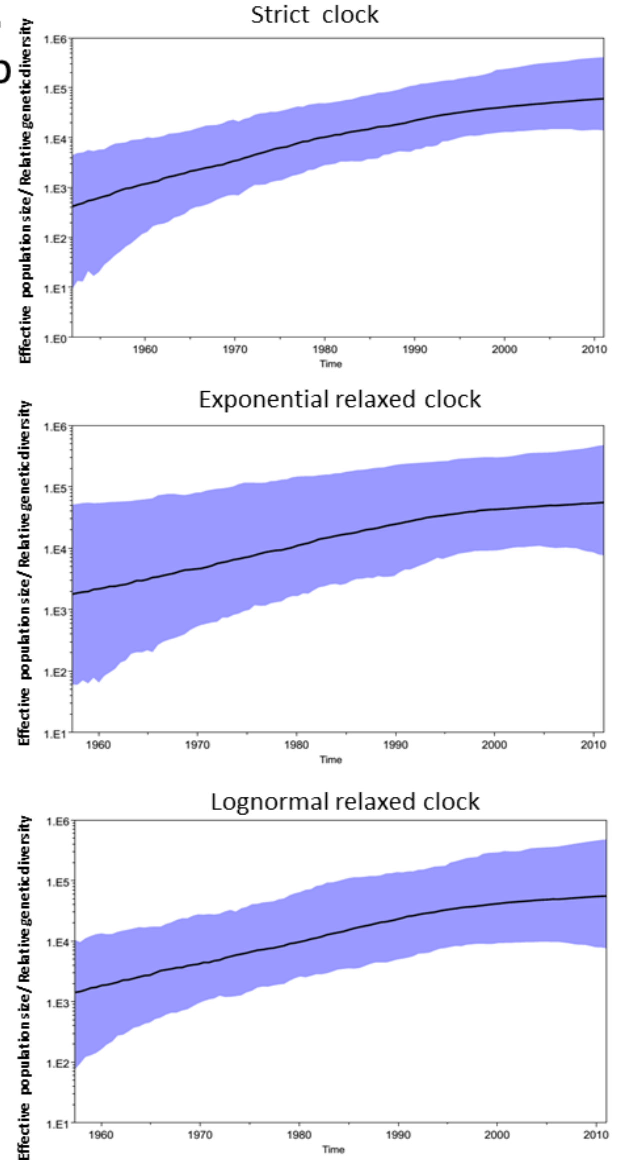


Figure 11. Phylodynamic and Bayesian tree with timescale of HIV-1 subtype B Tat. A) Maximum clade credibility tree with time scale obtained from the strict molecular clock. Time to the MRCA was shown by vertical lines and indicated in years at the bottom of the figure. B) Gaussian Markov random field (GMRF) skyride plot considering three molecular clocks (strict, exponential relaxed and lognormal relaxed). The X-axis represents the time in year. The Y-axis represents the HIV-1 Tat effective number of infections (genetic diversity). The black line marks the median estimate for effective population size and the blue shading showed region displays the 95% highest posterior density (HPD) interval.

Figure: 11

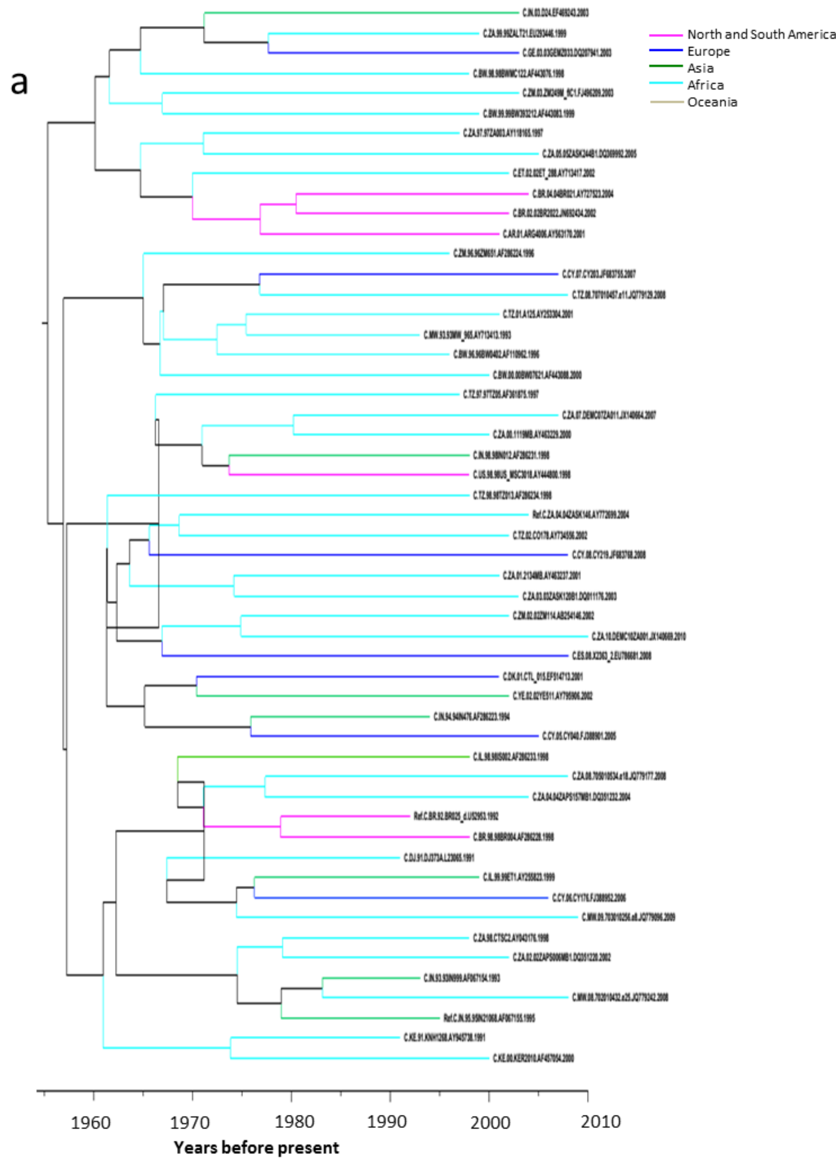


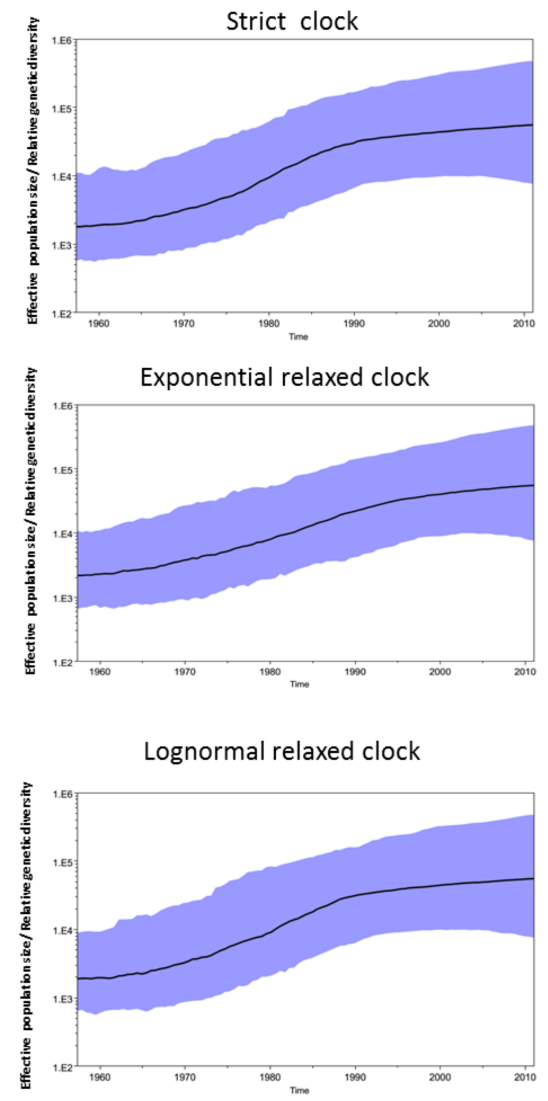


Roy *et al. PLoS ONE* 10(6): e0129896.)

Figure 12. Phylodynamic and Bayesian tree with timescale of HIV-1subtype C Tat. a) Maximum clade credibility tree with time scale obtained from the strict molecular clock. Time to the MRCA was shown by vertical lines and indicated in years at the bottom of the figure. b) Gaussian Markov random field (GMRF) skyride plot considering three molecular clocks (strict, exponential relaxed and lognormal relaxed). The X-axis represents the time in year. The Y-axis represents the HIV-1 Tat effective number of infections (genetic diversity). The black line marks the median estimate for effective population size and the blue shading showed region displays the 95% highest posterior density (HPD) interval.

Figure: 12





Roy *et al. PLoS ONE* 10(6): e0129896.)

Table 1: Genetic variability in different subtypes of HIV-1 Tat exon 1

Variable	Subtype								
	A	B	C	D	F	G	H	J	K
Sequence identity [a] (%)	83	83	85	82	81	86	81	78	80
Mean nucleotide divergence [b] (%)	88	89	90	88	86	89	88	97	97

[a] Sequence identity is calculated with corresponding consensus sequences for each subtype
[b] Mean nucleotide divergence is calculated by the number of base substitutions per site from averaging over all sequence pairs for each subtypes compared to consensus sequences .

Roy *et al.* *AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Table 2: Detail information about the amino acid variation in different subtypes of HIV-1 Tat exon 1

Domain	aa position of HIV-1_{HXB2} strain		Subtype A (N=128)	Subtype B (N=598)	Subtype C (N=336)	Subtype D (N=52)	Subtype F (N=27)	Subtype G (N=30)	Subtype H (N=4)	Subtype J (N=2)	Subtype K (N=2)
Acidic domain 1	1	M	M(128)	M(598)	M(336)	M(52)	M(27)	M(30)	M(4)	M (2)	M(2)
	2	E	D(109), E(19)	D(78), E(520)	D(33), E(303)	D(30), E(22)	D(1), E (26)	D(29), E(1)	D(4)	D(1), E(1)	E(2)
	3	P	L(2), P(126)	P(598)	L(12), P(320), Q(3), S(1)	L(3), P(49)	I(8), L(6), M(2), P(7), V(4)	P(30)	P(4)	P(2)	P(2)
	4	V	I(1),V(127)	A(1), I(14), V(583)	A(1), I(79), V(256)	I(2), V(50)	I(1)V(26)	V(30)	V(4)	V(2)	V(2)
	5	D	D(128)	D(598)	D(336)	D(52)	D(27)	D(29), E(1)	D(4)	D(2)	D(2)
	6	P	P(128)	H(5), P(591), S(1), Y(1)	H(6), P(330)	H(1), L(1), P(50)	P(27)	P(30)	P(4)	P(2)	P(2)
	7	R	D(1), K(6), N(113), S(8)	K(24), N(31), R(466), S(77)	D(2), H(1), K(55), N(249), R(4), S(25)	K(4), N(48)	N(17), S(2), K(5), R(2), D(1)	K(7), N(23)	N(2),K(2)	N(2)	N(2)
	8	L	L(128)	L(598)	I(5), L(331)	I(4), L(48)	I(2), L(25)	L(30)	Q(2),L(2)	I(1),R(1)	I(2)
	9	E	A(1), D(5), E(122)	A(4), D(4), E(588), K(1), Q(1)	A(1), D(5), E(330)	D(2), E(50)	D(25), E(2)	E(30)	E(4)	E(2)	E(2)
	10	P	P(128)	P(597), A(1)	P(336)	P(52)	P(27)	P(30)	P(4)	P(2)	P(2)
	11	W	W(128)	W(598)	W(336)	W(52)	W(27)	W(30)	W(4)	W(2)	W(2)
	12	K	E(1), H(1), K(7), Q(3), S(1),N(115)	E(12), K(525), N(25), Q(28), R(8)	E(4), H(1), K(29), N(299), Q(1), R(1), S(1)	K(5), N(47)	K(2), Q(1),N(24)	N(30)	N(4)	N(2)	N(2)
	13	H	H(128)	H(596), Q(2)	H(333), Q(1), R(2)	H(52)	H(26), Q(1)	H(30)	H(4)	Q(2)	Q(2)
	14	P	P(128)	P(598)	P(336)	P(52)	P(27)	P(30)	P(4)	P(2)	P(2)
	15	G	G(128)	G(598)	G(336)	G(52)	G(27)	G(30)	G(4)	G(2)	G(2)
	16	S	S(128)	S(598)	S(336)	C(2), (S50)	S(27)	S(30)	S(1), N(3)	S(2)	S(2)
	17	Q	K(1), R(4), Q(123)	K(7), Q(569), R(22)	H(1), K(2), R(6), Q(327)	Q(52)	Q(27)	K(2), Q(28)	Q(4)	Q(2)	Q(2)
	18	P	P(128)	P(598)	H(1), P(335)	P(52)	P(27)	P(30)	P(4)	P(2)	P(2)
	19	K	A(20), I(4), K(32), R(2), S(1), T(69)	E(6), G(1),K(400),L(2), N(1), Q(19), R(159), S(7), T(2), V(1)	A(5), D(2), E(18), G(1), H(1), I(4), K(229), L(2), N(12), Q(16),R(19), S(9), T(18)	E(2), G(4),K(11), N(1), Q(2), R(16), S(16)	A(4), E(5), K(1), P(1), S(1), T(15)	I(1), K(19), N(1), Q(1), R(3). T(5)	R(1),Q(1), K(2)	K(2)	K(2)
	20	T	T(128)	T(598)	A(2), I(1), T(333)	S(1), T(51)	N(1), T(26)	T(30)	T(4)	T(2)	T(2)
	21	A	A(75), E(1), P(52)	A(455), D(8), E(2), N(1), P(126), R(1), S(4), T(1)	A(142), E(1), P(189), S(2), T(1), V(1)	A(11), P(41)	P(27)	P(8), A(22)	A(1),P(3)	A(2)	A(2)
Cysteine-rich domain (domain 2)	22	C	C(128)	C(597), S(1)	C(335), G(1)	C(1), R(51)	C(27)	C(30)	C(4)	C(2)	C(2)
	23	T	N(44), S(80), T(4)	N(126), P(1), S(5), T(466)	N(269), P(1),S(8), T(58)	N(45), S(5), T(3)	N(11), T(16)	N(23), T(7)	N(4)	N(1),T(1)	N(2)
	24	N	A(1), K(81), N(33), P(3), R(4), S(2), T(4)	A(15), D(1), G(4), K(105), N(262), P(81), Q(3), R(1), S(62), T(64)	A(5), H(1), K(162), N(92), P(12), Q(13), R(4), S(19), T(28)	K(30), N(9), P(2), R(1), S(9)	K(16), N(5), R(3), S(1), T(2)	N(7), K(23)	N(3), G(1)	Q(2)	Q(2)
	25	C	C(1280	C597, R1	C(336)	C(52)	C(27)	C(30)	C(4)	C(2)	C(2)
	26	Y	F(6),Y(122)	F(26), H(3), Y(569)	C(1), F(43), H(1), M(1), W(1),Y(289)	H(4), Y(48)	F(2), Y(25)	Y(29), F(1)	Y(4)	Y(2)	Y(2)
	27	C	C(128)	C(598)	C(336)	C(52)	C(27)	C(30)	C(4)	C(2)	C(2)
	28	K	K(128)	K(598)	K(336)	K(52)	K(27)	K(30)	K(4)	K(2)	K(2)
	29	K	A(7), E(2), K(91), Y(1), V(9), T(1), R(13),Q(2), M(2),	A(13), E(3), G(4), I(4), K(439), M(3), N(1), Q(41), R(75), S(7), V(7), Y(1)	A(10), C(9), F(4), G(3), H(102), I(3), K(68), L(5), M(1), N(1), Q(6), R(64), S(14), V(2), Y(44)	I(1), K(38), N(1),Q(5),R(5),W(1),	A(2), K(11),R(8), Q(1), V(3), W(2)	A(3), I(2), K(6), N(1), R(1), V(17)	K(1), Q(2), N(1)	R(2)	R(2)

	30	C	C(128)	C(596), G(1), R(1)	C(332), R(3), Y(1)	C(52)	C(27)	C(30)	C(4)	C(2)	C(2)
	31	C	C(128)	C(592), S(6)	C(51), G(52), S(281), T(2)	C(52)	C(27)	C(30)	C(1), S(3)	C(2)	C(2)
	32	F	F(8), L(1), W(30), Y(89)	F(390), L(102), M(8), W(13), Y(85)	F(4), Y(332)	F(1), L(1), W(2), Y(48)	F(27)	W(28), Y(2)	Y(1),F(3)	Y(2)	Y(2)
	33	H	H(127), Y(1)	H(597), P(1)	H(336)	H(52)	H(27)	H(30)	H(4)	H(2)	H(2)
	34	C	C(128)	C(596), G(1), Y(1)	C(335), R(1)	C(52)	C(27)	C(30)	C(4)	C(2)	C(2)
	35	Q	Y(3), V(2), T(1), Q(76), P(31), L(10), I(5)	A(1), I(4),M(1), P(12), Q(578), V(1), Y(1)	A(1), F(1), I(6), L(267), P(22), Q(32), R(1), T(2), V(4)	L(1), Y(3), Q(48)	C(1), H(1), P(1), Q(8), Y(16)	Q(30)	Q(4)	P(2)	Q(2)
	36	V	A(15), D(2), H(1), I(9),K(2), L(10), M(4),R(1), S(7), V(73), Y(4)	A(23), F(2), H(2), I(2), K(8), L(12), M(1), N(1), R(2),S(2), T(2),V(539), W(1), Y(1)	A(25), D(1), F(1), H(1), I(2),K(2), L(1), R(2), S(6), T(1), V(294)	C(1), H(2), K(1), L(20), N(7), Q(1), R(1), V(18), Y(1)	C(1), H(3), L(5), R(4), W(10), Y(4)	I(2), L(4), S(1), W(1), Y(2), V(20)	L(2), M(2)	L(1),V(1)	L(1)I(1)
	37	C	C(128)	C(598)	C(336)	C(52)	C(27)	C(30)	C(4)	C(2)	C(2)
Core domain (domain 3)	38	F	L(1), V(1), F(126)	F(596), L(1), S(1)	F(332), L(3), S(1)	F(52)	F(26), L(1)	F(30)	F(4)	F(2)	F(2)
	39	I	H(2), I(2), L(111), Q(11), S(1), V(1)	E(1), G(1), I(258), L(52), M(68), S(2), T(207), V(9)	R(1), Y(1),H(3), L(60), Q(271)	I(36), L(5), M(2), T(2), V(7)	A(6), G(1), M(2), T(18)	L(30)	L(4)	L(2)	L(2)
	40	T	H(6), K(36), N(78), Q(1), R(1), S(1),T(5)	A(1), D(1), E(1), H(1), K(204), N(7), Q(20), R(51), S(12), T(300)	A(4), K(70), N(6), R(12), S(2), T(242)	K(5), S(3), T(46)	K(2), M(2), R(3), S(2), T(18)	D(1), H(1), N(28)	K(4)	K(1), Q(1)	Q(2)
	41	K	K(128)	K(597), Q(1)	K(335),T(1)	K(52)	K(27)	K(30)	K(4)	K(2)	K(2)
	42	A	G(128)	A(110), G(487), T(1)	D(1), G(335)	G(52)	G(27)	G(30)	G(4)	G(2)	G(2)
	43	L	L(128)	L(598)	L(335), S(1)	F(1), L(51)	L(27)	L(30)	L(4)	L(2)	L(2)
	44	G	G(128)	G(596), S(2)	A(1), C(1), G(333),S(1)	G(52)	G(27)	G(30)	G(4)	G(2)	G(2)
	45	I	I(127), V(1)	I(598)	T(3), V(1), I(332)	I(50), V(2)	I(27)	I(30)	I(4)	I(2)	I(2)
	46	S	F(1), S(127)	C(1), F(14), I(1), L(1), P(1), S(564), V(3), Y(13)	C(1), F(1), H(1), P(1), S(312), Y(20)	S(52)	S(27)	S(30)	Y(1), S(3)	S(2)	C(1),S(1)
	47	Y	Y(128)	F(1), H(13), N(6), Y(578)	H(2), N(4), Y(330)	Y(52)	H(3), Y(24)	Y(30)	Y(4)	H(2)	Y(1), N(1)
	48	G	G(128)	G(598)	C(1), D(2), G(333)	G(52)	G(27)	G(30)	G(4)	G(2)	G(2)
Basic domain or TAR binding domain (domain 4)	49	R	K(1), R(127)	R(598)	R(336)	R(52)	R(27)	R(30)	R(4)	R(2)	R(2)
	50	K	K(126), R(2)	K(598)	K(336)	K(52)	K(27)	K(30)	K(4)	K(2)	E(2)
	51	K	K(128)	K(597), N(1)	E(1), K(333), R(2)	K(52)	K(27)	K(30)	K(4)	K(2)	K(2)
	52	R	R(123), W(5)	G(1), Q(2),R(584), W(11)	R(327), W(9)	R(52)	R(27)	R(30)	R(4)	R(2)	R(2)
	53	R	K(34), R(91), S(2), T(1)	G(9), K(9), N(1), R(571), S(8)	G(7), K(4), R(317), S(8)	G(9), K(1), R(40), S(2)	R(27)	K(21), N(1), R(5), S(3)	R1, S3	R(2)	R(2)
	54	Q	H(18), K(1),L(2), P(13), Q(86), R(8)	H(1), K(4), P(12), Q(571), R(9), S(1)	K(3),P(5), Q(332), R(6)	K(1), L(1), P(1), R(1), Q(48)	Q(27)	P(2), Q(2), H(26)	Q(1), R(3)	Q(2)	Q(1), P(1)
	55	R	R(128)	G(1), H(1), M(1), Q(1), R(594)	R(336)	R(52)	R(27)	R(30)	R(4)	R(2)	R(2)
	56	R	R(127), Q(1)	H(1), P(1), Q(3), R(593)	H(1), R(335)	Q(1), R(51)	R(27)	R(30)	R(2),H(2)	R(2)	T(2)
	57	R	D(1), G(95), K(1), R(31)	A(2), G(18), K(4), R(566), S(3), T(5)	T(1), G(1), H(1), N(12), Q(1), R(25), S(295)	G(6), K(4), S(1), T(1), R(40)	R(27)	A(5), G(18), R(6), S(1)	G(1),A(1),R(2)	A(1),S(1)	T(2)
Domain 5	58	A	A(13), P(17), S(5), T(93)	A(313), D(2), H(1), L(1), N(1), P(121), S(54), T(105)	A(202), G(1), P(1), S(2), T(130)	A(5), L(1), P(19), S(8), T(19)	A(2),P(2), S(3), T(20)	A(3), P(4), S(1), T(22)	T(4)	A(2)	T(2)
	59	H	D(1), H(3), L(1), P(112), S(11)	A(4), D(7), H(60), N(3),P(486),S(32),T(5), Y(1),	H(5), P(321), T(3), S(7),	A(1), H(2), P(43), S(6)	P(24), S(3)	H(4), P(25), S(1)	P(4)	P(2)	P(2)
	60	Q	H(33), P(8), Q(81), R(3), Y(3)	D(5), E(34), H(13), K(13), L(6), N(5), P(52), Q(458), R(8), S(2), T(1), Y(1)	A(6), H(1), L(3), P(287), Q(12), R(4), S(22), T(1)	E(1), H(3), K(2), P(3), Q(43)	H(3), K(1), P(2), Q(19), R(1), Y(1)	A(1), H(4), P(7), Q(11), S(5), W(1)	K(1), A(3)	S(1),P(1)	Y(2)
	61	N	C(2), D(1), G(13),R(9), S(98)	A(2), G(500), H(16), N(48), S(31), Y(1)	D(2), G(10), N(9), R(9), S(305), T(1)	D(7), G(34), N(1), S(10)	D(1), G(2), S(24)	D(2), G(15), N(4), S(9)	S(3),G(1)	G(2)	A(1), N(1)
	62	S	C(1), D(2), N(39), R(1),	C(17), D(5),G(22),H(10) , N(57), R(20), S(465), T(1),	C(1), G(9), N(17)0, R(1), S(308)	D(16), G(23), N(9), S(3)	G(1), N(3), S(22), T(1)	N(6), S(24)	L(1),V(3)	S(2)	S(2)

			S(84), Y(1)	Y(1)							
	63	Q	E(7), G(1), K(115), Q(2), T(3)	A(2), E(91), K(76), P(49), Q(370), S(9), W(1)	A(5), E(260), G(1), K(68), T(2)	E(1), K(2), P(5), Q(43), T(1)	E(3), K(9), P(2), Q(12), T(1)	A(1), E(4), K(21), Q(1), T(3)	Q(4)	K(2)	K(1), E(1)
	64	T	A(1), D(120), N(7)	A(37), D(29),G(3), H(1),I(26), N(63), P(10), R(1), S(20), T(406), V(2)	T(1), S(8), N(8), G(12), E(2), D(299), A(6)	A(34), D(4), I(1), T(11), V(2)	D(1), I(9), N(6), S(4), T(2), V(5)	D(26), E(1), G(1), N(2)	D(4)	T(1),N(1)	N(2)
	65	H	H(128)	D(29), H(491), N(55), P(1), R(11), Y(11),	H(327), N(4), R(3), Y(1), S(1)	H(45), N(4), Y(3)	H(27)	H29, R1	H(3),N(1)	H(2)	H(2)
	66	Q	Q(128)	Q(598)	Q(336)	Q(52)	Q(27)	K(1), Q(29)	Q(4)	Q(2)	K(2)
	67	A	D(4), H(1), I(4), N(108), S(2), T(9)	A(167), D(54), E(16), G(36), I(11), K(1), L(1), N(3), R(1), S(10), T(9), V(289)	A(1), D(79), H(2), I(1), K(1), N(240), S(5), T(7)	A(1), H(2), P(43), S(6), D(35), G(1), V(15)	D(14), H(1), N(12)	D(3), N(26), T(1)	T(1),N(2),D(1)	D(2)	D(2)
	68	S	F(1), H(2), L(11), P(109), S(4), T(1)	A(18), D(12), F(3), H1(1),L(1), N(5), P(125),S(407), Y(11), T(5)	F(6), H(8), I(5), L(159), N(1), P(144), S(13)	A(1), H(3), L(1), S(6), P(41)	F(1), L(6), P(17), S(3)	P(30)	L(1),H(1),S(1),P(1)	P(2)	P(2)
	69	L	I(84), K(1), L(4), T(2), V(37)	I(14), V(5), S(1), P(7), L(571)	I(223), L(21), T(2), V(90)	I(34), L(4), V(14)	I(1), L(1), V(25)	I(3), V(27)	I(4)	I(2)	I(1),L(1)
	70	S	P(89), Q(9), R(4), S(26)	P(174), Q(5), S(419)	L(2), P(34), R(1), S(299)	P(47), Q(1), S(4)	P(20), R(1), S(6)	P(28), R(2)	P(2),S(2)	P(2)	P(1),R(1)
	71	K	E(5), K(119), Q(3), N(1)	A(1), D(1), E(44), K(537), N(8), Q(2), R(2), T(3)	E(15), K(318), N(3)	D(1), E(9), K(42)	E(7), K(20)	E(5), K(24), Q(1)	K(4)	G(1),E(1)	K(2)

Table 3: Positively selected sites selected by all three (SLAC, FEL, iFEL) methods in different subtypes of HIV-1 Tat exon 1

Subtype*	Codon [a]	Domain	p-value		
			SLAC [b]	FEL [c]	iFEL [d]
A	19	Acidic	<0.05	<0.001	<0.05
	36	Cysteine-rich	<0.001	<0.001	<0.001
	40	Core	<0.05	<0.001	<0.001
	58	Basic	<0.001	<0.001	<0.001
	62	Basic	<0.001	<0.001	<0.001
	69	Basic	<0.05	<0.001	<0.05
B	24	Cysteine-rich	<0.001	<0.05	<0.001
	32	Cysteine-rich	<0.001	<0.001	<0.001
	58	Basic	<0.001	<0.001	<0.001
	59	Basic	<0.001	<0.001	<0.001
	61	Basic	<0.05	<0.001	<0.001
	63	Basic	<0.05	<0.05	<0.05
	64	Basic	<0.001	<0.001	<0.001
	65	Basic	<0.001	<0.001	<0.001
	68	Basic	<0.001	<0.001	<0.001
	70	Basic	<0.001	<0.001	<0.001
C	4	Acidic	<0.001	<0.001	<0.001
	21	Acidic	<0.05	<0.05	<0.05
	36	Cysteine-rich	<0.05	<0.05	<0.05
	40	Core	<0.001	<0.05	<0.05
	58	Basic	<0.001	<0.001	<0.001
	67	Basic	<0.001	<0.05	<0.001
	68	Basic	<0.001	<0.001	<0.001
	69	Basic	<0.001	<0.001	<0.001
	70	Basic	<0.001	<0.001	<0.05
D	19	Acidic	<0.05	<0.001	<0.001
	62	Basic	<0.05	<0.05	<0.05
F	3	Acidic	Not detected[#]	<0.05	<0.05
	68	Basic	Not detected[#]	<0.05	<0.05
G	29	Cysteine-rich	<0.05	<0.05	<0.05

Foot notes: *In subtypes H, J and K, selection pressure analysis could not be performed because of insufficient number of eligible sequences.

[#]SLAC method could not detect any positively selected position

[a] according to HIV-1$_{HXB2}$ numbering

[b] single-likelihood ancestor counting (SLAC) method

[c] fixed effects likelihood (FEL) method

[d] interior branches likelihood (iFEL) method

Roy *et al. AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Table 4: Detail information about the positively selected codons with statistical significance in different subtypes of HIV-1 Tat exon 1 calculated by SLAC, FEL, and iFEL methods

Subtype	SLAC[a]		FEL[b]		iFEL[c]	
	Position*	p-value	Position*	p-value	Position*	p-value
Subtype A	19	0.0175253	19	0.00977211	19	0.0235536
	36	0.00270279	36	0.00704499	35	0.0450479
	40	0.024587	40	0.000358346	36	0.00145854
	58	5.94E-05	58	2.76E-05	40	0.00690747
	59	0.0196127	59	0.0321661	58	9.73E-06
	62	0.000124971	62	3.46E-05	62	0.00071199
	67	0.0297202	67	0.00753005	69	0.0135648
	68	0.0010292	68	0.000704223		
	69	0.0212549	69	0.0021765		
	70	0.0203055	70	0.0113633		
			71	0.0423675		
Subtype B	24	0.000259111	7	0.00936966	7	0.00255847
	32	1.06543E-05	24	0.012083148	24	0.009146051
	40	0.00341684	32	0.000254588	32	0.001172438
	58	2.38942E-17	40	0.00269402	58	7.34248E-13
	59	4.9532E-07	42	0.0135939	59	0.00862037
	61	0.0165545	58	9.65893E-15	61	0.001074532
	62	0.0021589	59	5.42127E-06	63	0.0312014
	63	0.01439107	61	0.000176651	64	7.99525E-05
	64	3.17905E-07	62	0.0132088	65	0.00998377
	65	1.28295E-05	63	0.00228988	68	2.28406E-09
	68	1.65095E-13	64	2.41611E-05	70	3.45268E-06
	70	7.81209E-08	65	6.4801E-05		
			68	1.91599E-11		
			70	1.53501E-07		
Subtype C	4	9.68E-07	4	3.93E-07	4	2.55E-05
	19	0.0112024	19	0.045844	21	0.0268247
	21	0.00559645	21	0.0470325	29	1.12E-07
	36	0.0137592	29	1.22E-07	31	0.0346563
	39	0.0173557	36	0.0309755	36	0.0313345
	40	0.000498035	40	0.00108228	39	0.0108953
	58	1.28E-08	58	7.91E-08	40	0.0101595
	59	0.00344666	59	0.00344738	58	5.96E-07
	67	0.000426635	67	0.00121861	67	0.00022994
	68	1.51E-17	68	7.77E-16	68	6.55E-15
	69	2.59E-06	69	1.38E-07	69	3.47E-07
	70	1.00E-05	70	6.95E-06	70	0.0242114
	71	0.021025	71	0.048103		
Subtype D	19	0.00174976	19	0.000281545	19	0.000636232
	58	0.000451093	53	0.0266086	62	0.00962657
	59	0.0267532	58	6.53E-05		
	62	0.0146033	59	0.00908178		
	68	0.026916	62	0.0185503		
			68	0.00887027		
			69	0.0273611		
			70	0.0427388		
Subtype F	Not detected		3	0.0418129	3	0.00895379
			39	0.0389108	58	0.0206786
			68	0.0499068	68	0.0207294
			70	0.0278739		
Subtype G	29	0.0359576	29	0.0358406	29	0.00445438
	61	0.0448835	58	0.0396735		
			61	0.0191456		

Foot notes: selection pressure analysis in subtypes H, J, and K could not be performed because of an insufficient number of eligible sequences

[*] according to HIV-1_{HXB2} numbering

[a] single-likelihood ancestor counting (SLAC) method

[b] fixed effects likelihood (FEL) method

[c] interior branches likelihood (iFEL) method

Roy *et al. AIDS Research and Human Retroviruses* 2015; 31(6): 641-648

Table 5: Positively selected sites selected by all three (SLAC, FEL, iFEL) methods in HIV-1 full-length Tat

Subtype	Codon[a]	Domain	p-value		
			SLAC[b]	FEL[c]	iFEL[d]
B	7	First (Acidic)	<0.05	<0.05	<0.05
	24	Second (Cysteine-rich)	<0.001	<0.001	<0.001
	32	Second (Cysteine-rich)	<0.001	<0.001	<0.001
	40	Third	<0.05	<0.001	<0.05
	59	Fifth (Basic)	<0.001	<0.001	<0.001
	61	Fifth (Basic)	<0.05	<0.001	<0.001
	62	Fifth (Basic)	<0.05	<0.05	<0.05
	63	Fifth (Basic)	<0.001	<0.05	<0.05
	64	Fifth (Basic)	<0.001	<0.001	<0.001
	65	Fifth (Basic)	<0.001	<0.001	<0.001
	68	Fifth (Basic)	<0.001	0	<0.001
	70	Fifth (Basic)	<0.001	<0.001	<0.001
	75	Sixth (Exon 2)	<0.001	<0.001	<0.001
	77	Sixth (Exon 2)	<0.001	<0.001	<0.05
	80	Sixth (Exon 2)	<0.05	<0.001	<0.05
	81	Sixth (Exon 2)	<0.001	<0.001	<0.001
	84	Sixth (Exon 2)	<0.001	<0.001	<0.05
	85	Sixth (Exon 2)	<0.001	<0.001	<0.001
	87	Sixth (Exon 2)	<0.001	0	<0.001
	88	Sixth (Exon 2)	<0.001	<0.001	<0.05
	90	Sixth (Exon 2)	<0.05	<0.05	<0.05
	93	Sixth (Exon 2)	<0.001	<0.001	<0.001
	98	Sixth (Exon 2)	<0.001	<0.001	<0.001
C	4	First (Acidic)	<0.001	<0.001	<0.001
	21	First (Acidic)	<0.05	<0.05	<0.05
	29	Second (Cysteine-rich)	<0.05	<0.001	<0.001
	39	Third	<0.05	<0.05	<0.001
	40	Third	<0.05	<0.001	<0.05
	58	Fifth (Basic)	<0.001	<0.001	<0.001
	67	Fifth (Basic)	<0.001	<0.001	<0.001
	68	Fifth (Basic)	<0.001	<0.001	<0.001
	69	Fifth (Basic)	<0.001	<0.001	<0.001
	70	Fifth (Basic)	<0.001	<0.001	<0.05
	75	Second Exon	<0.001	<0.001	<0.001
	77	Second Exon	<0.05	<0.001	<0.05
	80	Second Exon	<0.001	<0.001	<0.001
	81	Second Exon	<0.001	<0.001	<0.001
	87	Second Exon	<0.05	<0.05	Not selected
	95	Second Exon	<0.001	<0.001	<0.05
	97	Second Exon	<0.001	<0.001	<0.05
	100	Second Exon	<0.05	<0.001	<0.001

Foot note: [a] according to HIV-1HXB2 numbering
[b] single-likelihood ancestor counting (SLAC)
[c] fixed effects likelihood (FEL), and
[d] interior branches likelihood (iFEL) approach

Grey colored rows indicate commonly selected sites in both subtypes. p-values were shown as <0.05 and <0.001.

Table 6: Substitution rates and time of the most recent common ancestors (tMRCAs) of subtypes B and C for all tested molecular clock models

Subtype	Prior, distribution [standard deviation]	Molecular clock	Marginal likelihood	Bayes Factor(vs. Lognormal)	Substitution rate			tMRCA			
					Mean	95% HPD		Median	Mean	95% HPD	
						Lower	Upper			Lower	Upper
B	Normal [$2.5x10^{-3}$-$5x10^{-4}$]	Strict	-8861.6551	63.282	1.53E-03	1.03E-03	2.08E-03	1933	1931	1907	1952
		Exponential	-8910.0671	41.921	1.50E-03	6.57E-04	2.53E-03	1917	1907	1837	1958
		Lognormal	-8866.5294	-	1.49E-03	8.59E-04	2.08E-03	1931	1928	1896	1956
C	Normal [$2.5x10^{-3}$-$5x10^{-4}$]	Strict	-3641.8741	10.648	2.14E-03	1.35E-03	2.91E-03	1956	1954	1934	1970
		Exponential	-3636.6521	-26.721	2.13E-03	1.25E-03	3.05E-03	1946	1942	1911	1967
		Lognormal	-3638.9247	-	2.14E-03	1.39E-03	2.88E-03	1955	1954	1935	1971

Roy *et al. PLoS ONE* 10(6): e0129896.)

Appendix Table 1a: Distribution HIV-1 Tat sequences of different subtypes (N=1179) by continent of origin (Data set 1)

Continent	A	B	C	D	F	G	H	J	K	Total
Asia	1	48	18	3	-	2	-	-	-	**72 (6.1%)**
Africa	79	3	293	15	10	17	1	1	2	**421 (35.7)**
Europe	48	123	13	1	9	9	3	1	-	**207 (17.6%)**
South America	-	75	8	32	8	2	-	-	-	**125 (10.6%)**
North America	-	330	4	1	-	-	-	-	-	**335 (28.4%)**
Oceania	-	19	-	-	-	-	-	-	-	**19 (1.6%)**
Total	**128 (10.9%)**	**598 (50.7%)**	**336 (28.5%)**	**52 (4.4%)**	**27 (2.3%)**	**30 (2.5%)**	**4 (0.3%)**	**2 (0.2%)**	**2 (0.2%)**	**1179 (100%)**

Appendix Table 1b: Accession numbers of the different subtypes of HIV-1 Tat Exon-1 used in this study obtained from the Los Alamos National Laboratory (LANL) HIV sequence database (Data set 1)

Subtype A (n=128)				
A1.CD.02.02CD_KTB035.AM000055	A1.CD.97.97CD_KCC2.AM000053	A1.CD.97.97CD_KTB13.AM000054	A1.CH.03.HIV_CH_BID_V3538_2003.JQ403028	
A1.CY.05.CY021.FJ388892	A1.CY.05.CY022.FJ388893	A1.CY.05.CY051.FJ388903	A1.CY.05.CY057.FJ388906	A1.CY.05.CY064.FJ388909
A1.CY.05.CY106.FJ388925	A1.CY.05.CY121.FJ388932	A1.CY.05.CY140.FJ388938	A1.CY.06.CY164.FJ388946	A1.CY.06.CY171.FJ388950
A1.CY.06.CY173.FJ388951	A1.CY.07.CY182.JF683737	A1.CY.07.CY196.JF683748	A1.CY.07.CY207.JF683759	A1.CY.07.CY208.JF683760
A1.CY.07.CY213.JF683763	A1.CY.08.CY218.JF683767	A1.CY.08.CY230.JF683779	A1.CY.08.CY235.JF683782	A1.CY.08.CY236.JF683783
A1.CY.09.CY243.JF683789	A1.CY.09.CY255.JF683798	A1.ES.05.X1608_8.FJ670519	A1.ES.06.X2110.FJ670523	A1.IT.02.60000.EU861977
A1.KE.00.KER2008.AF457052	A1.KE.00.KER2009.AF457053	A1.KE.00.KER2012.AF457055	A1.KE.00.KNH1144.AF457066	A1.KE.00.KNH1207.AF457068
A1.KE.00.KNH1209.AF457069	A1.KE.00.KNH1211.AF457070	A1.KE.00.KSM4024.AF457077	A1.KE.00.KSM4030.AF457079	A1.KE.00.MSA4069.AF457080
A1.KE.00.MSA4070.AF457081	A1.KE.00.MSA4072.AF457083	A1.KE.00.MSA4076.AF457084	A1.KE.00.MSA4079.AF457086	A1.KE.00.NKU3005.AF457089
A1.KE.01.ML1901PCR2.EU110095	A1.KE.01.ML1945.EU110088	A1.KE.02.ML1990.EU110092	A1.KE.06.06KECst_001.FJ623487	A1.KE.06.06KECst_005.FJ623481
A1.KE.06.06KECst_006.FJ623475	A1.KE.06.06KECst_007.FJ623476	A1.KE.06.06KECst_009.FJ623480	A1.KE.06.06KECst_013.FJ623485	A1.KE.06.06KECst_016.FJ623483
A1.KE.06.06KECst_017.FJ623488	A1.KE.06.06KECst_019.FJ623478	A1.KE.06.06KECst_020.FJ623482	A1.KE.06.06KECst_021.FJ623477	A1.KE.06.06KECst_025.FJ623486
A1.KE.06.06KECst_028.FJ623479	A1.KE.11.DEMA111KE002.KF716474	A1.KE.11.DEMA11KE001.KF716475	A1.KE.86.ML013_10.AY322184	
A1.KE.86.ML170_1986.AF539405	A1.KE.94.Q23_17.AF004885	A1.KE.97.ML605_3.AY322190	A1.KE.97.ML752.AY322193	
A1.KE.99.KNH1088.AF457063	A1.KE.99.KNH1135.AF457065	A1.KE.99.KSM4021.AF457075	A1.RU.05.RU_560_1125_JA.JQ292895	
A1.RU.06.RU_915_1016.JQ292896	A1.RU.06.RU_915_1035.JQ292897	A1.RU.06.RU_915_1038.JQ292898	A1.RU.06.RU_SP_B_049.JQ292900	
A1.RU.07.Irkutsk_5.JQ292891	A1.RU.08.DEMA108RU003.KF716491	A1.RU.08.DEMA108RU004.KF716492	A1.RU.08.RUA001.JQ292893	
A1.RU.08.RUA007.JQ292894	A1.RU.10.10RU6617.JX500696	A1.RU.10.10RU6792.JX500695	A1.RU.11.11RU6950.JX500694	A1.RW.07.pR463F.JX236677
A1.RW.07.pR880F.JX236678	A1.RW.11.DEMA111RW002.KF716472	A1.RW.92.92RW008.AB253421	A1.RW.92.92RW025A.AB287377	
A1.RW.93.93RW_024.AY713406	A1.RW.93.93RW037A.AB287379	A1.SE.94.SE7253.AF069670	A1.SE.94.SE7535.AF069671	A1.SE.95.SE8538.AF069669
A1.SE.95.SE8891.AF069673	A1.SE.95.UGSE8131.AF107771	A1.SN.01.DDJ369.AY521631	A1.SN.96.DDJ360.AY521630	A1.TZ.01.A173.AY253305
A1.TZ.01.A341.AY253314	A1.TZ.97.97TZ02.AF361872	A1.TZ.97.97TZ03.AF361873	A1.UA.00.98UA0116.AF413987	A1.UA.01.01UADN139.DQ823357
A1.UA.01.01UAOD89.DQ823367	A1.UG.07.p191084.JX236669	A1.UG.07.p191845.JX236671	A1.UG.07.p9004SDM.JX236676	A1.UG.07.PP6_F3_A2.HM027823
A1.UG.09.DEMA109UG001.KF716478	A1.UG.11.DEMA110UG001.KF859745	A1.UG.11.DEMA110UG009.KF716486	A1.UG.85.U455_U455A.M62320	
A1.UG.90.UG273A.L22957	A1.UG.90.UG275A.L22951	A1.UG.92.UG029.AB098332	A1.UG.98.98UG57134.AF484507	A1.UG.98.98UG57135.AF484508
A1.UG.98.98UG57136.AF484509	A1.UG.98.98UG57142.AF484512	A1.UG.99.99UGG03379.AF484493	A1.UG.x.UG031.AB098330	
A1.UZ.02.02UZ0663.AY829210	A1.ZA.01.TV314.FJ647148	A1.ZA.04.04ZASK162B1.DQ396400	A2.CD.97.97CDKS10.AF286241	
A2.CD.97.97CDKTB48.AF286238	A2.CM.01.01CM_1445MV.GU201516	A2.CY.94.94CY017_41.AF286237		
Subtype B (n=598)				
B.AR.00.ARMS008.AY037269	B.AR.02.02AR114146.DQ383746	B.AR.03.03AR137681.DQ383748	B.AR.03.03AR138910.DQ383749	B.AR.04.04AR151263.DQ383751
B.AR.04.04AR151516.DQ383752	B.AR.99.ARMA132.AY037282	B.AU.03.PS2008_Day182.DQ676875	B.AU.04.MS2004_37_016.EF178314	
B.AU.04.MS2004_37_122.EF178420	B.AU.04.MS2004_37_129.EF178427	B.AU.04.Phcsffull04.AY818644	B.AU.04.PS2016_Day380.DQ676879	
B.AU.93.MBC18_MBCC18.AF042102	B.AU.95.C24.AF538304	B.AU.95.C42.AF538305	B.AU.95.C76.AF538306	B.AU.95.C92.AF538307
B.AU.95.MBCC54.AF042103	B.AU.96.MBCC98.AF042104	B.AU.96.MBCD36.AF042105	B.AU.99.1181.AF538302	B.AU.99.1182.AF538303
B.AU.x.15888_30.AY624304	B.AU.x.8634991.AY857144	B.AU.x.VH_VHPCR.AF146728	B.BO.99.BOL0122.AY037270	B.BR.02.02BR002.DQ358805
B.BR.02.02BR008.DQ358808	B.BR.02.02BR011.DQ358809	B.BR.02.02BR1013.JN692432	B.BR.02.02BR2025.JN692435	B.BR.02.02BR2032.JN692439
B.BR.02.02BR2033.JN692440	B.BR.02.02BR2041.JN692443	B.BR.02.02BR2042.JN692444	B.BR.02.04BR1064.JN692433	B.BR.02.05BR1094.JN692470

B.BR.03.03BR1020.JN692445 B.BR.03.03BR1046.JN692447 B.BR.03.BREPM1023.EF637057 B.BR.03.BREPM1024.EF637056 B.BR.03.BREPM1027.EF637054
B.BR.03.BREPM1032.EF637051 B.BR.03.BREPM1033.EF637050 B.BR.03.BREPM1035.EF637049 B.BR.03.BREPM1038.EF637048 B.BR.03.BREPM1040.EF637047
B.BR.03.BREPM2012.EF637046 B.BR.04.04BR1047.JN692450 B.BR.04.04BR1049.JN692451 B.BR.04.04BR1054.JN692453 B.BR.04.04BR1057.JN692455
B.BR.04.04BR1068.JN692457 B.BR.04.BREPM1070.FJ195086 B.BR.05.05BR1074.JN692459 B.BR.05.05BR1077.JN692460 B.BR.05.05BR1078.JN692461
B.BR.05.05BR1079.JN692462 B.BR.05.05BR1080.JN692463 B.BR.05.05BR1082.JN692465 B.BR.05.05BR1089.JN692467 B.BR.05.05BR1095.JN692471
B.BR.05.05BR1101.JN692473 B.BR.05.05BR1104.JN692474 B.BR.05.05BR1107.JN692475 B.BR.05.BREPM1081.FJ195091 B.BR.05.BREPM1084.FJ195088
B.BR.05.BREPM1093.FJ195089 B.BR.06.06BR1119.JN692480 B.BR.07.BP00047_RH01.JN687739 B.CA.06.502_1799_FL02.JF320427
B.CA.96.WC10P_1.AY314044 B.CA.98.CANA3FULL.AY779550 B.CA.98.CANC10FULL.AY779557 B.CA.x.HND_DRPI039V01.DQ322225
B.CH.00.CHLSM1SI0i0007251ppNFLG95.KC797171 B.CH.00.HIV_CH_BID_V3530_2000.JQ403023 B.CH.01.HIV_CH_BID_V3531_2001.JQ403024
B.CH.01.HIV_CH_BID_V3534_2001.JQ403026 B.CH.02.HIV_CH_BID_V3527_2002.JQ403021 B.CH.02.HIV_CH_BID_V3539_2002.JQ403029
B.CH.02.HIV_CH_BID_V4424_2002.JQ403044 B.CH.03.HIV_CH_BID_V4421_2003.JQ403043 B.CH.03.HIV_CH_BID_V4470_2003.JQ403045
B.CH.03.HIV_CH_BID_V4474_2003.JQ403046 B.CH.04.HIV_CH_BID_V4408_2004.JQ403042 B.CH.08.CHLSM2SI5082i0803101ppNFLG8.KC797225
B.CO.01.PCM034.AY561238 B.CU.99.Cu19.AY586542 B.CU.99.Cu43.AY586543 B.CY.05.CY018.FJ388890 B.CY.05.CY020.FJ388891
B.CY.05.CY030.FJ388956 B.CY.05.CY032.FJ388898 B.CY.05.CY033.FJ388957 B.CY.05.CY035.FJ388958 B.CY.05.CY036.FJ388959
B.CY.05.CY037.FJ388899 B.CY.05.CY055.FJ388904 B.CY.05.CY056.FJ388905 B.CY.05.CY065.FJ388910 B.CY.05.CY067.FJ388911
B.CY.05.CY068.FJ388912 B.CY.05.CY070.FJ388960 B.CY.05.CY072.FJ388914 B.CY.05.CY074.FJ388915 B.CY.05.CY082.FJ388962
B.CY.05.CY088.FJ388919 B.CY.05.CY093.FJ388923 B.CY.05.CY096.FJ388963 B.CY.05.CY110.FJ388927 B.CY.05.CY122.FJ403482
B.CY.05.CY124.FJ388933 B.CY.05.CY130.FJ388934 B.CY.05.CY131.FJ388935 B.CY.05.CY142.FJ388965 B.CY.05.CY149.FJ388940
B.CY.06.CY168.FJ388949 B.CY.07.CY181.JF683736 B.CY.07.CY184.JF683738 B.CY.07.CY188.JF683741 B.CY.07.CY189.JF683742
B.CY.07.CY190.JF683743 B.CY.07.CY195.JF683747 B.CY.07.CY197.JF683749 B.CY.07.CY198.JF683750 B.CY.07.CY199.JF683751
B.CY.07.CY201.JF683753 B.CY.07.CY202.JF683754 B.CY.07.CY204.JF683756 B.CY.07.CY214.JF683764 B.CY.07.CY216.JF683765
B.CY.08.CY220.JF683769 B.CY.08.CY224.JF683773 B.CY.08.CY226.JF683775 B.CY.08.CY229.JF683778 B.CY.08.CY232.JF683781
B.CY.08.CY237.JF683784 B.CY.08.CY238.JF683785 B.CY.09.CY241.JF683787 B.CY.09.CY244.JF683790 B.CY.09.CY250.JF683793
B.CY.09.CY251.JF683794 B.CY.09.CY253.JF683796 B.CY.09.CY258.JF683801 B.CY.09.CY263.JF683805 B.CY.09.CY266.JF683807
B.DE.03.HIV_DE_BID_V3307_2003.JQ403048 B.DE.86.HAN.U43141 B.DE.x.DEMBXXDE001.KC596067 B.DK.01.CTL_017.EF514705 B.DK.01.CTL_018.EF514706
B.DK.01.CTL_030.EF514708 B.DK.01.CTL_035.EF514710 B.DK.01.CTL_041.EF514711 B.DK.04.PMVL_012.EF514699 B.DK.04.PMVL_013.EF514700
B.DK.04.PMVL_018.EF514697 B.DK.04.PMVL_039.EF514703 B.DK.04.PMVL_049.EF514701 B.DK.07.PMVL_011.FJ694790 B.EC.89.EC003.AY173959
B.EC.89.EC102.AY173960 B.ES.05.X1890.EU786672 B.ES.06.X1998.EU786676 B.ES.06.X2102.EU786677 B.ES.07.X2210_3.EU786679
B.ES.07.X2231.EU786680 B.ES.08.ES_X2515_3.GQ372988 B.ES.08.P2008.FJ670531 B.ES.08.P2021_3.FJ853620 B.ES.08.X2425_2.FJ670525
B.ES.08.X2510_2.FJ853622 B.ES.08.X2555_2.GU362883 B.ES.08.X2574_2.GU362886 B.ES.09.P2149_3.GU362881 B.ES.09.X2689_2.GU362885
B.ES.10.DEMB10ES002.KC473842 B.ES.10.DEMB10VE001.JX140659B.FR.08.DEMB08FR002.JX140654B.FR.09.DEMB09FR001.KF716494
B.FR.09.DEMB09FR002.KF716495 B.FR.11.DEMB11FR001.KF716496 B.FR.83.HXB2_LAI_IIIB_BRU.K03455 B.GA.88.OYI_397.M26727
B.GB.03.MM33d12p.HM586187 B.GB.04.MM39d11p.HM586193 B.GB.05.MM43d368_GN1.HM586209 B.GB.05.MM45d22_GN1.HM586210 B.GB.83.CAM1.D10112
B.GB.86.GB8_46R.AJ271445 B.GB.x.MANC.U23487B.GE.03.03GEMZ004.DQ207940 B.GE.03.03GEMZ010.DQ207942 B.HK.06.HK002.FJ460499
B.HT.05.05HT_129473.EU839603 B.HT.05.05HT_129696.EU839601 B.IN.x.11807.EF694037 B.JM.05.05JM_KJ108.EU839605
B.JM.09.09JM_PF09XN.HM030559 B.JM.09.09JM_PF09YT.HM030560 B.JM.09.09JM_PF0B8J.HM030561 B.JM.09.09JM_PF0B97.HM030562
B.JM.09.09JM_PF0B9L.HM030565 B.JP.00.DR2508.AB289588 B.JP.00.DR2510.AB287372 B.JP.01.DR388.AB289590 B.JP.03.285.AB428558
B.JP.04.04JPDR6075B.AB221126 B.JP.04.DR5913.AB480696 B.JP.04.DR6174.AB480692 B.JP.04.DR6175.AB480694 B.JP.05.DR6538.AB287363
B.JP.05.DR6737.AB287364 B.JP.08.08JP_NMC104_clone_01.AB731663 B.JP.09.09JP_NMC127_clone_07.AB731667 B.JP.11.DEMB11JP002.KF716497
B.JP.12.DEMB12JP001.KF716498 B.JP.x.DR1673.AB564745 B.JP.x.DR1712.AB604946 B.JP.x.JRC03B.AB565496 B.JP.x.pJPDR0796B02.AB565478
B.JP.x.pJRC57B09.AB641836 B.KR.03.03HJY8.JQ316131 B.KR.03.03KDE11.JQ316128 B.KR.03.03KGS5.JQ316132 B.KR.03.03YGS3.JQ316135
B.KR.04.04CWS5.JQ316133 B.KR.04.04KJin8_1955.DQ295195 B.KR.04.04KJS8.JQ316130 B.KR.04.04KMK5.JQ316126 B.KR.04.04KYR8.DQ295196
B.KR.05.05CSR3.DQ837381 B.KR.05.05YJN2.JQ316134 B.KR.12.12KYY10_10742.KF561441 B.KR.91.91OSG10.KF561442 B.KR.97.WK.AF224507
B.NL.99.671_99T12.AY423381 B.PE.06.502_0491_wg5.JF320183 B.PE.06.502_0524_FL04.JF320008 B.PE.06.502_0648_FL02.JF320215
B.PE.06.502_0841_FL04.JF320208 B.PE.06.502_2717_RH03.JF320230 B.PE.06.502_2794_FL05.JF320244 B.PE.07.502_0525_wg5.JF320191

B.PE.07.502_1047_wg5.JF320226 B.PE.07.502_2254_FL6.JF320018 B.PE.07.502_2622_wg1.JF320189 B.PL.08.BP00074_RH02.JN687774
B.PL.x.DEMBXXPL001.KC596069 B.PY.02.02PY_PSP0019.JN251896 B.PY.02.02PY_PSP0090.JN251901 B.PY.03.03PY_PSP0115.JN251906
B.RU.04.04RU128005.AY682547 B.RU.04.04RU129005.AY751406 B.RU.04.04RU139089.AY751407 B.RU.09.09RU4457.JX500709 B.RU.10.10RU6629.JX500707
B.RU.11.11RU21n.JX500708 B.TH.04.04TH317223.JN248321 B.TH.04.04TH803686.JN248333 B.TH.05.05TH356764.JN248344 B.TH.05.05TH357801.JN248346
B.TH.05.05TH645189.JN248353 B.TH.05.05TH736580.JN248354 B.TH.06.AA010a_WG3.JX446800 B.TH.08.AA115c03R.JX448103 B.TH.90.BK132.AY173951
B.TH.96.M149.DQ354119 B.TT.00.00TT_CRC50018.EU839607 B.TW.94.TWCYS_LM49.AF086817 B.UA.01.01UAKV167.DQ823362
B.UA.01.01UAKV252.DQ823363 B.UA.01.01UAKV259.DQ823364 B.US.00.ES1_20.EF363123 B.US.00.RHPA_TF1.JN944917 B.US.00.THRO_TF1.JN944930
B.US.00.WITO_TF1.JN944938 B.US.01.L805.FJ469738 B.US.01.REJO_TF1.JN944911 B.US.01.TRJO_TF1.JN944936 B.US.02.04013226_2_flH11.FJ496078
B.US.02.CR0017Q.FJ469687 B.US.02.CR0058S.FJ469693 B.US.02.F706.FJ469728 B.US.02.F714.FJ469730 B.US.02.F719.FJ469734
B.US.02.F797.FJ469737 B.US.02.HIV_US_BID_V5249_2002.JQ403102 B.US.02.HIV_US_BID_V5276_2002.JQ403105 B.US.02.L8124P.FJ469741
B.US.02.L861P.FJ469748 B.US.02.MTF.FJ469751 B.US.02.PRLS08.FJ469757 B.US.02.PRLS09.FJ469758 B.US.02.PRLS16.FJ469760
B.US.03.1BRHGA_E13_DP.KC312470 B.US.03.AT01480.FJ469686 B.US.03.CR0154X.FJ469701 B.US.03.DEMB03JP004.KC473846 B.US.03.F7157.FJ469731
B.US.03.HDNDRPI032B2.GU562135 B.US.03.HIV_US_BID_V5279_2003.JQ403106 B.US.03.L8107.FJ469739 B.US.03.L8116.FJ469740
B.US.03.L8152.FJ469742 B.US.03.PRLS01.FJ469753 B.US.03.PRLS12.FJ469759 B.US.03.PRLS24.FJ469764 B.US.03.PSL024B10.GU562272
B.US.03.SH8233.FJ469770 B.US.04.014837G.FJ469684 B.US.04.CR0068P.FJ469695 B.US.04.CR0080N.FJ469697 B.US.04.CR0116Y.FJ469698
B.US.04.CR0345Q.FJ469722 B.US.04.ES10_53.EF363127 B.US.04.ES11_2004_culture.KC935957 B.US.04.ES8_43.EF363126
B.US.04.HIV_US_BID_V5258_2004.JQ403103 B.US.04.HIV_US_BID_V5261_2004.JQ403104 B.US.04.HIV_US_BID_V5282_2004.JQ403107
B.US.04.SAMI_WGA1.EU547186 B.US.04.USPI71101EI7y04051pcWG2B9.JN024210 B.US.04.USPI83747EI6y04121pcWG2B5.JN024100 B.US.05.01144G.FJ469682
B.US.05.012286G.FJ469683 B.US.05.04013396_0_flE6.FJ496081 B.US.05.05US_SAJ_NVS12.JF689852 B.US.05.05US_SAJ_NVS16.JF689854
B.US.05.05US_SAJ_NVS3.JF689856 B.US.05.05US_SAJ_NVS5.JF689857 B.US.05.05US_SAJ_NVS8.JF689859 B.US.05.05US_SAJ_NVS9.JF689860
B.US.05.502_0223_wg1.JF320059 B.US.05.502_1400_FL02.JF320043 B.US.05.502_2008_FL04.JF320484 B.US.05.502_2136_FL02.JF320185
B.US.05.502_2495_wg02.JF320054 B.US.05.CR0175S.FJ469703 B.US.05.CR0307R.FJ469717 B.US.05.CR0312W.FJ469718 B.US.05.L8249.DQ886037
B.US.05.L827.FJ469747 B.US.05.STCO_3_A2.KC312435 B.US.05.USPI38417EI33y05051pcWG2B2.JN024363 B.US.05.USPI88403EI14y05121pcWG2B7.JN024344
B.US.06.06US_SAJ_C164_SC.JF689862 B.US.06.06US_SAJ_C165_TJ.JF689863 B.US.06.06US_SAJ_C166_SG.JF689864 B.US.06.06US_SAJ_C167_LH.JF689865
B.US.06.06US_SAJ_C168_LS.JF689866 B.US.06.06US_SAJ_NVS22.JF689870 B.US.06.06US_SAJ_NVS23.JF689871 B.US.06.06US_SAJ_NVS27.JF689872
B.US.06.06US_SAJ_NVS32.JF689874 B.US.06.06US_SAJ_NVS35.JF689875 B.US.06.06US_SAJ_NVS39.JF689876 B.US.06.502_0062_FL04.JF320613
B.US.06.502_0176_FL06.JF320363 B.US.06.502_0227_FL05.JF320036 B.US.06.502_0309_wg13.JF320048 B.US.06.502_0322_RH04.JF320308
B.US.06.502_0341_FL05.JF320003 B.US.06.502_0346_wg02.JF320097 B.US.06.502_0572_FL06.JF320038 B.US.06.502_0717_FL02.JF320526
B.US.06.502_0762_RH05.JF320329 B.US.06.502_0839_wg01.JF320263 B.US.06.502_0897_RH05.JF320502 B.US.06.502_0923_wg07.JF320160
B.US.06.502_0961_RH08.JF320011 B.US.06.502_1046_FL04.JF320564 B.US.06.502_1055_FL01.JF320169 B.US.06.502_1174_FL09.JF320053
B.US.06.502_1211_FL01.JF320151 B.US.06.502_1512_FL01.JF320356 B.US.06.502_1619_FL06.JF320126 B.US.06.502_1919_RH01.JF320184
B.US.06.502_2437_RH01.JF320592 B.US.06.502_2667_FL03.JF320145 B.US.06.701010068_E_J1.GU331147 B.US.06.BP00054_RH01.JN687749
B.US.06.BP00055_RH01.JN687750 B.US.06.BP00057_RH02.JN687758 B.US.06.CH106_TF1.JN944897 B.US.06.CH40_TF1.JN944905
B.US.06.CH58_TF1.JN944907 B.US.06.CH77_TF1.JN944909 B.US.06.CR0047U.FJ469691 B.US.06.CR0078.FJ469696 B.US.06.CR0131.FJ469700
B.US.06.CR0215.FJ469708 B.US.06.CR0222X.FJ469709 B.US.06.CR0276Z.FJ469714 B.US.06.CR0361T.FJ469723 B.US.06.HIV_US_BID_V3047_2006.JQ403068
B.US.06.HIV_US_BID_V4503_2006.JQ403095 B.US.06.HIV_US_BID_V4506_2006.JQ403097 B.US.06.HTM385B7.GU562266 B.US.07.07US_SAJ_C154.JF689877
B.US.07.07US_SAJ_C156.JF689879 B.US.07.07US_SAJ_C161_H1.JF689883 B.US.07.07US_SAJ_C163_H3.JF689885 B.US.07.07US_SAJ_C166_MS.JF689886
B.US.07.07US_SAJ_NVS42.JF689889 B.US.07.07US_SAJ_NVS48.JF689890 B.US.07.07US_SAJ_NVS54.JF689892 B.US.07.07US_SAJ_NVS55.JF689893
B.US.07.306344_FL.JX863921 B.US.07.502_0938_RH04.JF320631 B.US.07.502_0965_RH01.JF320385 B.US.07.502_1115_wg1.JF320045
B.US.07.502_1368_RH02.JF320173 B.US.07.502_1478_wg4.JF320150 B.US.07.502_1500_RH01.JF320387 B.US.07.502_1504_RH07.JF320394
B.US.07.502_1518_RH03.JF320117 B.US.07.502_1709_04.JF320467 B.US.07.502_1897_wg6.JF320182 B.US.07.502_2000_RH03.JF320279
B.US.07.502_2241_RH13.JF320539 B.US.07.502_2289_05.JF320197 B.US.07.502_2586_RH04.JF320131 B.US.07.BP00058_RH01.JN687759
B.US.07.BP00063_RH01.JN687761 B.US.07.BP00064_RH01.JN687762 B.US.07.BP00067_RH01.JN687763 B.US.07.CR0027M.FJ469689
B.US.07.CR0275.FJ469713 B.US.07.CR0339X.FJ469721 B.US.07.HIV_US_BID_V3010_2007.JQ403058 B.US.07.HIV_US_BID_V3020_2007.JQ403059
B.US.07.HIV_US_BID_V3021_2007.JQ403060 B.US.07.HIV_US_BID_V3036_2007.JQ403064 B.US.07.HIV_US_BID_V3053_2007.JQ403071

B.US.07.HIV_US_BID_V3115_2007.JQ403075 B.US.07.HIV_US_BID_V3120_2007.JQ403078 B.US.07.HIV_US_BID_V3512_2007.JQ403092
B.US.07.HIV_US_BID_V3515_2007.JQ403093 B.US.07.HIV_US_BID_V4516_2007.JQ403096 B.US.07.MCST_B17.KC312583 B.US.07.WARO_A13.KC312386
B.US.08.08US_SAJ_C202.JF689894 B.US.08.08US_SAJ_C203.JF689895 B.US.08.08US_SAJ_C204.JF689896 B.US.08.B.700010607.S.0dps.BF13.JX974238
B.US.08.BP00061_RH01.JN687760 B.US.08.HIV_US_BID_V3024_2008.JQ403061 B.US.08.HIV_US_BID_V3032_2008.JQ403063
B.US.08.HIV_US_BID_V3048_2008.JQ403069 B.US.08.HIV_US_BID_V3050_2008.JQ403070 B.US.08.HIV_US_BID_V3114_2007.JQ403074
B.US.08.HIV_US_BID_V3128_2008.JQ403082 B.US.08.HIV_US_BID_V4120_2008.JQ403031 B.US.08.HIV_US_BID_V4124_2008.JQ403035
B.US.08.HIV_US_BID_V4388_2008.JQ403083 B.US.08.HIV_US_BID_V4389_2008.JQ403084 B.US.08.HIV_US_BID_V4391_2008.JQ403086
B.US.08.HIV_US_BID_V4392_2008.JQ403087 B.US.08.HIV_US_BID_V4393_2008.JQ403088 B.US.08.HIV_US_BID_V4394_2008.JQ403089
B.US.08.HIV_US_BID_V4397_2008.JQ403091 B.US.08.HIV_US_BID_V4489_2008.JQ403094 B.US.09.C1P.GU733713 B.US.09.DEMB09US002.JX140657
B.US.09.DEMB09US003.KC473824 B.US.09.ES38.KC935958 B.US.09.LTNP1.KC935959 B.US.10.10CB4_45E6.KF526141 B.US.10.C.HQ846913
B.US.10.CP1.JN397365 B.US.10.DEMB10US001.KC473825 B.US.10.DEMB10US003.KC473826 B.US.10.DEMB10US004.KC473827
B.US.10.DEMB10US007.KC473828 B.US.10.DEMB10US009.KC473829 B.US.10.DEMB10US011.KC473830 B.US.10.VC1.JN397364
B.US.11.19CB1_induced.KF526228 B.US.11.20CB4_46F1.KF526265 B.US.11.22CC9_induced.KF526312 B.US.11.23CB6_induced.KF526323
B.US.11.CP10_3A.KF384798 B.US.11.CP12_10.KF384799 B.US.11.CP13_2.KF384800 B.US.11.CP3_6.KF384801 B.US.11.CP4_2B.KF384802
B.US.11.CP5_3A.KF384803 B.US.11.CP6_2E.KF384804 B.US.11.CP7_2B.KF384805 B.US.11.CP8_4.KF384806 B.US.11.CP9_1A.KF384807
B.US.11.DEMB11US002.KC473831 B.US.11.DEMB11US004.KC473832 B.US.11.DEMB11US011.KC473834 B.US.11.DEMB11US015.KC473835
B.US.11.ES22_27.KF384808 B.US.11.ES38.JN397362 B.US.11.ES39_42.KF384810 B.US.11.ES40_24.KF384811 B.US.11.VC3_14.KF384812
B.US.11.VC5_3.KF384813 B.US.11.VC6_2.KF384814 B.US.83.SF2_LAV2_ARV2.K02007 B.US.84.MNCG_MN.M17449 B.US.84.SC.M17450 B.US.84.SF33.AY352275
B.US.85.RJS4_15_ENVBB.M37573 B.US.86.5096_86.AY835749 B.US.86.5127_86.AY835774 B.US.86.AD87_ADA.AF004394 B.US.86.YU_2.M93258
B.US.87.5113_87.AY835758 B.US.88.5160_88.AY835763 B.US.88.WR27.AF286365 B.US.89.P896_89_6.U39362 B.US.90.90US_873.AY713412
B.US.90.US1.AY173952 B.US.90.US2.AY173953 B.US.90.WEAU160_GHOSH.U21135 B.US.91.5048_91.AY835761 B.US.91.DH12_3.AF069140
B.US.91.R3A.AY608577 B.US.91.SUMA_TF1.JN944928 B.US.92.92US657_1_301657_1.U04908 B.US.94.5082_94.AY835773 B.US.94.94US_33931N.AY713410
B.US.95.5073_95.AY835768 B.US.95.USPI90770EI72y95091pcWG2B7.JN024303 B.US.96.1027_05.AY331291 B.US.96.1057_01.AY331292
B.US.96.5155_96.AY835753 B.US.96.USPI55751EI32y96071pcWG2B15.JN024428 B.US.97.1001_07.AY331282 B.US.97.1006_08.AY331284
B.US.97.1012_08.AY331285 B.US.97.Ac_06.AY247251 B.US.97.ACTDM580208A15.GU562033 B.US.97.CQLDR03A1.GU562058 B.US.97.ZP6248_07_3A1.JN400469
B.US.98.1058_08.AY331294 B.US.98.15384_1.DQ853463 B.US.98.98USHVTN1925c1.AY560107 B.US.98.98USHVTN3605c9.AY560108
B.US.98.98USHVTN8229c6.AY560109 B.US.98.98USHVTN941c1.AY560110 B.US.98.ACT54869022A2.GU562001 B.US.98.HDNDRPI001B10.GU562080
B.US.98.WC3_0498_4.EF175212 B.US.99.169_1999.JN599165 B.US.99.AD17_3C_TA4.GU331247 B.US.99.HIV_US_BID_V5239_1999.JQ403100
B.US.99.PRB959_03.AY331296 B.US.x.1229I.FJ469685 B.US.x.AC_16_0_Days_Consen_fa.DQ127537 B.US.x.AC_59_41_Days_Seq1_fa.DQ127548
B.US.x.AC160_T9_Day_1034_Dom.EU616649 B.US.x.CR0036W.FJ469690 B.US.x.CR0050Z.FJ469692 B.US.x.CR0059T.FJ469694
B.US.x.CR0164U.FJ469702 B.US.x.CR0192W.FJ469704 B.US.x.CR0206U.FJ469705 B.US.x.CR0234.FJ469711 B.US.x.CR0248X.FJ469712
B.US.x.CR0289Z.FJ469715 B.US.x.CR0317N.FJ469719 B.US.x.CR0382N.FJ469725 B.US.x.CR0413T.FJ469726 B.US.x.F701.FJ469727 B.US.x.F703.DQ886031
B.US.x.F710.FJ469729 B.US.x.L8146.DQ886034 B.US.x.L8157.DQ886035 B.US.x.L8185.DQ886036 B.US.x.L8188.FJ469744 B.US.x.L896.FJ469750
B.US.x.MACS2BR_9.AF491739 B.US.x.NC7.AF049495 B.US.x.NCQ.FJ469752 B.US.x.PRLS05.FJ469756 B.US.x.PRLS17.FJ469761 B.US.x.PRLS19.FJ469763
B.US.x.PRLS28.FJ469766 B.US.x.sample_C_BID_D617.JX503075 B.US.x.SH8183.FJ469768 B.US.x.SH8229.FJ469769 B.US.x.SH8241.FJ469771
B.UY.01.01UYTRA1092.AY781126 B.UY.01.01UYTRA1179.AY781127 B.UY.02.02UY_TSU1290.JN235958 B.UY.99.99UY_TRA0177.JN235965
B.YE.02.02YE507.AY795904 B.YE.02.02YE508.AY795905 B.ZA.03.03ZAPS045MB2.DQ396398 B.ZA.85.R84.FJ647145

Subtype C (n=336)

C.AR.01.ARG4006.AY563170 C.BR.02.02BR2022.JN692434 C.BR.04.04BR021.AY727523 C.BR.04.04BR038.AY727524 C.BR.07.BP00040_RH01.JN687737
C.BR.07.DEMC07BR003.JX140663 C.BR.92.BR025_d.U52953 C.BR.98.98BR004.AF286228 C.BW.00.00BW07621.AF443088 C.BW.00.00BW076820.AF443089
C.BW.00.00BW087421.AF443090 C.BW.00.00BW147127.AF443091 C.BW.00.00BW16162.AF443092 C.BW.00.00BW1686.AF443093
C.BW.00.00BW17593.AF443094 C.BW.00.00BW17732.AF443095 C.BW.00.00BW17956.AF443097 C.BW.00.00BW18113.AF443098 C.BW.00.00BW18595.AF443099
C.BW.00.00BW18802.AF443100 C.BW.00.00BW192113.AF443101 C.BW.00.00BW20872.AF443104 C.BW.00.00BW2127214.AF443105
C.BW.00.00BW22767.AF443107 C.BW.00.00BW38193.AF443108 C.BW.00.00BW38428.AF443109 C.BW.00.00BW38713.AF443110 C.BW.00.00BW3876_9.AF443111

C.BW.00.00BW3886_8.AF443112 C.BW.00.00BW3970_2.AF443114 C.BW.00.00BW5031_1.AF443115 C.BW.96.96BW0504.AF110968
C.BW.96.96BW06K18.AF290030 C.BW.96.96BW1104.AF110969 C.BW.96.96BW1210.AF110972 C.BW.96.96BW15B03.AF110973 C.BW.96.96BW16B01.AF110976
C.BW.96.96BW17.AF110979 C.BW.96.96BWM032.AF443075 C.BW.96.96BWMO1_5.AF443074 C.BW.98.98BWMC122.AF443076 C.BW.98.98BWMC134.AF443077
C.BW.98.98BWMC14A3.AF443078 C.BW.98.98BWMO36A5.AF443081 C.BW.99.99BW393212.AF443083 C.BW.99.99BW4745.AF443085
C.BW.99.99BW47547.AF443086 C.BW.99.99BWMC168.AF443087 C.CN.06.06YNLC028sg.KC898980 C.CN.06.06YNLC32sg.KC899004 C.CN.07.07CNYN328.KF835515
C.CN.07.07CNYN336.KF835522 C.CN.09.09YNLX090sg.KC898995 C.CN.09.09YNLX133sg.KC898996 C.CY.05.CY040.FJ388901 C.CY.05.CY069.FJ388913
C.CY.06.CY166.FJ388948 C.CY.06.CY176.FJ388952 C.CY.07.CY187.JF683740 C.CY.07.CY205.JF683757 C.CY.08.CY219.JF683768
C.CY.08.CY221.JF683770 C.CY.09.CY260.JF683803 C.DJ.91.DJ373A.L23065 C.DK.01.CTL_015.EF514713 C.ES.06.DEMC06ES003.KC473844
C.ES.07.X2118_2.EU884500 C.ES.08.X2363_2.EU786681
C.ET.02.02ET_288.AY713417 C.GE.03.03GEMZ033.DQ207941 C.IL.98.98IS002.AF286233 C.IL.99.99ET1.AY255823 C.IN.03.D24.EF469243
C.IN.05.C.IN.05.NIRT723.1.KF766541 C.IN.09.T125_2139.KC156210 C.IN.93.93IN999.AF067154 C.IN.94.94IN476.AF286223
C.IN.98.98IN012.AF286231 C.IN.98.98IN022.AF286232 C.IN.x.VB39.EF694033 C.IN.x.VB49.EF694035 C.KE.00.KER2010.AF457054
C.KE.91.KNH1268.AY945738 C.MW.07.703010085_CH085.w4a.KC156211 C.MW.07.703010167_CH167.w8.KC156213 C.MW.07.703010228_CH228_TFa.KC156120
C.MW.07.CH0010.w24.e11.KC148599 C.MW.07.CH0047.w2.e11.KC148663 C.MW.08.4403bmLwk4_fl11.KF527172 C.MW.08.4419_301.KF527081
C.MW.08.702010293_CH293.w8a.KC156216 C.MW.08.702010432_CH432.w4.KC156218 C.MW.08.702010440_CH440.w4.KC156219 C.MW.08.CH0275.w24.e1.KC149493
C.MW.09.703010256_CH256.w96.KC156214 C.TZ.01.A125.AY253304 C.TZ.01.A207.AY253307 C.TZ.01.A246.AY253308 C.TZ.01.A301.AY253312
C.TZ.01.A388.AY253317 C.TZ.01.A93.AY253303 C.TZ.01.BD16_10.AY253320 C.TZ.01.BD9_11.AY253322 C.TZ.02.CO178.AY734556
C.TZ.02.CO328.AY734551 C.TZ.02.CO3305.AY734558 C.TZ.02.CO6130.AY734559 C.TZ.02.CO6721.AY734560 C.TZ.97.97TZ04.AF361874
C.TZ.97.97TZ05.AF361875 C.TZ.98.98TZ013.AF286234 C.TZ.98.98TZ017.AF286235 C.US.11.17TB4_4G8.KF526226 C.US.98.98US_MSC3018.AY444800
C.US.98.98US_MSC5016.AY444801 C.YE.02.02YE511.AY795906 C.ZA.00.1119MB.AY463229 C.ZA.00.1134MB.AY463217 C.ZA.00.1162MB.AY463224
C.ZA.00.1165MB.AY463230 C.ZA.00.1168MB.AY463231 C.ZA.00.1170MB.AY463225 C.ZA.00.1171MB.AY463232 C.ZA.00.1189MB.AY838565
C.ZA.00.1192M3M.AY463219 C.ZA.00.1195MB.AY463220 C.ZA.00.1210MB.AY463221 C.ZA.00.1217MB.AY463226 C.ZA.00.1225MB.AY463227
C.ZA.01.2004MB.AY463235 C.ZA.01.2134MB.AY463237 C.ZA.02.02ZAPS005MB1.DQ351235 C.ZA.02.02ZAPS006MB1.DQ351220
C.ZA.02.02ZAPS008MB1.DQ275647 C.ZA.02.02ZAPS015MB1.DQ369995 C.ZA.03.03ZAPS017MB1.DQ351224 C.ZA.03.03ZAPS020MB1.DQ275653
C.ZA.03.03ZAPS021MB1.DQ369978 C.ZA.03.03ZAPS023MB1.DQ351225 C.ZA.03.03ZAPS024MB1.DQ396367 C.ZA.03.03ZAPS027MB1.DQ351223
C.ZA.03.03ZAPS030MB1.DQ369996 C.ZA.03.03ZAPS032MB1.DQ445633 C.ZA.03.03ZAPS044MB1.DQ396384 C.ZA.03.03ZAPS046MB1.DQ369984
C.ZA.03.03ZAPS049MB1.DQ369986 C.ZA.03.03ZAPS054MB2.DQ369988 C.ZA.03.03ZAPS055MB1.DQ396373 C.ZA.03.03ZAPS056MB1.DQ396374
C.ZA.03.03ZAPS057MB2.DQ369989 C.ZA.03.03ZAPS059MB2.DQ445634 C.ZA.03.03ZAPS063MB1.DQ396388 C.ZA.03.03ZAPS066MB2.DQ396375
C.ZA.03.03ZAPS067MB2.DQ396389 C.ZA.03.03ZAPS074MB2.DQ351228 C.ZA.03.03ZAPS081MB1.DQ351219 C.ZA.03.03ZAPS083MB1.DQ351229
C.ZA.03.03ZAPS086MB1.DQ275654 C.ZA.03.03ZAPS097MB1.DQ351230 C.ZA.03.03ZAPS099MB1.DQ275655 C.ZA.03.03ZAPS103MB2.DQ275656
C.ZA.03.03ZAPS105MB2.DQ445632 C.ZA.03.03ZAPS108MB1.DQ396378 C.ZA.03.03ZAPS112MB2.DQ396386 C.ZA.03.03ZAPS113MB2.DQ396365
C.ZA.03.03ZAPS118MB1.DQ396368 C.ZA.03.03ZAPS124MB1.DQ369976 C.ZA.03.03ZAPS130MB1.DQ275658 C.ZA.03.03ZAPS136MB1.DQ351231
C.ZA.03.03ZAPS140MB1.DQ369981 C.ZA.03.03ZAPS143MB1.DQ396391 C.ZA.03.03ZAPS151MB1.DQ396392 C.ZA.03.03ZAPS152MB1.DQ396399
C.ZA.03.03ZAPS155MB1.DQ396371 C.ZA.03.03ZASK005B2.DQ011175 C.ZA.03.03ZASK010B2.DQ164104 C.ZA.03.03ZASK011B2.AY901965
C.ZA.03.03ZASK016MB2.DQ351233 C.ZA.03.03ZASK019B2.AY878063 C.ZA.03.03ZASK020B2.AY878064 C.ZA.03.03ZASK034B1.AY878065 C.ZA.03.03ZASK039B2.AY878068
C.ZA.03.03ZASK061B1.AY901968 C.ZA.03.03ZASK062B1.DQ164113 C.ZA.03.03ZASK066B1.AY901969 C.ZA.03.03ZASK067B1.DQ275642
C.ZA.03.03ZASK073B1.AY901970 C.ZA.03.03ZASK076B1.AY901975 C.ZA.03.03ZASK092B1.AY878057 C.ZA.03.03ZASK097B1.AY878060 C.ZA.03.03ZASK098B1.AY878061
C.ZA.03.03ZASK103B1.DQ164106 C.ZA.03.03ZASK104B1.DQ396395 C.ZA.03.03ZASK107B1.DQ056410 C.ZA.03.03ZASK110B1.DQ056411
C.ZA.03.03ZASK117B1.DQ056408 C.ZA.03.03ZASK120B1.DQ011176 C.ZA.03.03ZASK215M6W.DQ275661 C.ZA.03.03ZASK232B1.DQ093589
C.ZA.03.03ZASK233B1.DQ351234 C.ZA.03.04ZASK082B1.AY901972 C.ZA.03.SK029B2.AY772691 C.ZA.03.SK040B1.AY703908 C.ZA.03.SK041B1.AY772693
C.ZA.03.SK043B1.AY772700 C.ZA.03.SK065B1.AY772694 C.ZA.03.SK091B1.AY772701 C.ZA.03.SK112B1.AY772695 C.ZA.03.SK116B1.AY772696
C.ZA.04.04ZAPS157MB1.DQ351232 C.ZA.04.04ZAPS160B1.DQ164107 C.ZA.04.04ZAPS161B1.DQ164115 C.ZA.04.04ZAPS168B1.DQ164121
C.ZA.04.04ZAPS172MB1.DQ275659 C.ZA.04.04ZAPS177MB1.DQ396394 C.ZA.04.04ZAPS188B1.DQ164122 C.ZA.04.04ZAPS189B1.DQ164117
C.ZA.04.04ZAPS190B1.DQ093602 C.ZA.04.04ZAPS195B1.DQ164118 C.ZA.04.04ZAPS197MB1.DQ369997 C.ZA.04.04ZAPS198MB1.DQ445637
C.ZA.04.04ZAPS199B1.DQ164110 C.ZA.04.04ZAPS202B1.DQ093598 C.ZA.04.04ZAPS205B1.DQ093599 C.ZA.04.04ZAPS206B1.DQ164126

C.ZA.04.04ZAPS211B1.DQ164112 C.ZA.04.04ZAPS214B1.DQ093588 C.ZA.04.04ZAPS216B1.DQ093600 C.ZA.04.04ZAPS217B1.DQ164119
C.ZA.04.04ZASK031B2.AY901974 C.ZA.04.04ZASK083B2.DQ093585 C.ZA.04.04ZASK127B1.DQ011172 C.ZA.04.04ZASK128B1.DQ093586
C.ZA.04.04ZASK131B1.DQ056412 C.ZA.04.04ZASK132B1.DQ093594 C.ZA.04.04ZASK135B1.DQ011166 C.ZA.04.04ZASK136B1.DQ011177
C.ZA.04.04ZASK139B1.AY878072 C.ZA.04.04ZASK142B1.DQ011167 C.ZA.04.04ZASK145B1.AY901976 C.ZA.04.04ZASK146.AY772699 C.ZA.04.04ZASK147B1.DQ164114
C.ZA.04.04ZASK150B1.DQ011178 C.ZA.04.04ZASK154B1.AY878071 C.ZA.04.04ZASK159B1.DQ011179 C.ZA.04.04ZASK161B1.DQ011170
C.ZA.04.04ZASK167B1.DQ164127 C.ZA.04.04ZASK170B1.DQ093595 C.ZA.04.04ZASK173B1.DQ093604 C.ZA.04.04ZASK175B1.DQ164129
C.ZA.04.04ZASK176B1.DQ056416 C.ZA.04.04ZASK180B1.AY878059 C.ZA.04.04ZASK181B1.AY878062 C.ZA.04.04ZASK182B1.AY878054 C.ZA.04.04ZASK184B1.DQ056418
C.ZA.04.04ZASK190B1.DQ056409 C.ZA.04.04ZASK192B1.DQ396382 C.ZA.04.04ZASK193B1.DQ396396 C.ZA.04.04ZASK196B1.DQ056413
C.ZA.04.04ZASK201B1.DQ396397 C.ZA.04.04ZASK202B1.DQ011180 C.ZA.04.04ZASK206B1.DQ056415 C.ZA.04.04ZASK217B1.DQ056417
C.ZA.04.04ZASK234B1.DQ093605 C.ZA.04.CAP30_5w_F4.GQ999973 C.ZA.04.CAP61_8w_F3.GQ999975 C.ZA.04.SK134B1.AY703909
C.ZA.04.SK143B1.AY703910 C.ZA.04.SK144B1.AY703911 C.ZA.05.05ZASK244B1.DQ369992 C.ZA.05.05ZASK245B1.DQ369982
C.ZA.05.05ZASK247B1.DQ369994 C.ZA.05.CAP206_8w_F1.GQ999982 C.ZA.05.CAP210_5w.GQ999983 C.ZA.05.CAP228_8w_F2.GQ999984
C.ZA.05.CAP229_7w.GQ999985 C.ZA.05.CAP244_8w_F1.GQ999986 C.ZA.05.CAP248_9w.GQ999987 C.ZA.05.CAP256_6w.GQ999989 C.ZA.05.CAP257_7w_F1.GQ999990
C.ZA.05.CAP63_5w_F4.GQ999976 C.ZA.05.CAP65_6w.GQ999977 C.ZA.05.CAP8_3w_F2.GQ999972 C.ZA.05.CAP85_5w_F1.GQ999979
C.ZA.05.CAP88_5w_F2.GQ999980 C.ZA.07.704010042_CH042_TF.KC156123 C.ZA.07.705010067_CH067_TF.KC156125 C.ZA.07.705010162_CH162.mo6.KC156115
C.ZA.07.705010198_CH198_TF.KC156130 C.ZA.07.706010164_CH164_TF.KC156127 C.ZA.07.BP00005_RH03.JN687704 C.ZA.07.BP00008_RH01.JN687706
C.ZA.07.BP00009_SUP_RH01.JN687717 C.ZA.07.BP00035_RH01.JN687736 C.ZA.07.DEMC07ZA011.JX140664 C.ZA.08.705010185_CH185.mo6.KC156116
C.ZA.08.BP00012_RH01.JN687724 C.ZA.08.BP00014_RH01.JN687725 C.ZA.08.BP00018_RH01.JN687728 C.ZA.08.BP00023_RH02.JN687730
C.ZA.08.BP00025_RH02.JN687731 C.ZA.08.DEMC08ZA011.JX140666 C.ZA.09.DEMC09ZA008.JX140667 C.ZA.09.DEMC09ZA009.JX140668
C.ZA.10.DEMC10ZA001.JX140669 C.ZA.90.pZAC_R3714.JN188292 C.ZA.97.97ZA003.AY118165 C.ZA.97.97ZA009.AY118166 C.ZA.97.97ZA012.AF286227
C.ZA.98.98ZA502.AY158534 C.ZA.98.98ZA528.AY158535 C.ZA.98.CTSC2.AY043176 C.ZA.98.DU151.AY043173 C.ZA.98.TV012.AY162225
C.ZA.99.99ZALT21.EU293446 C.ZA.99.99ZALT39.EU293447 C.ZA.99.99ZALT4.EU293444 C.ZA.99.99ZALT42.EU293448 C.ZA.99.99ZALT45.EU293449
C.ZA.99.99ZALT46.EU293450 C.ZA.99.99ZALT5.EU293445 C.ZA.99.99ZATM10.AY228556 C.ZA.99.DU179.AY043174 C.ZA.99.ZASW7.AF411966
C.ZM.02.02ZM114.AB254146 C.ZM.02.02ZM115.AB254148 C.ZM.02.02ZMBC.AB254149 C.ZM.02.02ZMDB.AB254153 C.ZM.02.02ZMJM.AB254156
C.ZM.03.ZM246F_flA10.FJ496186 C.ZM.03.ZM247F_flA1.FJ496195 C.ZM.03.ZM249M_flC1.FJ496209C.ZM.11.DEMC11ZM006.KF716467
C.ZM.96.96ZM651.AF286224 C.ZM.96.96ZM751.AF286225 C.ZM.x.HIV1084i.AY805330

Subtype D (n=52)

D.CD.84.84ZR085.U88822 D.CD.85.Z2Z6_Z2_CDC_Z34.M22639 D.CM.01.01CM_0009BBY.AY371155 D.CM.01.01CM_0175BA.AY371156
D.CM.01.01CM_4412HAL.AY371157 D.CM.10.DEMD10CM009.JX140670 D.CY.06.CY163.FJ388945 D.KE.01.NKU3006.AF457090
D.KE.11.DEMD11KE003.KF716476 D.KE.97.ML415_2.AY322189 D.KR.04.04KBH8.DQ054367 D.SN.90.SE365A2.L22945 D.TD.99.MN011.AJ488926
D.TD.99.MN012.AJ488927 D.TZ.01.A280.AY253311 D.UG.05.p190049.JX236668 D.UG.07.p191647.JX236670 D.UG.07.p191882.JX236673
D.UG.07.pSC191727.JX236679 D.UG.08.p191859.JX236672 D.UG.10.DEMD10UG004.KF716479 D.UG.11.DEMD11UG003.KF716480
D.UG.91.UG270.AB485650 D.UG.93.93UG_065.AY713418 D.UG.94.94UG114.U88824 D.UG.98.98UG57128.AF484502 D.UG.98.98UG57130.AF484504
D.UG.98.98UG57131.AF484505 D.UG.98.98UG57132.AF484506 D.UG.98.98UG57140.AF484511 D.UG.98.98UG57143.AF484514 D.UG.98.98UG57146.AF484513
D.UG.98.99UGJ32228.AF484516 D.UG.99.99UGA03349.AF484518 D.UG.99.99UGA08483.AY304496 D.UG.99.99UGB25647.AF484481 D.UG.99.99UGB32394.AF484483
D.UG.99.99UGD23550.AF484485 D.UG.99.99UGD26830.AF484486 D.UG.99.99UGE08364.AF484487 D.UG.99.99UGE13613.AF484515 D.UG.99.99UGF03726.AF484519
D.UG.99.99UGF05734.AF484490 D.UG.99.99UGG10555.AF484494 D.UG.99.99UGG35093.AF484495 D.UG.99.99UGK09259.AF484498 D.UG.99.99UGK09958.AF484499
D.US.07.HIV_US_BID_V3121_2007.JQ403079 D.YE.01.01YE386.AY795903 D.YE.02.02YE516.AY795907 D.ZA.85.R286.AY773340 D.ZA.90.R1.EF633445

Subtype F (n=27)

F1.AO.06.AO_06_ANG125.FJ900269 F1.AO.06.AO_06_ANG32.FJ900266 F1.AO.06.AO_06_ANG40.FJ900267 F1.AO.06.AO_06_ANG58.FJ900268
F1.AR.02.ARE933.DQ189088 F1.BE.93.VI850.AF077336 F1.BR.02.02BR082.FJ771006 F1.BR.02.02BR170.FJ771007 F1.BR.06.06BR564.FJ771008
F1.BR.06.06BR579.FJ771009 F1.BR.07.07BR844.FJ771010 F1.BR.89.BZ163.AY173958 F1.BR.93.93BR020_1.AF005494 F1.CY.08.CY222.JF683771
F1.DE.x.MVP_30846.EU446022 F1.ES.02.ES_X845_4.FJ670516 F1.ES.x.X1093_2.DQ979025 F1.ES.x.X1670.DQ979024 F1.FI.93.FIN9363.AF075703
F1.FR.96.96FR_MP411.AJ249238 F1.RO.96.BCI_R07.AB485658 F2.CM.02.02CM_0016BBY.AY371158 F2.CM.10.DEMF210CM001.JX140672

F2.CM.10.DEMF210CM007.JX140673 F2.CM.95.95CM_MP255.AJ249236 F2.CM.95.95CM_MP257.AJ249237 F2.CM.97.CM53657.AF377956
Subtype G (n=30)
G.CM.01.01CM_4049HAN.AY371121 G.CM.04.178_15.FJ389363 G.CM.04.944_5.FJ389366 G.CM.10.DEMG10CM008.JX140676 G.CN.06.sh52.HM067749 G.CN.08.GX_2084_08.JN106043 G.CU.99.Cu74.AY586547 G.CU.99.Cu85.AY586548 G.ES.00.X558.AF423760 G.ES.05.P962.EU786670 G.ES.08.P1981_2.FJ670530 G.ES.09.X2634_2.GU362882 G.GH.03.03GH175G.AB287004 G.KE.09.DEMG09KE001.KF716477 G.KE.93.HH8793_12_1.AF061641 G.NG.01.01NGPL0669.DQ168576 G.NG.01.01NGPL0674.DQ168575 G.NG.01.01NGPL0760.DQ168579 G.NG.01.PL0567.DQ168573 G.NG.08.08NG_SC13.JN248582 G.NG.09.09NG_SC21.JN248584 G.NG.09.09NG_SC26.JN248586 G.NG.09.09NG_SC31.JN248591 G.NG.09.09NG_SC62.JN248593 G.NG.92.92NG083_JV10832.U88826 G.PT.x.PT2695.AY612637 G.PT.x.PT3037.FR846408 G.PT.x.PT3306.FR846409 G.PT.x.PT988.FR846410 G.SE.93.SE6165_G6165.AF061642
Subtype H (n=4)
H.BE.93.VI991.AF190127 H.BE.93.VI997.AF190128 H.CF.90.056.AF005496 H.GB.00.00GBAC4001.FJ711703
Subtype J (n=2)
J.CD.97.J_97DC_KTB147.EF614151 J.SE.94.SE9173_7022.AF082395
Subtype K (n=2)
K.CD.97.97ZR_EQTB11.AJ249235 K.CM.96.96CM_MP535.AJ24923

Appendix Table 2a: Distribution of HIV-1 subtypes B (n=493) and C (n=280) of sequences by continent of origin (Data set 2)

Continent	B	C	Total
Asia	80	9	**89 (11.5%)**
Africa	2	254	**256 (33.1%)**
Europe	94	8	**102 (13.2%)**
South America	78	7	**85 (11.0%)**
North America	228	2	**230 (29.8%)**
Oceania	11	-	**11 (1.4%)**
Total	**493 (63.8%)**	**280 (36.2%)**	**773 (100%)**

Appendix Table 2b: Accession numbers of the Full-length Tat in subtypes B and C obtained from the Los Alamos National Laboratory (LANL) HIV sequence (Data set 2)

Subtype B (n=493)	Subtype C (n=298)
B.AR.00.ARMS008.AY037269	C.AR.01.ARG4006.AY56317001
B.AR.02.02AR114146.DQ383746	C.BR.02.02BR2022.JN69243402
B.AR.03.03AR137681.DQ383748	C.BR.04.04BR021.AY72752304
B.AR.03.03AR138910.DQ383749	C.BR.04.04BR038.AY72752404
B.AR.04.04AR143170.DQ383750	C.BR.04.04BR073.AY72752504
B.AR.04.04AR151263.DQ383751	C.BR.98.98BR004.AF28622898
B.AR.98.ARCH054.AY037268	C.BW.00.00BW07621.AF44308800
B.AR.99.ARMA132.AY037282	C.BW.00.00BW076820.AF44308900
B.AU.03.PS2008_Day182.DQ676875	C.BW.00.00BW087421.AF44309000
B.AU.04.MS2004_37_016.EF178314	C.BW.00.00BW147127.AF44309100
B.AU.04.MS2004_37_122.EF178420	C.BW.00.00BW16162.AF44309200
B.AU.04.MS2004_37_129.EF178427	C.BW.00.00BW1686.AF44309300
B.AU.04.Phcsffull04.AY818644	C.BW.00.00BW17593.AF44309400
B.AU.87.MBC925.AF042101	C.BW.00.00BW17956.AF44309700
B.AU.93.MBC18_MBCC18.AF042102	C.BW.00.00BW18595.AF44309900
B.AU.95.MBCC54.AF042103	C.BW.00.00BW18802.AF44310000
B.AU.96.MBCC98.AF042104	C.BW.00.00BW192113.AF44310100
B.AU.96.MBCD36.AF042105	C.BW.00.00BW20872.AF44310400
B.AU.99.1181.AF538302	C.BW.00.00BW2127214.AF44310500
B.BO.09.DEMB09BO001.JX140656	C.BW.00.00BW22767.AF44310700
B.BO.99.BOL0122.AY037270	C.BW.00.00BW38193.AF44310800
B.BR.02.02BR002.DQ358805	C.BW.00.00BW38428.AF44310900
B.BR.02.02BR008.DQ358808	C.BW.00.00BW38713.AF44311000
B.BR.02.02BR011.DQ358809	C.BW.00.00BW3876_9.AF44311100
B.BR.02.02BR2025.JN692435	C.BW.00.00BW3886_8.AF44311200
B.BR.02.02BR2033.JN692440	C.BW.00.00BW3891_6.AF44311300
B.BR.02.02BR2041.JN692443	C.BW.00.00BW3970_2.AF44311400
B.BR.02.02BR2042.JN692444	C.BW.00.00BW5031_1.AF44311500
B.BR.02.05BR1094.JN692470	C.BW.96.96BW0402.AF11096296
B.BR.03.03BR1020.JN692445	C.BW.96.96BW0504.AF11096896
B.BR.03.03BR1046.JN692447	C.BW.96.96BW06K18.AF29003096
B.BR.03.BREPM1032.EF637051	C.BW.96.96BW1104.AF11096996
B.BR.03.BREPM1033.EF637050	C.BW.96.96BW1210.AF11097296
B.BR.03.BREPM1035.EF637049	C.BW.96.96BW15B03.AF11097396
B.BR.03.BREPM1040.EF637047	C.BW.96.96BW16B01.AF11097696
B.BR.03.BREPM2012.EF637046	C.BW.96.96BW17.AF11097996
B.BR.04.04BR1049.JN69245104	C.BW.96.96BWM032.AF44307596
B.BR.04.04BR1051.JN69245204	C.BW.96.96BWMO1_5.AF44307496

B.BR.04.04BR1054.JN69245304	C.BW.98.98BWMC122.AF44307698
B.BR.04.04BR1057.JN69245504	C.BW.98.98BWMC134.AF44307798
B.BR.04.04BR1068.JN69245704	C.BW.98.98BWMC14A3.AF44307898
B.BR.04.BREPM1066.FJ19509004	C.BW.98.98BWMO1410.AF44307998
B.BR.04.BREPM1070.FJ19508604	C.BW.99.99BW393212.AF44308399
B.BR.05.05BR1074.JN69245905	C.BW.99.99BW4745.AF44308599
B.BR.05.05BR1078.JN69246105	C.BW.99.99BW47547.AF44308699
B.BR.05.05BR1080.JN69246305	C.BW.99.99BWMC168.AF44308799
B.BR.05.05BR1082.JN69246505	C.CY.05.CY040.FJ38890105
B.BR.05.05BR1095.JN69247105	C.CY.05.CY069.FJ38891305
B.BR.05.05BR1104.JN69247405	C.CY.06.CY176.FJ38895206
B.BR.05.05BR1107.JN69247505	C.CY.07.CY187.JF68374007
B.BR.05.BREPM1084.FJ19508805	C.CY.07.CY203.JF68375507
B.BR.05.BREPM1093.FJ19508905	C.CY.08.CY219.JF68376808
B.BR.06.06BR1115.JN69247906	C.DJ.91.DJ373A.L2306591
B.BR.06.06BR1119.JN69248006	C.DK.01.CTL_015.EF51471301
B.BR.07.BP00047_RH01.JN68773907	C.ES.08.X2363_2.EU78668108
B.CA.06.502_1799_FL02.JF32042706	C.ET.02.02ET_288.AY71341702
B.CA.07.502_1191_03.JF32042407	C.GE.03.03GEMZ033.DQ20794103
B.CA.96.WC10P_1.AY31404496	C.IL.98.98IS002.AF28623398
B.CA.98.CANA3FULL.AY77955098	C.IL.99.99ET1.AY25582399
B.CA.98.CANC10FULL.AY77955798	C.IN.03.D24.EF46924303
B.CH.00.HIV_CH_BID_V3529_2000.JQ40302200	C.IN.93.93IN999.AF06715493
B.CH.00.HIV_CH_BID_V3530_2000.JQ40302300	C.IN.94.94IN476.AF28622394
B.CH.01.HIV_CH_BID_V3531_2001.JQ40302401	C.IN.98.98IN012.AF28623198
B.CH.01.HIV_CH_BID_V3533_2001.JQ40302501	C.IN.98.98IN022.AF28623298
B.CH.01.HIV_CH_BID_V3534_2001.JQ40302601	C.KE.00.KER2010.AF45705400
B.CH.02.HIV_CH_BID_V3511_2002.JQ40301902	C.KE.91.KNH1268.AY94573891
B.CH.02.HIV_CH_BID_V3527_2002.JQ40302102	C.MW.08.702010432.e25.JQ77924208
B.CH.02.HIV_CH_BID_V4424_2002.JQ40304402	C.MW.09.703010256.e8.JQ77909609
B.CH.03.HIV_CH_BID_V4421_2003.JQ40304303	C.MW.93.93MW_965.AY71341393
B.CH.03.HIV_CH_BID_V4470_2003.JQ40304503	C.TZ.01.A125.AY25330401
B.CH.04.HIV_CH_BID_V4408_2004.JQ40304204	C.TZ.01.A207.AY25330701
B.CN.01.CNHN24.AY18090501	C.TZ.01.A246.AY25330801
B.CN.02.02HNsq4.DQ00790202	C.TZ.01.A301.AY25331201
B.CN.03.SHXDC0081.JF93249203	C.TZ.01.A388.AY25331701
B.CN.05.05CNHB_hp3.DQ99088005	C.TZ.01.A93.AY25330301
B.CN.06.CC056.JF93248206	C.TZ.01.BD16_10.AY25332001
B.CN.07.AH070011.JF93246807	C.TZ.01.BD22_11.AY25332101
B.CN.07.AH070014.JF93246907	C.TZ.01.BD9_11.AY25332201
B.CN.07.AH070018.JF93247107	C.TZ.02.CO178.AY73455602
B.CN.07.BJ070030.JF93247307	C.TZ.02.CO3305.AY73455802
B.CN.07.CBJC261.JF93247407	C.TZ.02.CO6130.AY73455902

B.CN.07.CBJC394.JF93247607	C.TZ.02.CO6721.AY73456002
B.CN.07.CBJC396.JF93247707	C.TZ.08.707010457.e11.JQ77912908
B.CN.07.FJ070016.JF93248307	C.TZ.97.97TZ05.AF36187597
B.CN.07.GS070017.JF93248407	C.TZ.98.98TZ013.AF28623498
B.CN.07.GZ070030.JF93248607	C.TZ.98.98TZ017.AF28623598
B.CN.07.HB070006.JF93248707	C.US.98.98US_MSC3018.AY44480098
B.CN.07.HB070022.JF93248807	C.YE.02.02YE511.AY79590602
B.CN.07.hb070025.JF93249907	C.ZA.00.1119MB.AY46322900
B.CN.07.hen1345.JF93250007	C.ZA.00.1134MB.AY46321700
B.CN.07.JL070038.JF93249007	C.ZA.00.1162MB.AY46322400
B.CN.07.JS070389.JF93249107	C.ZA.00.1165MB.AY46323000
B.CN.07.SX070080.JF93249307	C.ZA.00.1168MB.AY46323100
B.CN.08.cbjc468.JF93249808	C.ZA.00.1170MB.AY46322500
B.CN.08.CBJC476.JF93247808	C.ZA.00.1171MB.AY46323200
B.CN.08.CBJC489.JF93247908	C.ZA.00.1192M3M.AY46321900
B.CN.08.CBJC503.JF93248108	C.ZA.00.1195MB.AY46322000
B.CN.09.09LNA014.JX96059709	C.ZA.00.1210MB.AY46322100
B.CN.09.09LNA439.JX96059809	C.ZA.00.1214MB.AY46323600
B.CN.09.1121.HQ21555609	C.ZA.00.1225MB.AY46322700
B.CN.09.YN09P0014.JF93249409	C.ZA.01.2134MB.AY46323701
B.CN.09.ZK042.JF93249709	C.ZA.02.02ZAPS006MB1.DQ35122002
B.CN.10.DEMB10CN002.JX14065810	C.ZA.02.02ZAPS015MB1.DQ36999502
B.CN.98.YN9802.JF93249598	C.ZA.03.03ZAPS017MB1.DQ35122403
B.CN.98.YN9838.JF93249698	C.ZA.03.03ZAPS020MB1.DQ27565303
B.CO.01.PCM013.AY56123701	C.ZA.03.03ZAPS023MB1.DQ35122503
B.CO.01.PCM034.AY56123801	C.ZA.03.03ZAPS024MB1.DQ39636703
B.CU.99.Cu19.AY58654299	C.ZA.03.03ZAPS026MB1.DQ36998503
B.CU.99.Cu43.AY58654399	C.ZA.03.03ZAPS027MB1.DQ35122303
B.CY.05.CY020.FJ38889105	C.ZA.03.03ZAPS030MB1.DQ36999603
B.CY.05.CY030.FJ38895605	C.ZA.03.03ZAPS032MB1.DQ44563303
B.CY.05.CY036.FJ38895905	C.ZA.03.03ZAPS044MB1.DQ39638403
B.CY.05.CY065.FJ38891005	C.ZA.03.03ZAPS046MB1.DQ36998403
B.CY.05.CY067.FJ38891105	C.ZA.03.03ZAPS049MB1.DQ36998603
B.CY.05.CY070.FJ38896005	C.ZA.03.03ZAPS055MB1.DQ39637303
B.CY.05.CY088.FJ38891905	C.ZA.03.03ZAPS056MB1.DQ39637403
B.CY.05.CY096.FJ38896305	C.ZA.03.03ZAPS057MB2.DQ36998903
B.CY.05.CY113.FJ38893005	C.ZA.03.03ZAPS059MB2.DQ44563403
B.CY.05.CY122.FJ40348205	C.ZA.03.03ZAPS063MB1.DQ39638803
B.CY.05.CY130.FJ38893405	C.ZA.03.03ZAPS066MB2.DQ39637503
B.CY.05.CY131.FJ38893505	C.ZA.03.03ZAPS074MB2.DQ35122803
B.CY.05.CY149.FJ38894005	C.ZA.03.03ZAPS081MB1.DQ35121903
B.CY.06.CY168.FJ38894906	C.ZA.03.03ZAPS083MB1.DQ35122903
B.CY.06.CY180.FJ38895506	C.ZA.03.03ZAPS086MB1.DQ27565403

B.CY.07.CY189.JF68374207	C.ZA.03.03ZAPS094MB1.DQ39637703
B.CY.07.CY197.JF68374907	C.ZA.03.03ZAPS097MB1.DQ35123003
B.CY.07.CY198.JF68375007	C.ZA.03.03ZAPS103MB2.DQ27565603
B.CY.07.CY199.JF68375107	C.ZA.03.03ZAPS105MB2.DQ44563203
B.CY.07.CY201.JF68375307	C.ZA.03.03ZAPS108MB1.DQ39637803
B.CY.07.CY202.JF68375407	C.ZA.03.03ZAPS112MB2.DQ39638603
B.CY.07.CY204.JF68375607	C.ZA.03.03ZAPS113MB2.DQ39636503
B.CY.07.CY216.JF68376507	C.ZA.03.03ZAPS118MB1.DQ39636803
B.CY.08.CY220.JF68376908	C.ZA.03.03ZAPS124MB1.DQ36997603
B.CY.08.CY224.JF68377308	C.ZA.03.03ZAPS131MB1.DQ39638003
B.CY.08.CY226.JF68377508	C.ZA.03.03ZAPS136MB1.DQ35123103
B.CY.08.CY229.JF68377808	C.ZA.03.03ZAPS140MB1.DQ36998103
B.CY.08.CY232.JF68378108	C.ZA.03.03ZAPS143MB1.DQ39639103
B.CY.08.CY237.JF68378408	C.ZA.03.03ZAPS151MB1.DQ39639203
B.CY.08.CY238.JF68378508	C.ZA.03.03ZASK010B2.DQ16410403
B.CY.09.CY241.JF68378709	C.ZA.03.03ZASK011B2.AY90196503
B.CY.09.CY242.JF68378809	C.ZA.03.03ZASK016MB2.DQ35123303
B.CY.09.CY244.JF68379009	C.ZA.03.03ZASK020B2.AY87806403
B.CY.09.CY250.JF68379309	C.ZA.03.03ZASK034B1.AY87806503
B.CY.09.CY253.JF68379609	C.ZA.03.03ZASK039B2.AY87806803
B.CY.09.CY254.JF68379709	C.ZA.03.03ZASK062B1.DQ16411303
B.CY.09.CY262.JF68380409	C.ZA.03.03ZASK067B1.DQ27564203
B.CY.09.CY263.JF68380509	C.ZA.03.03ZASK073B1.AY90197003
B.CY.09.CY266.JF68380709	C.ZA.03.03ZASK076B1.AY90197503
B.DE.03.HIV_DE_BID_V3307_2003.JQ40304803	C.ZA.03.03ZASK092B1.AY87805703
B.DE.04.9213_d0.JQ41615804	C.ZA.03.03ZASK097B1.AY87806003
B.DE.04.HIV_DE_BID_V4131_2004.JQ40303704	C.ZA.03.03ZASK098B1.AY87806103
B.DE.86.HAN.U4314186	C.ZA.03.03ZASK103B1.DQ16410603
B.DK.01.CTL_017.EF51470501	C.ZA.03.03ZASK104B1.DQ39639503
B.DK.01.CTL_018.EF51470601	C.ZA.03.03ZASK107B1.DQ05641003
B.DK.01.CTL_041.EF51471101	C.ZA.03.03ZASK117B1.DQ05640803
B.DK.04.PMVL_012.EF51469904	C.ZA.03.03ZASK120B1.DQ01117603
B.DK.04.PMVL_013.EF51470004	C.ZA.03.03ZASK215M6W.DQ27566103
B.DK.04.PMVL_018.EF51469704	C.ZA.03.03ZASK232B1.DQ09358903
B.DK.04.PMVL_039.EF51470304	C.ZA.03.03ZASK233B1.DQ35123403
B.DK.07.PMVL_011.FJ69479007	C.ZA.03.04ZAPS194MB1.DQ27565003
B.EC.89.EC003.AY17395989	C.ZA.03.SK029B2.AY77269103
B.EC.89.EC102.AY17396089	C.ZA.03.SK040B1.AY70390803
B.ES.05.X1890.EU78667205	C.ZA.03.SK041B1.AY77269303
B.ES.06.X1998.EU78667606	C.ZA.03.SK065B1.AY77269403
B.ES.06.X2102.EU78667706	C.ZA.03.SK091B1.AY77270103
B.ES.07.X2210_3.EU78667907	C.ZA.03.SK116B1.AY77269603
B.ES.07.X2231.EU78668007	C.ZA.04.04ZAPS157MB1.DQ35123204

B.ES.08.ES_X2515_3.GQ37298808	C.ZA.04.04ZAPS160B1.DQ16410704
B.ES.08.ES_X2556_3.GQ37299008	C.ZA.04.04ZAPS161B1.DQ16411504
B.ES.08.P2008.FJ67053108	C.ZA.04.04ZAPS168B1.DQ16412104
B.ES.08.P2021_3.FJ85362008	C.ZA.04.04ZAPS172MB1.DQ27565904
B.ES.08.X2425_2.FJ67052508	C.ZA.04.04ZAPS177MB1.DQ39639404
B.ES.08.X2510_2.FJ85362208	C.ZA.04.04ZAPS188B1.DQ16412204
B.ES.09.P2149_3.GU36288109	C.ZA.04.04ZAPS189B1.DQ16411704
B.ES.09.X2689_2.GU36288509	C.ZA.04.04ZAPS190B1.DQ09360204
B.FR.08.DEMB08FR002.JX14065408	C.ZA.04.04ZAPS195B1.DQ16411804
B.GA.88.OYI_397.M2672788	C.ZA.04.04ZAPS197MB1.DQ36999704
B.GB.03.MM33d12p.HM58618703	C.ZA.04.04ZAPS198MB1.DQ44563704
B.GB.04.MM39d11p.HM58619304	C.ZA.04.04ZAPS199B1.DQ16411004
B.GB.05.MM43d368_GN1.HM58620905	C.ZA.04.04ZAPS202B1.DQ09359804
B.GB.05.MM45d22_GN1.HM58621005	C.ZA.04.04ZAPS206B1.DQ16412604
B.GB.83.CAM1.D1011283	C.ZA.04.04ZAPS211B1.DQ16411204
B.GB.86.GB8_46R.AJ27144586	C.ZA.04.04ZAPS214B1.DQ09358804
B.GE.03.03GEMZ010.DQ20794203	C.ZA.04.04ZASK031B2.AY90197404
B.HK.06.HK002.FJ46049906	C.ZA.04.04ZASK052MB2.DQ39636604
B.HT.05.05HT_129473.EU83960305	C.ZA.04.04ZASK083B2.DQ09358504
B.HT.05.05HT_129696.EU83960105	C.ZA.04.04ZASK127B1.DQ01117204
B.JM.05.05JM_KJ108.EU83960505	C.ZA.04.04ZASK128B1.DQ09358604
B.JM.09.09JM_PF09XN.HM03055909	C.ZA.04.04ZASK131B1.DQ05641204
B.JM.09.09JM_PF09YT.HM03056009	C.ZA.04.04ZASK132B1.DQ09359404
B.JM.09.09JM_PF0B8J.HM03056109	C.ZA.04.04ZASK135B1.DQ01116604
B.JM.09.09JM_PF0B97.HM03056209	C.ZA.04.04ZASK136B1.DQ01117704
B.JM.09.09JM_PF0B9L.HM03056509	C.ZA.04.04ZASK139B1.AY87807204
B.JM.09.09JM_PF0BB5.HM03056309	C.ZA.04.04ZASK145B1.AY90197604
B.JP.00.117.AB42855100	C.ZA.04.04ZASK147B1.DQ16411404
B.JP.00.DR2508.AB28958800	C.ZA.04.04ZASK150B1.DQ01117804
B.JP.00.DR2510.AB28737200	C.ZA.04.04ZASK154B1.AY87807104
B.JP.01.DR388.AB28959001	C.ZA.04.04ZASK159B1.DQ01117904
B.JP.024.AB42855302	C.ZA.04.04ZASK161B1.DQ01117004
B.JP.04.04JPDR6075B.AB22112604	C.ZA.04.04ZASK167B1.DQ16412704
B.JP.04.DR5913.AB48069604	C.ZA.04.04ZASK170B1.DQ09359504
B.JP.04.DR6175.AB48069404	C.ZA.04.04ZASK173B1.DQ09360404
B.JP.05.426.AB42855605	C.ZA.04.04ZASK175B1.DQ16412904
B.JP.05.DR6737.AB28736405	C.ZA.04.04ZASK176B1.DQ05641604
B.JP.05.DR7060.AB28736705	C.ZA.04.04ZASK180B1.AY87805904
B.JP.05.DR7065.AB28736805	C.ZA.04.04ZASK181B1.AY87806204
B.KR.02.02OSG1.JQ42943302	C.ZA.04.04ZASK182B1.AY87805404
B.KR.03.03HJY8.JQ31613103	C.ZA.04.04ZASK184B1.DQ05641804
B.KR.03.03KDE11.JQ31612803	C.ZA.04.04ZASK190B1.DQ05640904
B.KR.03.03KGS5.JQ31613203	C.ZA.04.04ZASK191B1.DQ36999304

B.KR.03.03YGS3.JQ31613503	C.ZA.04.04ZASK192B1.DQ39638204
B.KR.04.04KJin8_1955.DQ29519504	C.ZA.04.04ZASK196B1.DQ05641304
B.KR.04.04KJS8.JQ31613004	C.ZA.04.04ZASK201B1.DQ39639704
B.KR.04.04KMK5.JQ31612604	C.ZA.04.04ZASK202B1.DQ01118004
B.KR.04.04KYR8.DQ29519604	C.ZA.04.04ZASK204B1.DQ05641404
B.KR.05.05CSR3.DQ83738105	C.ZA.04.04ZASK206B1.DQ05641504
B.KR.05.05YJN2.JQ31613405	C.ZA.04.04ZASK217B1.DQ05641704
B.KR.07.07KYY4.JQ34141107	C.ZA.04.04ZASK234B1.DQ09360504
B.KR.97.WK.AF22450797	C.ZA.04.CAP30_5w_F4.GQ99997304
B.NL.99.671_99T12.AY42338199	C.ZA.04.CAP61_8w_F3.GQ99997504
B.PE.06.502_0491_wg5.JF32018306	C.ZA.04.SK133B1.AY77269804
B.PE.06.502_0648_FL02.JF32021506	C.ZA.04.SK134B1.AY70390904
B.PE.06.502_0841_FL04.JF32020806	C.ZA.04.SK143B1.AY70391004
B.PE.06.502_2794_FL05.JF32024406	C.ZA.04.SK144B1.AY70391104
B.PE.07.502_0525_wg5.JF32019107	C.ZA.05.05ZASK244B1.DQ36999205
B.PE.07.502_1047_wg5.JF32022607	C.ZA.05.05ZASK245B1.DQ36998205
B.PE.07.502_1399_wg4.JF32001307	C.ZA.05.05ZASK247B1.DQ36999405
B.PE.07.502_2254_FL6.JF32001807	C.ZA.05.CAP206_8w_F1.GQ99998205
B.PE.07.502_2349_wg2.JF32002807	C.ZA.05.CAP228_8w_F2.GQ99998405
B.PE.07.502_2622_wg1.JF32018907	C.ZA.05.CAP229_7w.GQ99998505
B.PE.07.502_2649_wg8.JF32001907	C.ZA.05.CAP239_5w_F1.GQ99999105
B.PL.08.BP00074_RH02.JN68777408	C.ZA.05.CAP244_8w_F1.GQ99998605
B.PY.02.02PY_PSP0019.JN25189602	C.ZA.05.CAP248_9w.GQ99998705
B.PY.02.02PY_PSP0090.JN25190102	C.ZA.05.CAP255_8w_F1.GQ99998805
B.PY.03.03PY_PSP0115.JN25190603	C.ZA.05.CAP256_6w.GQ99998905
B.RU.04.04RU128005.AY68254704	C.ZA.05.CAP257_7w_F1.GQ99999005
B.RU.04.04RU139089.AY75140704	C.ZA.05.CAP8_3w_F2.GQ99997205
B.RU.09.09RU4457.JX50070909	C.ZA.05.CAP88_5w_F2.GQ99998005
B.RU.10.10RU6629.JX50070710	C.ZA.07.BP00005_RH03.JN68770407
B.RU.11.11RU21n.JX50070811	C.ZA.07.BP00008_RH01.JN68770607
B.TH.04.04TH317223.JN24832104	C.ZA.07.BP00009_SUP_RH01.JN68771707
B.TH.04.04TH803686.JN24833304	C.ZA.07.BP00035_RH01.JN68773607
B.TH.04.04TH808998.JN24833504	C.ZA.07.DEMC07ZA011.JX14066407
B.TH.04.04TH821921.JN24833704	C.ZA.08.705010534.e18.JQ77917708
B.TH.05.05TH356764.JN24834405	C.ZA.08.BP00012_RH01.JN68772408
B.TH.05.05TH357801.JN24834605	C.ZA.08.BP00014_RH01.JN68772508
B.TH.05.05TH429730.JN24834705	C.ZA.08.BP00018_RH01.JN68772808
B.TH.05.05TH645189.JN24835305	C.ZA.08.BP00019_RH01.JN68772908
B.TH.05.05TH736580.JN24835405	C.ZA.08.BP00023_RH02.JN68773008
B.TH.06.AA010a_WG3.JX44680006	C.ZA.08.DEMC08ZA011.JX14066608
B.TH.06.AA011a08R.JX44681806	C.ZA.10.DEMC10ZA001.JX14066910
B.TH.06.AA093a_RH1.JX44779506	C.ZA.97.97ZA003.AY11816597
B.TH.07.AA040a_WG11.JX44715606	C.ZA.97.97ZA009.AY11816697

B.TH.08.AA115c03R.JX44810308	C.ZA.97.97ZA012.AF28622797
B.TH.08.MERLBDTRC10.JN86076908	C.ZA.98.98ZA445.AY15853398
B.TH.96.M145.DQ35411896	C.ZA.98.98ZA502.AY15853498
B.TH.96.M149.DQ35411996	C.ZA.98.98ZA528.AY15853598
B.TT.00.00TT_CRC50018.EU83960702	C.ZA.98.CTSC2.AY04317698
B.TW.94.TWCYS_LM49.AF08681794	C.ZA.98.DU151.AY04317398
B.UA.01.01UAKV167.DQ82336201	C.ZA.99.99ZALT21.EU29344699
B.UA.01.01UAKV252.DQ82336301	C.ZA.99.99ZALT39.EU29344799
B.UA.01.01UAKV259.DQ82336401	C.ZA.99.99ZALT4.EU29344499
B.US.00.ES1_20.EF36312300	C.ZA.99.99ZALT42.EU29344899
B.US.00.RHPA_TF1.JN94491700	C.ZA.99.99ZALT45.EU29344999
B.US.00.THRO_TF1.JN94493000	C.ZA.99.99ZALT46.EU29345099
B.US.00.WITO_TF1.JN94493800	C.ZA.99.99ZALT5.EU29344599
B.US.01.L805.FJ46973801	C.ZA.99.99ZATM10.AY22855699
B.US.01.REJO_TF1.JN94491101	C.ZA.99.DU179.AY04317499
B.US.01.TRJO_TF1.JN94493601	C.ZM.02.02ZM114.AB25414602
B.US.02.04013226_2_flH11.FJ49607802	C.ZM.02.02ZM115.AB25414802
B.US.02.CR0017Q.FJ46968702	C.ZM.02.02ZMDB.AB25415302
B.US.02.CR0058S.FJ46969302	C.ZM.02.02ZMJM.AB25415602
B.US.02.F706.FJ46972802	C.ZM.03.ZM246F_flA10.FJ49618603
B.US.02.F762P.FJ46973502	C.ZM.03.ZM247F_flA1.FJ49619503
B.US.02.F797.FJ46973702	C.ZM.03.ZM249M_flC1.FJ49620903
B.US.02.HIV_US_BID_V5249_2002.JQ40310202	C.ZM.96.96ZM651.AF28622496
B.US.02.HIV_US_BID_V5276_2002.JQ40310502	C.ZM.96.96ZM751.AF28622596
B.US.02.L8124P.FJ46974102	Ref.C.BR.92.BR025_d.U5295392
B.US.02.L861P.FJ46974802	Ref.C.IN.95.95IN21068.AF06715595
B.US.02.MTF.FJ46975102	Ref.C.ZA.04.04ZASK146.AY77269904
B.US.02.PRLS08.FJ46975702	
B.US.02.PRLS09.FJ46975802	
B.US.02.PRLS16.FJ46976002	
B.US.03.CR0154X.FJ46970103	
B.US.03.F7157.FJ46973103	
B.US.03.HIV_US_BID_V5279_2003.JQ40310603	
B.US.03.L8107.FJ46973903	
B.US.03.L8116.FJ46974003	
B.US.03.L8152.FJ46974203	
B.US.03.L8180.FJ46974303	
B.US.03.PRLS01.FJ46975303	
B.US.03.PRLS12.FJ46975903	
B.US.03.PRLS24.FJ46976403	
B.US.03.PSL024B10.GU56227203	
B.US.03.SH8233.FJ46977003	
B.US.04.014837G.FJ46968404	

B.US.04.CR0068P.FJ46969504
B.US.04.CR0080N.FJ46969704
B.US.04.CR0116Y.FJ46969804
B.US.04.CR0345Q.FJ46972204
B.US.04.ES10_53.EF36312704
B.US.04.ES8_43.EF36312604
B.US.04.F7165.FJ46973204
B.US.04.HDNDRPI034A4.GU56215504
B.US.04.HIV_US_BID_V5258_2004.JQ40310304
B.US.04.HIV_US_BID_V5261_2004.JQ40310404
B.US.04.HIV_US_BID_V5282_2004.JQ40310704
B.US.04.USPI71101EI7y04051pcWG2B9.JN02421004
B.US.04.USPI83747EI6y04121pcWG2B5.JN02410004
B.US.05.01144G.FJ46968205
B.US.05.012286G.FJ46968305
B.US.05.04013396_0_flE6.FJ49608105
B.US.05.05US_SAJ_NVS12.JF68985205
B.US.05.05US_SAJ_NVS16.JF68985405
B.US.05.05US_SAJ_NVS3.JF68985605
B.US.05.05US_SAJ_NVS5.JF68985705
B.US.05.05US_SAJ_NVS8.JF68985905
B.US.05.05US_SAJ_NVS9.JF68986005
B.US.05.502_0223_wg1.JF32005905
B.US.05.502_1400_FL02.JF32004305
B.US.05.502_1926_FL01.JF32036105
B.US.05.502_2008_FL04.JF32048405
B.US.05.502_2136_FL02.JF32018505
B.US.05.502_2495_wg02.JF32005405
B.US.05.CR0175S.FJ46970305
B.US.05.CR0307R.FJ46971705
B.US.05.CR0312W.FJ46971805
B.US.05.L8249.DQ88603705
B.US.05.L827.FJ46974705
B.US.05.USPI38417EI33y05051pcWG2B2.JN02436305
B.US.05.USPI88403EI14y05121pcWG2B7.JN02434405
B.US.06.06US_SAJ_C164_SC.JF68986206
B.US.06.06US_SAJ_C165_TJ.JF68986306
B.US.06.06US_SAJ_C168_LS.JF68986606
B.US.06.06US_SAJ_C169_JS.JF68986706
B.US.06.06US_SAJ_NVS22.JF68987006
B.US.06.06US_SAJ_NVS23.JF68987106
B.US.06.06US_SAJ_NVS27.JF68987206
B.US.06.06US_SAJ_NVS32.JF68987406

B.US.06.06US_SAJ_NVS35.JF68987506
B.US.06.06US_SAJ_NVS39.JF68987606
B.US.06.502_0053_wg06.JF32061506
B.US.06.502_0062_FL04.JF32061306
B.US.06.502_0176_FL06.JF32036306
B.US.06.502_0227_FL05.JF32003606
B.US.06.502_0309_wg13.JF32004806
B.US.06.502_0322_RH04.JF32030806
B.US.06.502_0341_FL05.JF32000306
B.US.06.502_0346_wg02.JF32009706
B.US.06.502_0572_FL06.JF32003806
B.US.06.502_0717_FL02.JF32052606
B.US.06.502_0762_RH05.JF32032906
B.US.06.502_0839_wg01.JF32026306
B.US.06.502_0897_RH05.JF32050206
B.US.06.502_0923_wg07.JF32016006
B.US.06.502_1055_FL01.JF32016906
B.US.06.502_1174_FL09.JF32005306
B.US.06.502_1512_FL01.JF32035606
B.US.06.502_1619_FL06.JF32012606
B.US.06.502_2437_RH01.JF32059206
B.US.06.502_2667_FL03.JF32014506
B.US.06.BP00054_RH01.JN68774906
B.US.06.CH40_TF1.JN94490506
B.US.06.CH58_TF1.JN94490706
B.US.06.CH77_TF1.JN94490906
B.US.06.CR0047U.FJ46969106
B.US.06.CR0078.FJ46969606
B.US.06.CR0131.FJ46970006
B.US.06.CR0215.FJ46970806
B.US.06.CR0222X.FJ46970906
B.US.06.CR0228Q.FJ46971006
B.US.06.CR0276Z.FJ46971406
B.US.06.CR0361T.FJ46972306
B.US.06.HIV_US_BID_V3047_2006.JQ40306806
B.US.06.HIV_US_BID_V4506_2006.JQ40309706
B.US.06.HTM385B7.GU56226606
B.US.07.07US_SAJ_C154.JF68987707
B.US.07.07US_SAJ_C161_H1.JF68988307
B.US.07.07US_SAJ_C163_H3.JF68988507
B.US.07.07US_SAJ_C166_MS.JF68988607
B.US.07.07US_SAJ_NVS48.JF68989007
B.US.07.07US_SAJ_NVS54.JF68989207

B.US.07.07US_SAJ_NVS55.JF68989307
B.US.07.502_0287_RH1.JF32037507
B.US.07.502_0364_wg2.JF32056307
B.US.07.502_0388_RH06.JF32031507
B.US.07.502_0823_05.JF32053007
B.US.07.502_0938_RH04.JF32063107
B.US.07.502_0965_RH01.JF32038507
B.US.07.502_1115_wg1.JF32004507
B.US.07.502_1368_RH02.JF32017307
B.US.07.502_1478_wg4.JF32015007
B.US.07.502_1500_RH01.JF32038707
B.US.07.502_1897_wg6.JF32018207
B.US.07.502_2000_RH03.JF32027907
B.US.07.502_2241_RH13.JF32053907
B.US.07.502_2289_05.JF32019707
B.US.07.502_2305_01.JF32057707
B.US.07.502_2586_RH04.JF32013107
B.US.07.BP00058_RH01.JN68775907
B.US.07.BP00064_RH01.JN68776207
B.US.07.BP00067_RH01.JN68776307
B.US.07.CR0027M.FJ46968907
B.US.07.CR0214.FJ46970707
B.US.07.CR0275.FJ46971307
B.US.07.CR0295S.FJ46971607
B.US.07.CR0339X.FJ46972107
B.US.07.HIV_US_BID_V3010_2007.JQ40305807
B.US.07.HIV_US_BID_V3020_2007.JQ40305907
B.US.07.HIV_US_BID_V3021_2007.JQ40306007
B.US.07.HIV_US_BID_V3036_2007.JQ40306407
B.US.07.HIV_US_BID_V3053_2007.JQ40307107
B.US.07.HIV_US_BID_V3115_2007.JQ40307507
B.US.07.HIV_US_BID_V3120_2007.JQ40307807
B.US.07.HIV_US_BID_V3515_2007.JQ40309307
B.US.07.HIV_US_BID_V4516_2007.JQ40309607
B.US.08.08US_SAJ_C203.JF68989508
B.US.08.08US_SAJ_C204.JF68989608
B.US.08.BP00061_RH01.JN68776008
B.US.08.HIV_US_BID_V3024_2008.JQ40306108
B.US.08.HIV_US_BID_V3027_2008.JQ40306208
B.US.08.HIV_US_BID_V3046_2008.JQ40306708
B.US.08.HIV_US_BID_V3048_2008.JQ40306908
B.US.08.HIV_US_BID_V3050_2008.JQ40307008
B.US.08.HIV_US_BID_V3114_2007.JQ40307408

B.US.08.HIV_US_BID_V3122_2008.JQ40308008
B.US.08.HIV_US_BID_V3128_2008.JQ40308208
B.US.08.HIV_US_BID_V4124_2008.JQ40303508
B.US.08.HIV_US_BID_V4388_2008.JQ40308308
B.US.08.HIV_US_BID_V4389_2008.JQ40308408
B.US.08.HIV_US_BID_V4390_2008.JQ40308508
B.US.08.HIV_US_BID_V4391_2008.JQ40308608
B.US.08.HIV_US_BID_V4392_2008.JQ40308708
B.US.08.HIV_US_BID_V4393_2008.JQ40308808
B.US.08.HIV_US_BID_V4394_2008.JQ40308908
B.US.08.HIV_US_BID_V4489_2008.JQ40309408
B.US.09.DEMB09US002.JX14065709
B.US.10.CP1.JN39736510
B.US.10.VC1.JN39736410
B.US.11.ES38.JN39736211
B.US.83.RF_HAT3.M1745183
B.US.84.MNCG_MN.M1744984
B.US.84.SC.M1745084
B.US.84.SF33.AY35227584
B.US.85.RJS4_15_ENVBB.M3757385
B.US.86.5127_86.AY83577486
B.US.86.AD87_ADA.AF00439486
B.US.86.YU_2.M9325886
B.US.87.5113_87.AY83575887
B.US.88.5160_88.AY83576388
B.US.88.WR27.AF28636588
B.US.89.P896_89_6.U3936289
B.US.90.US1.AY17395290
B.US.90.US2.AY17395390
B.US.90.US4.AY17395590
B.US.90.WEAU160_GHOSH.U2113590
B.US.91.5048_91.AY83576191
B.US.91.DH12_3.AF06914091
B.US.91.R3A.AY60857791
B.US.91.SUMA_TF1.JN94492891
B.US.92.92US657_1_301657_1.U0490892
B.US.94.5082_94.AY83577394
B.US.94.94US_33931N.AY71341094
B.US.95.5073_95.AY83576895
B.US.96.1027_05.AY33129196
B.US.96.1057_01.AY33129296
B.US.96.5155_96.AY83575396
B.US.96.USPI55751EI32y96071pcWG2B15.JN02442896

B.US.97.1006_08.AY33128497
B.US.97.1012_08.AY33128597
B.US.97.1018_06.AY33128997
B.US.97.Ac_06.AY24725197
B.US.97.ACTDM580208A15.GU56203397
B.US.97.CQLDR03A1.GU56205897
B.US.97.HTM319C1.GU56223697
B.US.97.ZP6248_07_3A1.JN40046997
B.US.98.1058_08.AY33129498
B.US.98.98USHVTN3605c9.AY56010898
B.US.98.98USHVTN941c1.AY56011098
B.US.98.HDNDRPI001B10.GU56208098
B.US.98.WC3_0498_4.EF17521298
B.US.99.169_1999.JN59916599
B.US.99.HIV_US_BID_V5239_1999.JQ40310099
B.US.99.PRB959_03.AY33129699
B.UY.01.01UYTRA1092.AY78112601
B.UY.01.01UYTRA1179.AY78112701
B.UY.02.02UY_TSU1290.JN23595802
B.VE.10.DEMB10VE001.JX14065910
B.YE.02.02YE508.AY79590502
B.ZA.03.03ZAPS045MB2.DQ39639803
Ref.B.NL.00.671_00T36.AY42338700
Ref.B.TH.90.BK132.AY17395190
Ref.B.US.98.1058_11.AY33129598

Appendix Table 3a: Distribution of HIV-1 subtypes B (n=135) and C (n=53) of sequences by continent of origin (Data set 3)

Continent	B	C	Total
Asia	29	8	**37 (19.7%)**
Africa	2	32	**34 (18.1%)**
Europe	36	6	**42 (22.3%)**
South America	24	5	**29 (15.4%)**
North America	37	2	**39 (20.7%)**
Oceania	7	-	**7 (3.7%)**
Total	**135 (71.8%)**	**53 (28.2%)**	**188 (100%)**

Appendix Table 3b: Accession numbers of the Full-length Tat in subtypes B and C obtained from the Los Alamos National Laboratory (LANL) HIV sequence (Data set 3)

Subtype B (n=135)	Subtype C (n=53)
B.AR.00.ARMS008.AY037269.2000	C.AR.01.ARG4006.AY563170.2001
B.AR.02.02AR114146.DQ383746.2002	C.BR.02.02BR2022.JN692434.2002
B.AR.03.03AR137681.DQ383748.2003	C.BR.04.04BR021.AY727523.2004
B.AR.04.04AR143170.DQ383750.2004	C.BR.98.98BR004.AF286228.1998
B.AR.98.ARCH054.AY037268.1998	C.BW.00.00BW07621.AF443088.2000
B.AR.99.ARMA132.AY037282.1999	C.BW.96.96BW0402.AF110962.1996
B.AU.03.PS2008_Day182.DQ676875.2003	C.BW.98.98BWMC122.AF443076.1998
B.AU.04.MS2004_37_016.EF178314.2004	C.BW.99.99BW393212.AF443083.1999
B.AU.87.MBC925.AF042101.1987	C.CY.05.CY040.FJ388901.2005
B.AU.93.MBC18_MBCC18.AF042102.1993	C.CY.06.CY176.FJ388952.2006
B.AU.95.MBCC54.AF042103.1995	C.CY.07.CY203.JF683755.2007
B.AU.96.MBCC98.AF042104.1996	C.CY.08.CY219.JF683768.2008
B.AU.99.1181.AF538302.1999	C.DJ.91.DJ373A.L23065.1991
B.BO.09.DEMB09BO001.JX140656.2009	C.DK.01.CTL_015.EF514713.2001
B.BO.99.BOL0122.AY037270.1999	C.ES.08.X2363_2.EU786681.2008
B.BR.02.02BR002.DQ358805.2002	C.ET.02.02ET_288.AY713417.2002
B.BR.03.03BR1046.JN692447.2003	C.GE.03.03GEMZ033.DQ207941.2003
B.BR.05.05BR1074.JN692459.2005	C.IL.98.98IS002.AF286233.1998
B.BR.06.06BR1115.JN692479.2006	C.IL.99.99ET1.AY255823.1999
B.BR.07.BP00047_RH01.JN687739.2007	C.IN.03.D24.EF469243.2003
B.CA.06.502_1799_FL02.JF320427.2006	C.IN.93.93IN999.AF067154.1993
B.CA.07.502_1191_03.JF320424.2007	C.IN.94.94IN476.AF286223.1994
B.CA.96.WC10P_1.AY314044.1996	C.IN.98.98IN012.AF286231.1998
B.CA.98.CANA3FULL.AY779550.1998	C.KE.00.KER2010.AF457054.2000
B.CH.00.HIV_CH_BID_V3529_2000.JQ403022.2000	C.KE.91.KNH1268.AY945738.1991
B.CH.01.HIV_CH_BID_V3533_2001.JQ403025.2001	C.MW.08.702010432.e25.JQ779242.2008
B.CH.02.HIV_CH_BID_V3511_2002.JQ403019.2002	C.MW.09.703010256.e8.JQ779096.2009
B.CH.03.HIV_CH_BID_V4470_2003.JQ403045.2003	C.MW.93.93MW_965.AY713413.1993
B.CH.04.HIV_CH_BID_V4408_2004.JQ403042.2004	C.TZ.01.A125.AY253304.2001
B.CN.01.CNHN24.AY180905.2001	C.TZ.02.CO178.AY734556.2002
B.CN.02.02HNsq4.DQ007902.2002	C.TZ.08.707010457.e11.JQ779129.2008
B.CN.03.SHXDC0081.JF932492.2003	C.TZ.97.97TZ05.AF361875.1997
B.CN.05.05CNHB_hp3.DQ990880.2005	C.TZ.98.98TZ013.AF286234.1998
B.CN.06.CC056.JF932482.2006	C.US.98.98US_MSC3018.AY444800.1998
B.CN.07.AH070011.JF932468.2007	C.YE.02.02YE511.AY795906.2002
B.CN.08.cbjc468.JF932498.2008	C.ZA.00.1119MB.AY463229.2000
B.CN.09.09LNA014.JX960597.2009	C.ZA.01.2134MB.AY463237.2001
B.CN.10.DEMB10CN002.JX140658.2010	C.ZA.02.02ZAPS006MB1.DQ351220.2002
B.CN.98.YN9802.JF932495.1998	C.ZA.03.03ZASK120B1.DQ011176.2003
B.CO.01.PCM013.AY561237.2001	C.ZA.04.04ZAPS157MB1.DQ351232.2004
B.CU.99.Cu19.AY586542.1999	C.ZA.05.05ZASK244B1.DQ369992.2005
B.CY.05.CY020.FJ388891.2005	C.ZA.07.DEMC07ZA011.JX140664.2007

B.CY.05.CY065.FJ388910.2005	C.ZA.08.705010534.e18.JQ779177.2008
B.CY.06.CY180.FJ388955.2006	C.ZA.10.DEMC10ZA001.JX140669.2010
B.CY.07.CY189.JF683742.2007	C.ZA.97.97ZA003.AY118165.1997
B.CY.08.CY238.JF683785.2008	C.ZA.98.CTSC2.AY043176.1998
B.CY.09.CY241.JF683787.2009	C.ZA.99.99ZALT21.EU293446.1999
B.DE.03.HIV_DE_BID_V3307_2003.JQ403048.2003	C.ZM.02.02ZM114.AB254146.2002
B.DE.04.9213_d0.JQ416158.2004	C.ZM.03.ZM249M_flC1.FJ496209.2003
B.DE.86.HAN.U43141.1986	C.ZM.96.96ZM651.AF286224.1996
B.DK.01.CTL_017.EF514705.2001	Ref.C.BR.92.BR025_d.U52953.1992
B.DK.04.PMVL_012.EF514699.2004	Ref.C.IN.95.95IN21068.AF067155.1995
B.DK.07.PMVL_011.FJ694790.2007	Ref.C.ZA.04.04ZASK146.AY772699.2004
B.EC.89.EC003.AY173959.1989	
B.ES.05.X1890.EU786672.2005	
B.ES.06.X2102.EU786677.2006	
B.ES.07.X2210_3.EU786679.2007	
B.ES.08.ES_X2515_3.GQ372988.2008	
B.ES.09.X2689_2.GU362885.2009	
B.FR.08.DEMB08FR002.JX140654.2008	
B.GA.88.OYI_397.M26727.1988	
B.GB.03.MM33d12p.HM586187.2003	
B.GB.04.MM39d11p.HM586193.2004	
B.GB.05.MM43d368_GN1.HM586209.2005	
B.GB.83.CAM1.D10112.1983	
B.GB.86.GB8_46R.AJ271445.1986	
B.GE.03.03GEMZ010.DQ207942.2003	
B.HK.06.HK002.FJ460499.2006	
B.HT.05.05HT_129473.EU839603.2005	
B.JM.05.05JM_KJ108.EU839605.2005	
B.JM.09.09JM_PF09XN.HM030559.2009	
B.JP.00.DR2508.AB289588.2000	
B.JP.01.DR388.AB289590.2001	
B.JP.04.DR5913.AB480696.2004	
B.JP.05.426.AB428556.2005	
B.KR.02.02OSG1.JQ429433.2002	
B.KR.03.03HJY8.JQ316131.2003	
B.KR.04.04KJS8.JQ316130.2004	
B.KR.05.05CSR3.DQ837381.2005	
B.KR.07.07KYY4.JQ341411.2007	
B.KR.97.WK.AF224507.1997	
B.NL.99.671_99T12.AY423381.1999	
B.PE.06.502_0491_wg5.JF320183.2006	
B.PE.07.502_0525_wg5.JF320191.2007	
B.PL.08.BP00074_RH02.JN687774.2008	
B.PY.02.02PY_PSP0019.JN251896.2002	
B.PY.03.03PY_PSP0115.JN251906.2003	
B.RU.04.04RU128005.AY682547.2004	
B.RU.09.09RU4457.JX500709.2009	
B.RU.11.11RU21n.JX500708.2011	

B.TH.04.04TH317223.JN248321.2004
B.TH.05.05TH356764.JN248344.2005
B.TH.06.AA010a_WG3.JX446800.2006
B.TH.07.AA040a_WG11.JX447156.2006
B.TH.08.AA115c03R.JX448103.2008
B.TH.96.M145.DQ354118.1996
B.TT.00.00TT_CRC50018.EU839607.2002
B.TW.94.TWCYS_LM49.AF086817.1994
B.UA.01.01UAKV167.DQ823362.2001
B.US.00.ES1_20.EF363123.2000
B.US.01.L805.FJ469738.2001
B.US.02.F706.FJ469728.2002
B.US.03.L8107.FJ469739.2003
B.US.04.CR0345Q.FJ469722.2004
B.US.05.012286G.FJ469683.2005
B.US.06.06US_SAJ_NVS23.JF689871.2006
B.US.07.07US_SAJ_C154.JF689877.2007
B.US.08.08US_SAJ_C203.JF689895.2008
B.US.09.DEMB09US002.JX140657.2009
B.US.10.CP1.JN397365.2010
B.US.11.ES38.JN397362.2011
B.US.83.RF_HAT3.M17451.1983
B.US.84.MNCG_MN.M17449.1984
B.US.85.RJS4_15_ENVBB.M37573.1985
B.US.86.5127_86.AY835774.1986
B.US.87.5113_87.AY835758.1987
B.US.88.5160_88.AY835763.1988
B.US.89.P896_89_6.U39362.1989
B.US.90.US1.AY173952.1990
B.US.91.DH12_3.AF069140.1991
B.US.92.92US657_1_301657_1.U04908.1992
B.US.94.5082_94.AY835773.1994
B.US.95.5073_95.AY835768.1995
B.US.96.5155_96.AY835753.1996
B.US.97.1006_08.AY331284.1997
B.US.98.WC3_0498_4.EF175212.1998
B.US.99.169_1999.JN599165.1999
B.UY.01.01UYTRA1179.AY781127.2001
B.UY.02.02UY_TSU1290.JN235958.2002
B.VE.10.DEMB10VE001.JX140659.2010
B.YE.02.02YE508.AY795905.2002
B.ZA.03.03ZAPS045MB2.DQ396398.2003
Ref.B.FR.83.HXB2_LAI_IIIB_BRU.K03455
Ref.B.NL.00.671_00T36.AY423387.2000
Ref.B.TH.90.BK132.AY173951.1990
Ref.B.US.98.1058_11.AY331295.1998