


Article

# HIV-1 Subtypes and 5′LTR-Leader Sequence Variants Correlate with Seroconversion Status in Pumwani Sex Worker Cohort

Raghavan Sampathkumar <sup>1,2</sup>, Joel Scott-Herridge <sup>2</sup> , Binhua Liang <sup>2,3</sup>, Joshua Kimani <sup>1,4</sup>, Francis A. Plummer <sup>1,2</sup> and Ma Luo <sup>1,2,\*</sup>

<sup>1</sup> Department of Medical Microbiology, University of Manitoba, Winnipeg, MB, R3E 0J9, Canada; raghavans1@gmail.com (R.S.); jkimani@csrkenya.org (J.K.); Francis.Plummer@umanitoba.ca (F.A.P.)

<sup>2</sup> National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, R3E 3R2, Canada; jascottherridge@gmail.com (J.S.-H.); Binhua.Liang@umanitoba.ca (B.L.)

<sup>3</sup> Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, MB, R3E 0J9, Canada

<sup>4</sup> Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya

\* Correspondence: Ma.Luo@umanitoba.ca; Tel.: +1-204-789-5072; Fax: +1-204-784-4835

Received: 13 October 2017; Accepted: 20 December 2017; Published: 23 December 2017

**Abstract:** Within the Pumwani sex worker cohort, a subgroup remains seronegative, despite frequent exposure to HIV-1; some of them seroconverted several years later. This study attempts to identify viral variations in 5′LTR-leader sequences (5′LTR-LS) that might contribute to the late seroconversion. The 5′LTR-LS contains sites essential for replication and genome packaging, viz, primer binding site (PBS), major splice donor (SD), and major packaging signal (PS). The 5′LTR-LS of 20 late seroconverters (LSC) and 122 early seroconverters (EC) were amplified, cloned, and sequenced. HelixTree 6.4.3 was employed to classify HIV subtypes and sequence variants based on seroconversion status. We find that HIV-1 subtypes A1.UG and D.UG were overrepresented in the viruses infecting the LSC ( $P < 0.0001$ ). Specific variants of PBS ( $P < 0.0001$ ), SD1 ( $P < 0.0001$ ), and PS ( $P < 0.0001$ ) were present only in the viral population from EC or LSC. Combinations of PBS [PBS-2 ( $P < 0.0001$ ) and PBS-3 ( $P < 0.0001$ )] variants with specific SD sequences were only seen in LSC or EC. Combinations of A1.KE or D with specific PBS and SD variants were only present in LSC or EC ( $P < 0.0001$ ). Furthermore, PBS variants only present in LSC co-clustered with PBS references utilizing tRNA<sup>Arg</sup>; whereas, the PBS variants identified only in EC co-clustered with PBS references using tRNA<sup>Lys</sup>,<sup>3</sup> and its variants. This is the first report that specific PBS, SD1, and PS sequence variants within 5′LTR-LS are associated with HIV-1 seroconversion, and it could aid designing effective anti-HIV strategies.

**Keywords:** HIV/AIDS; seroconversion; 5′LTR-leader sequence; genetic diversity; primer binding site sequences; splicing donor sequences; packaging signal; HIV subtypes

## 1. Introduction

In 2016, there were 62,000 new HIV infections, and 1,600,000 people living with HIV in Kenya [1]. Efforts are underway, globally, to find ways to prevent infection, as well as to explore practical cures for HIV [2,3]. The most at-risk individuals for infection by HIV are commercial sex workers (CSW), intravenous drug users, and men who have sex with men (MSM). The CSW population is at increased risk, as they may have hundreds of sexual partners each year. Compounding the risk of infection and transmission, many of them could be intravenous drug users, and/or may be infected with other sexually transmitted pathogens that could enhance HIV transmission [4]. Kenya has around 133,675 sex workers [1]. The percentage of female sex workers of a population has been reported to strongly correlate with total HIV/AIDS prevalence [4]. Female sex workers have a 13.5-fold higher

risk of being HIV infected compared to other women [5]. The high infection risk of this population can provide critical insights into HIV infection, disease progression, and transmission. The research on the HIV infection of this group could provide clues for designing effective HIV-1 control strategies.

The Pumwani sex worker cohort in Kenya was established in 1985. The overall HIV-1 prevalence in the cohort is over 73.7%. The majority of HIV-1 negative women at cohort entry seroconverted within three years. A small group of women remain HIV-1 negative, despite heavy exposure through active sex work [6,7]. This observed resistance to HIV-1 infection was not due to safer sexual practices, altered cellular susceptibility to HIV-1, or known chemokine-receptor polymorphisms [8]. Polymorphism in HLA and non-HLA genes have been shown to influence HIV-1 resistance in this cohort [7,9–15], as well as in other populations [16]. Some of the HIV “resistant” sex workers seroconverted after being seronegative for many years; these women were designated as late seroconverters (LSC). Late seroconversion may occur in HIV-1-resistant sex workers, despite preceding HIV-specific CD8<sup>+</sup> T cell responses [17]. Viral factors, such as subtype and functional genetic differences, have not been properly examined. It is conceivable that these women, previously resistant to HIV-1 infection, were infected by a more infectious, pathogenic viral species. This study intends to examine the viral factors infecting this group of late seroconverters.

Among the major groups of HIV-1, M, N, O, and P, M group viruses have been responsible for the majority of HIV-1 infections worldwide [18]. Nine major subtypes identified within group M viruses are A-D, F-H, J, and K. Sub-subtypes have been observed for clade A (A1, A2) as well as F (F1, F2) viruses. Additionally, group M includes 90 circulating recombinant forms. Globally, more prevalent subtypes are B (56.0%), C (17.0%), A (5.8%), D (3.1%), G (1.0%) and F (1.0%). In Kenya, subtypes A (68.0%), D (13.5%), and C (5.8%) were more common [19]. Analysis of 41 near full-length HIV-1 sequences from Kenya reported 56.1% subtype A, 2.4% each of subtypes C and D, and 39% recombinant [20]. Another study of 176 Kenyan patients observed 73.9% A1, 10.8% C, 10.2% D, and 0.6% of G and A2 clades [21]. Our previous analysis of HIV gag of 468 HIV-1 positive women also showed that the Pumwani sex worker cohort is primarily infected with clade A1 at 71%, 65%, 67%, and 63%, followed by clade D at 14%, 22%, 23%, and 20% for p17, p24, p7, and p6, respectively [22]. Different subtypes exhibit unique influences over viral transmission, replication, disease progression, virulence, and susceptibility to antiretroviral drugs [23–26]. Studies have shown that debilitated HIV-1 viruses needed only few mutations to attain fitness recovery, and these events most commonly involved the 5' untranslated leader sequence [27]. This region contains three important sites for viral replication: primer-binding site (PBS), major splice donor site (SD), and major packaging signal (PS) [28,29]. HIV-1 loses infectivity upon complete deletion of PBS, and this highlights the functional importance [30]. Reverse transcription initiation involves the binding of cellular tRNA<sup>Lys,3</sup> to the 18 nucleotide PBS that is located upstream of *gag*. This sequence is complementary to the 3' terminal 18 nucleotides of this particular tRNA molecule [31]. Usage of tRNA<sup>Lys,5</sup>, though infrequent, as primer in HIV replication, has also been reported [32]. While all retroviruses make use of tRNA as a primer for reverse transcription, specific viruses are known to exhibit preference for particular tRNA molecules, as exemplified by usage of tRNA<sup>Pro</sup> and tRNA<sup>Trp</sup> by murine leukemia virus and avian sarcoma virus-avian leukosis virus groups, respectively [33]. The preference for usage of specific tRNA primers by HIV-1 for reverse transcription could be linked to its fitness [34]. HIV-1 splicing aids in optimal expression of its proteins, facilitating infection and subsequent generation of new infectious viral progenies. HIV-1 uses several splice sites in its genome to produce more than 40 different mRNA transcripts [35]. Major splice donor site, SD1, joins to a 3' splice acceptor region downstream of *pol*, and this results in a transcript encoding envelope glycoproteins. Furthermore, all downstream splicing events become silenced if SD1 is mutated [36]. The PS is a GGAG tetraloop found at the end of stem-loop SL3 [37]. This structure is seen only among unspliced HIV-1 RNAs, and it interacts with nucleocapsid protein, a breakdown product of gag polyprotein [37]. It has been suggested that the newly synthesized gag protein could bind PS, leading to translation inhibition of part of unspliced RNAs, which in turn might ensure full-length viral RNA molecules are available for packaging [38]. Deletions in PS are known to cause substantial reduction in

genome packaging capability [29]. Primer-binding site, SD1, and PS, are therefore pivotal sequence elements for the replication and proliferation of HIV-1.

The presence of these three essential sites in 5′LTR-leader sequence led us to choose this region to examine the viruses infecting the late seroconverters. We compared 5′LTR-leader sequences from the late seroconverters with those from women who were seropositive at enrollment, or seroconverted within the first three years of enrollment in the Pumwani sex worker cohort. We hypothesized that late seroconverters were infected with specific variants of HIV-1, whose distinct 5′ leader sequence profile could confer potential replicative advantages, besides efficient genome packaging capability.

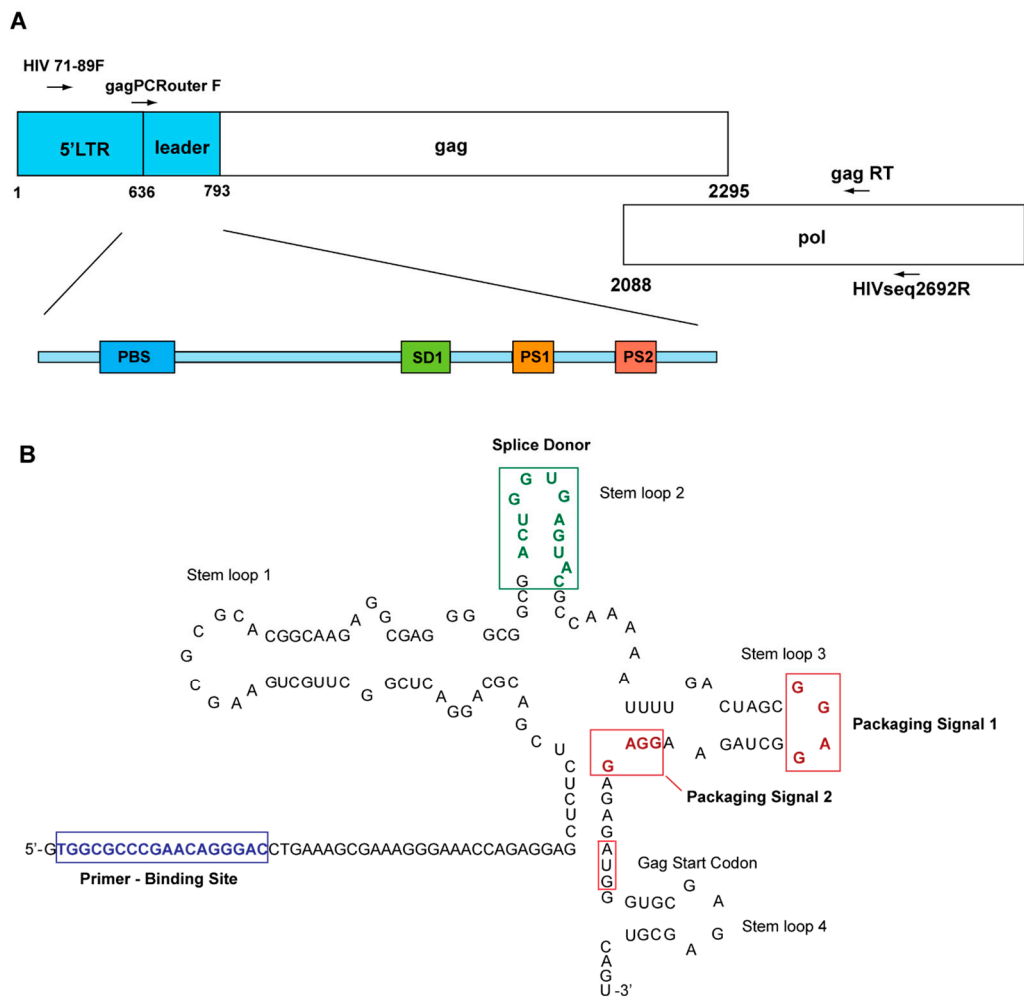
## 2. Materials and Methods

### 2.1. Sample Collection

HIV-1 positive sex workers and late seroconverters from the Pumwani sex worker cohort were selected for this study. No anti-retroviral treatments (ARTs) were available during the sample collection period in Kenya, thus, none of the samples analyzed in this study were confounded by ARTs. Informed written consent was obtained from all study subjects. The University of Manitoba, as well as University of Nairobi ethics review panels, have approved studies with these subjects. Women in this cohort are routinely screened for HIV-1 infection by serology and PCR amplification for the *env*, *nef*, and *vif* genes. Women were defined as resistant to HIV-1 infection if they remain HIV-1 seronegative and PCR negative for a minimum of three years of follow up after enrollment [6]. The late seroconverters were defined as those who seroconverted after meeting the defined resistance criteria [6,17]. In this study, 20 patients met this criterion. Seven of these patients also had samples collected at different dates since seroconversion. The control population consisted of 122 seropositive patients, of which 101 women were positive at enrollment, and 21 seroconverted within three years after enrollment. Sixteen control subjects had more than one timepoint sample. The average seronegative time of the late seroconverters is  $5.94 \pm 2.92$  years, compared to an average of  $0.80 \pm 0.70$  seronegative years of the 21 seroconverters in the positive control group.

### 2.2. Genomic DNA Isolation and Nested PCR Amplification of Partial 5′LTR of HIV-1

Genomic DNA was isolated from peripheral blood mononuclear cells of the study subjects using QIAamp DNA Mini Kit (Qiagen Inc., Mississauga, ON, Canada). Nested PCR was carried out, using Expand High Fidelity PCR system (Roche Diagnostics, Mannheim, Germany), to amplify a 2 kb fragment containing partial 5′LTR, HIV-1 *gag*, and partial protease gene (found in *pol*) (Figure 1A,B). Primers HIV71-89F (5′-CTTCCCTGATTGGCAGAAY-3′) and HIVseq2692R (5′-GGATTTTCAGG CCCAATTTTG-3′) were used for the first round of amplification. The PCR cycle conditions were 2 min initial denaturation at 94 °C, followed by 35 cycles of 15 s at 94 °C, 30 s at 53 °C, and 68 °C for 5 min, with final extension at 68 °C for 15 min. Primers Gag PCR outerF (5′-AATCTCTAGCAGTGGCGCCCGAACAG-3′) and GagRT (5′-CCATTGTTTAACCTTTGGGCCATCCA-3′) were used for the second round PCR reaction. One microliter of PCR product from first round amplification was used as template. Thermal cycler parameters were set as 94 °C for 2 min, 35 cycles of 94 °C for 15 s, 59 °C for 30 s and 68 °C for 4 min, with final extension at 68 °C for 10 min. All PCR products were examined using 1% agarose gel electrophoresis.



**Figure 1.** HIV-1 5'LTR leader sequences. (A). A schematic sketch of 5'LTR-leader sequence variant positions analyzed in this study; (B). a schematic sketch of secondary structure of HIV-1 5'leader sequence.

### 2.3. Cloning and Sequencing of Amplified Partial 5'LTR Sequences

Prior to cloning, the PCR products were TA extended. Each TA extended PCR product was ligated into pCR®4-TOPO®vector (TOPO TA Cloning Kit for Sequencing, Invitrogen Life Technologies, Carlsbad, CA, USA) and transformed into One Shot®TOP10 Chemically Competent *E. coli*. Forty-eight clones were picked from each sample and cultured for 16–20 h in 2 mL LB medium with ampicillin (200 µg/mL). Bacteria cultures were pelleted by centrifugation for 6 min at 1900 g. QIAprep 96 Turbo Miniprep Kit protocol was used to isolate plasmids containing the amplified HIV-1 fragment. EcoR1 restriction digestion and agarose gel electrophoresis were conducted to detect the presence of insert DNA. T3 and T7 sequencing primers were used to sequence the clones using BigDye version 3.1 Cycle sequencing kit (Applied Biosystems™, Carlsbad, CA, USA), and analyzed with an ABI3730XL DNA Analyzer, available at DNACORE facility of the National Microbiology Laboratory, Winnipeg, MB, Canada.

### 2.4. Sequence and Phylogenetic Analyses

The sequences were examined using Sequencher version 4.6 (Gene Codes Corporation, MI, USA). HIV *gag* sequences were removed and 160 nucleotide sequences of partial 5'LTR region, including the part of *U5* and untranslated leader sequence, were retained for further analysis. Close to 4000

5′LTR leader sequences have been generated. Phylogenetic analysis using MEGA 3.1 [39] was done to classify viral subtypes. Briefly, partial 5′LTR sequences were aligned with 51 reference sequences obtained from HIV sequence database [19]. Alignment was done with ClustalW and phylogenetic trees were generated. Alignment and phylogenetic relatedness to reference sequences permitted subtype identification for each clone. To confirm the subtype assignment of 5′LTR sequences by phylogenetic analysis, we also conducted phylogenetic analysis of p17 sequences of these cloned sequences. The results confirmed the subtype assignment using the sequences of the partial 5′LTR region. Two examples of the phylogenetics analysis using p17 sequences are shown in Supplemental Figures S1 and S2. To assess the possible function of PBS variants observed in the present study, 19 published PBS sequences, that used different tRNA primers for reverse transcription, were taken as reference to construct a maximum likelihood method based phylogenetic tree, using MEGA 6. The 19 PBS sequences in the reference alignment included those corresponding to tRNA<sup>Lys,3</sup> (wild-type), tRNA<sup>Lys,1,2</sup>, tRNA<sup>Lys,5a</sup>, tRNA<sup>Lys,1</sup>, EctRNA<sup>Lys,3</sup> (*E. coli* tRNA), tRNA<sup>Pro</sup>, tRNA<sup>Ile</sup>, tRNA<sup>Met</sup>, tRNA<sup>Met(e)</sup> (used in elongation), tRNA<sup>Met(i)</sup> (used in initiation), tRNA<sup>Met(i)</sup> AG (contains a transition), tRNA<sup>Ser</sup>, tRNA<sup>Phe</sup>, tRNA<sup>Thr</sup>, tRNA<sup>Gln,1</sup>, tRNA<sup>Gln,3</sup>, tRNA<sup>His</sup>, tRNA<sup>Arg(ACG)</sup> and tRNA<sup>Arg(CCU)</sup> [32,34,40–47].

### 2.5. Sequence Variant Classification by Recursive Partitioning Analysis

Recursive partitioning methods have become popular and widely used tools for non-parametric regression and classification in many scientific fields [48]. They can deal with large numbers of predictor variables, even in the presence of complex interactions, and have been applied successfully in genetics, clinical medicine, and bioinformatics within the past few years [48]. In this study, we used the recursive partitioning methods based interactive tree analysis tool in HelixTree SNP and Variation Suite version 6.4.3 (Golden Helix, Inc., Bozeman, MT, USA) to analyze the large pool of sequence variants of the three important sites (PBS, SD, and PS) within the 5′LTR leader region. The interactive tree analysis tool was developed based on formal inference recursive modeling (FIRM) technology by Dr. Douglas Hawkins [48–57] accessed 21 December 2017), and has taken the statistical foundations of FIRM and augmented it with faster and more exact segmenting algorithms. It has also extended FIRM methods to include multivariate response. Recursive partitioning uses a set of data and, based on some criterion, partitions or splits the original set into smaller sets. These smaller sets are, in turn, split into still smaller sets. This process continues (recursively) until additional splitting of the data into smaller sets gives no statistically meaningful information.

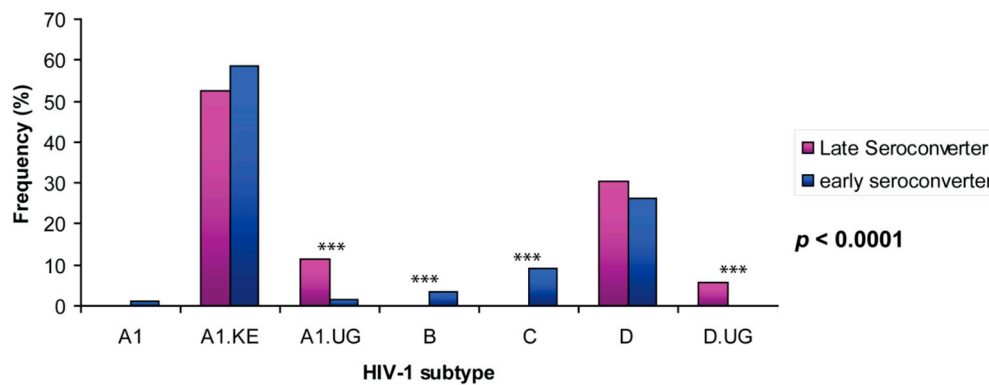
For example, because the aim of our study is to identify the sequence variants of the three sites within the HIV 5′LTR leader region that are predominantly detected among late seroconverters, we designated the *u*-value of late seroconverters as 1.0 and the *u*-value for early seroconverters as 0. For example, when analyzing sequence variants of PBS using the tree analysis tool, the sequence variants were partitioned based on whether they are detected in the early or late seroconverters and the *p* value. PBS sequence variants in the tree node with *u*-value equal to 1 indicate that the PBS sequence variants were identified only in late seroconverters, whereas the PBS sequence variants in the tree node with *u*-value equal to 0 were only identified in early seroconverters. The PBS sequence variants in the tree nodes with *u*-value varying between 1 and 0 indicate that the sequences exist in both early and late seroconverters. Because it is possible that not only specific sequence variants of PBS can influence seroconversion, but also the combinations of the PBS sequence variants with specific sequences of SD or PS may play a role in seroconversion, the PBS sequences in the nodes with *u*-values between 0 and 1 can be further classified by sequence variants of SD or PS. At each step of analysis, a combination of *u*-value and *p* value was used to define the sequences associated with late or early seroconverters.

Differences in subtype distributions of the sequence variants between late seroconverters and controls were analyzed by Pearson  $\chi^2$  analysis using SPSS version 13.0. *p* values equal to or less than 0.05 were considered statistically significant.

### 3. Results

#### 3.1. Uganda A1 and D Subtype 5'LTR-Leader Sequences Were Significantly Enriched in HIV Viral Population from Late Seroconverters

A total of 3678 sequences from 20 late seroconverters and 122 early seroconverters were phylogenetically analyzed to determine their HIV-1 subtypes. This analysis only included the sequences of the earliest sampling date of the available samples from each patient. Similar to previous studies, subtype A predominates in the HIV viral population of this Kenyan population, followed by subtype D. The frequencies of subtypes A1.KE, A1.UG, D, and D.UG were 57.2%, 3.7%, 27.2%, and 1.3%, respectively (Figure 2 and Table 1). There is a significant difference in overall subtype distribution of 5'LTR leader sequences between viral population in early and late seroconverters ( $p < 0.0001$ ). While subtypes B (0% versus 3.4%,  $p < 0.0001$ ) and C (0% versus 9.2%,  $p < 0.0001$ ) sequences were not observed among the late seroconverters, subtype A1.UG sequences were significantly enriched in the late seroconverters compared to the ones in early seroconverters (11.4% versus 1.5%,  $p < 0.0001$ ). Further, subtype D.UG sequences were absent in early seroconverters (5.7% versus 0%,  $p < 0.0001$ ). It is apparent that the viral population infecting late seroconverters was enriched with subtype A1.UG and D.UG 5'LTR leader sequences.



**Figure 2.** HIV-1 subtype distribution among late seroconverters and early seroconverters. \*\*\* :  $p$  value is less the 0.0001.

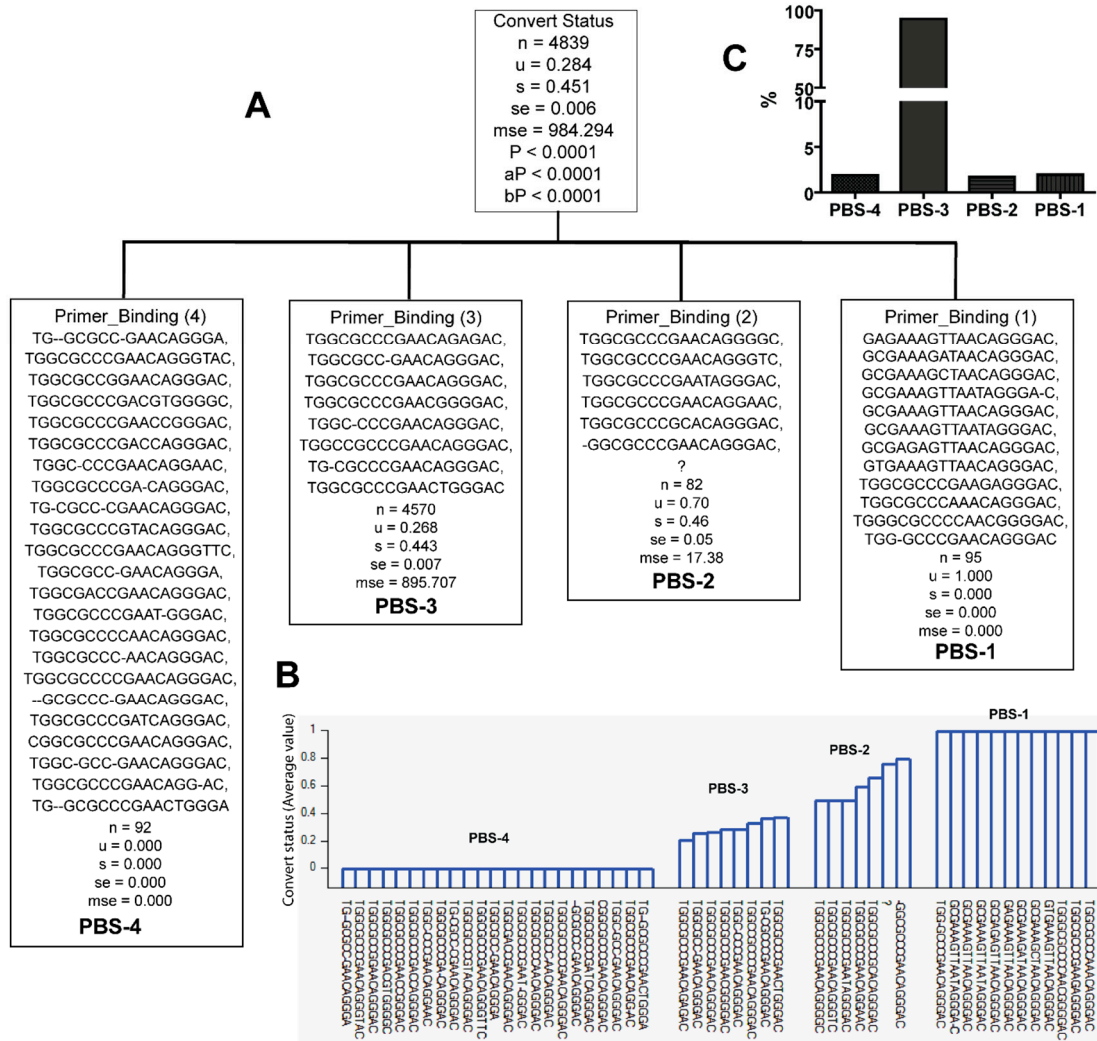
**Table 1.** Comparison of HIV-1 subtype distribution among late seroconverters and early seroconverters.

		HIV-1 Subtype	A1	A1.KE	A1.UG	B	C	D	D.UG	Total	Ave Seq./ind.	$p$ Value
A	All 142 individuals	No.	30	2162	135	97	263	945	46	3678		
		%	0.8	58.8	3.7	2.6	7.2	25.7	1.3	100		
B	Late seroconverter ( $n = 20$ )	No.	0	452	91	0	0	211	46	800	40	<0.0001
		%	0	56.5	11.4	0	0	26.4	5.8	100	(19–47)	
	Early seroconverter ( $n = 122$ )	No.	30	1710	44	97	263	734	0	2878	30	
		%	1.0	59.4	1.5	3.4	9.1	25.5	0	100	(11–75)	

#### 3.2. Unique Sequences and Combinations of PBS, SD, and Ps Sequences in Late Seroconverters

We then examined whether specific sequences of primer binding site (PBS), splice donor (SD), and packaging signal (PS), and their combinations, are more likely to be associated with HIV viral population in late seroconverters. For this, we included all 4839 sequences from multiple sample dates of the patients in recursive partition analysis using the Tree analysis tool of HelixTree 6.4.3. Recursive partitioning analysis classifies the 5'LTR-leader sequence variants based on their nucleotide sequences, subtypes, and their origin, into early (designated as 0) or late seroconverters (designated as 1) (Figures 3–5 and Table 2). The analysis showed that specific sequence variants of PBS were

only identified in the viral population of either early or late seroconverters ( $P_c < 0.0001$ ) (Figure 3 and Table 2). Specifically, 12 PBS sequence variants were only found in the viral population of late seroconverters (PBS-1, Figure 3 and Table 2), and 23 PBS sequence variants were only identified in the viral population of early seroconverters (PBS-4, Figure 3A,B, and Table 2). Some PBS sequence variants were identified in the viral population of both early and later seroconverters (PBS-2 and PBS-3, Figure 3 and Table 2).



**Figure 3.** Classification of primer binding sequence (PBS) variants according to whether they were identified from late seroconverters ( $u = 1$ ) or early seroconverters ( $u = 0$ ). Note: question mark denotes lack of sequence.  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value. (A) Classification tree; (B) PS variants and their distribution with seroconversion status; (C) PBS variants frequencies.

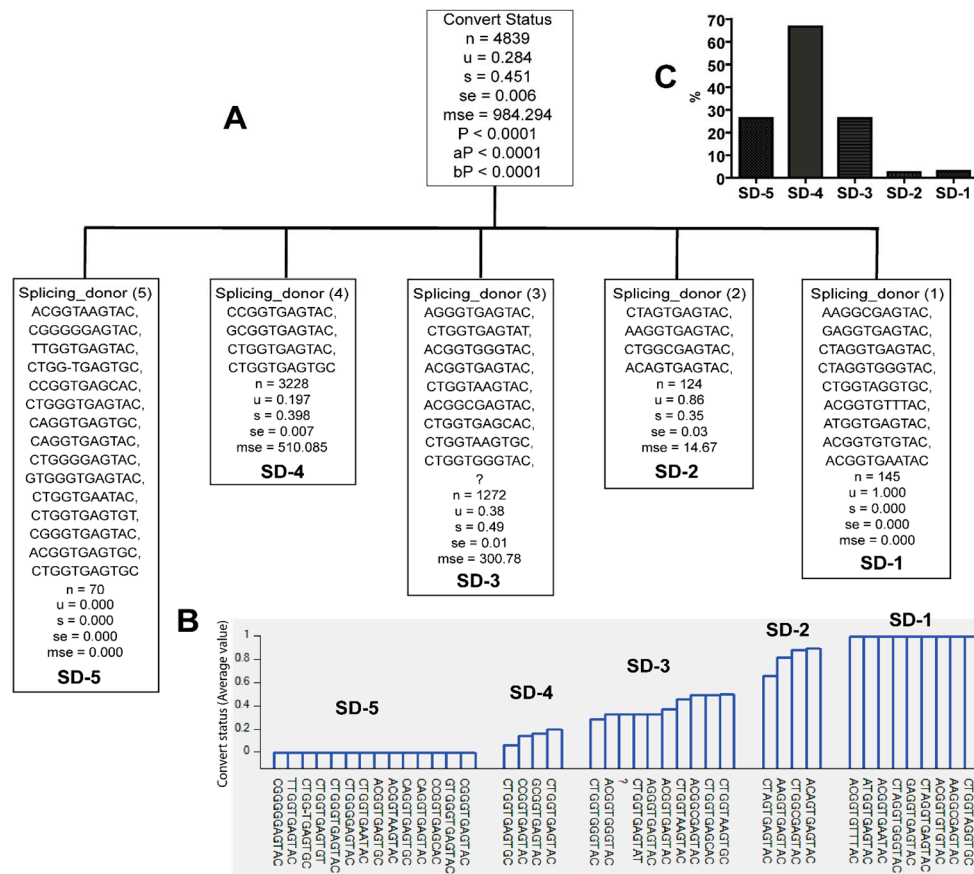
Table 2. PBS, SD1, and PS sequences listed in Figures 3–5.

Group	Sequences	Subtypes	Frequency in EC or LSC
PBS-1	GAGAAAGTTAACAGGGAC, GCGAAAAGATAACAGGGAC, GCGAAAAGCTAACAGGGAC, GCGAAAAGTTAATAGGGA-C, GCGAAAAGTTAACAGGGAC, GCGAAAAGTTAATAGGGAC, GCGAGAGTTAACAGGGAC, GTGAAAAGTTAACAGGGAC, TGGCGCCCCAAGAGGGAC, TGGCGCCCCAACAGGGAC, TGGCGCCCCAACAGGGAC, TGG-GCCCCAACAGGGAC	A1.KE (1.1%), A1.UG (93.67%), D (3.2%), D.UG (2.1%)	0/122 EC, 5/12 LSC
PBS-4	TG-GCGCC-GAACAGGGA, TGGCGCCCCAACAGGGTAC, TGGCGCCCGAACAGGGAC, TGGCGCCCCGACGTGGGGC, TGGCGCCCGAACCCGGGAC, TGGCGCCCCGACCAGGGAC, TGGC-CCCGAACAGGAAC, TGGCGCCCCA-CAGGGAC, TG-CGCC-CGAACAGGGAC, TGGCGCCCCGTACAGGGAC, TGGCGCCCCAACAGGGTTC, TGGCGCC-GAACAGGGA, TGGCGACCGAACAGGGAC, TGGCGCCCCAAT-GGGAC, TGGCGCCCCAACAGGGAC, TGGCGCCCCAACAGGGAC, TGGCGCCCCGAACAGGGAC, -GCGCCC-GAACAGGGAC, TGGCGCCCCGATCAGGGAC, CGGCGCCCCAACAGGGAC, TGGC-GCC-GAACAGGGAC, TGGCGCCCCGAACAGG-AC, TG-GCGCCCCGAACACTGGGA	A1 (2.2%), A1.KE (38%), B (2.2%), C (12.0%), D (42%)	36/122 EC, 0/20 LSC
SD1-1	AAGGCGAGTAC, GAGGTGAGTAC, CTAGGTGAGTAC, CTAGGTGGGTAC, CTGGTAGGTGC, ACGGTGTTTAC, ATGGTGAGTAC, ACGGTGAGTAC, ACGGTGAATAC	A1.KE (92.4), D (6.2%), D.UG (1.4%)	0/122 EC, 6/20 LSC
SD1-5	ACGGTAAGTAC, CGGGGGAGTAC, TTGGTGAGTAC, CTGG-TGAGTGC, CCGGTGAGCAC, CTGGGTGAGTAC, CAGGTGAGTGC, CAGGTGAGTAC, CTGGGGAGTAC, GTGGGTGAGTAC, CTGGTGAATAC, CTGGTGAGTGT, CGGGTGAGTAC, ACGGTGAGTGC, CTGGTGAGTGC	A1 (1.4%), A1.KE (90%), B (1.4%), C (2.9%), D (4.3%)	15/122 EC, 0/20 LSC
PS-1	AGTG, GGAC, CGAG, GGCG, AGGG	A1.KE (50%), A1.UG (16.7%), D (33.3%)	0/122 EC, 6/20 LSC
PS-3	GGAA, GAAG, GGAT, AGAG, ?	A1.KE (66.8%), A1.UG (1.3%), C (24.1%), D (7.8%)	30/122, EC, 4/20 LSC

Notes: ? denotes lack of sequence.

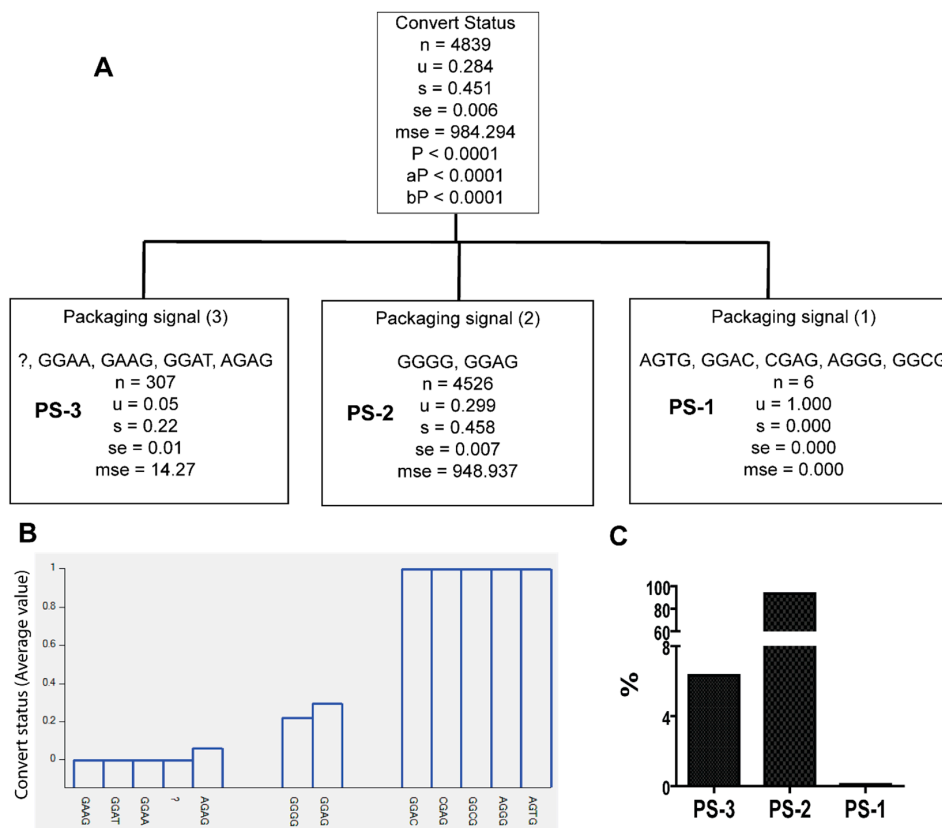


Similarly, specific sequence variants of SD were only identified in the viral population of either early or late seroconverters ( $P_c < 0.0001$ ) (Figure 4 and Table 2). Nine SD sequence variants were only found in the viral population of late seroconverters (SD-1, Figure 4A,B and Table 2), while 14 SD sequence variants were only found in the viral population of early seroconverters (SD-5, Figure 4A,B and Table 2). Some SD sequence variants were identified in the viral population of early as well as late seroconverters (SD-2, 3, 4, Figure 4 and Table 2).



**Figure 4.** Classification of splicing donor (SD) sequence variants according to whether they were identified from late seroconverters ( $u = 1$ ) or early seroconverters ( $u = 0$ ). Note: question mark denotes lack of sequence.  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value. (A) Classification tree; (B) PS variants, and their distribution with seroconversion status; (C) SD variants frequencies.

Likewise, specific sequence variants of PS were only identified in the viral population of either early or late seroconverters ( $P_c < 0.0001$ ) (Figure 5 and Table 2). Five PS sequence variants were only identified in the viral population of late seroconverters (PS-1, Figure 5A,B and Table 2), while four PS sequence variants were only seen in the viral population of early converters (PS-3, Figure 5A,B and Table 2). Some PS sequence variants were identified in the viral population of both early and late seroconverters (PS-2, Figure 5A,B and Table 2).



**Figure 5.** Classification of packaging signal (PS) sequence variants according to whether they were identified from late seroconverters ( $u = 1$ ) or early seroconverters ( $u = 0$ ). Note: question mark denotes lack of sequence.  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value. (A) Classification tree; (B) PS variants and their distribution with seroconversion status; (C) PS variants frequencies. “?” indicate the absence of the sequences”.

For the primer binding site sequence variants that existed in both early and late seroconverters (PBS-2 and PBS-3, Figure 3 and Table 2), we conducted further analysis to see whether combinations of specific PBS, SD sequence variants were more likely to exist in the viral population of late seroconverters. Further recursive analysis for the six sequence variants in the PBS-2 node with sequence variants of SD showed that the combinations of four specific SD sequence variants with the six PBS variants were only identified in the late seroconverters (PBS-2-SD-1; Figure 6A,B and Table 3). Similarly, PBS-3 node sequences in combinations with seven specific SD sequences occurred only in the viral population of late seroconverters (PBS-3-SD-1; Figure 7A,B and Table 3). In contrast, PBS-3 and 14 specific SD sequence variants (PBS-3-SD-4) existed only in the early seroconverters (Figure 7A,B and Table 3).

**Table 3.** Specific subtype, PBS, and SD variant combinations in late or early seroconverters.

Subtypes or PBS	SD or PBS	Seroconverter
TGGCGCCCGAACAGGGGC	CTGGTGAGTAC	LSC
TGGCGCCCGAACAGGGTC	AAGGTGAGTAC	
TGGCGCCCGAATAGGGAC	ACGGTGTTTAC	
TGGCGCCCGAACAGGAAC	ACAGTGAGTAC	
TGGCGCCCGCACAGGGAC? (PBS-2)	CAGGTGAGTAC	EC
	CAGGTGAGTGC	

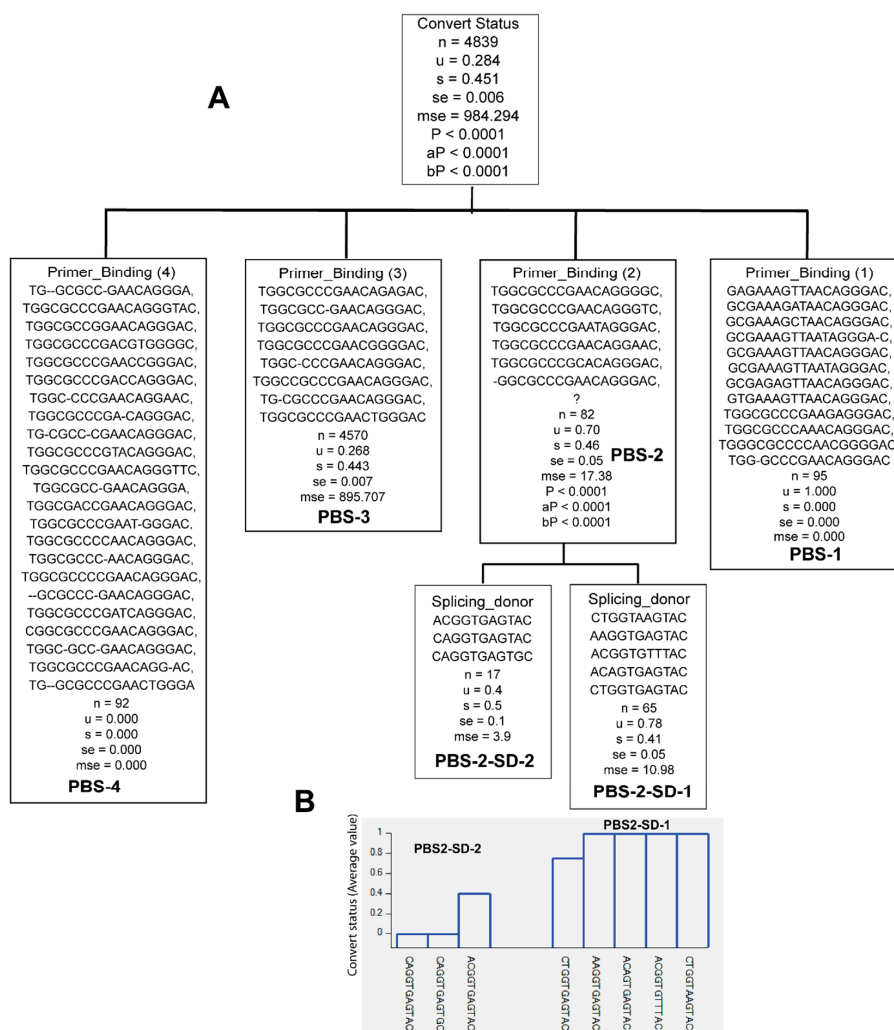
Table 3. Cont.

Subtypes or PBS	SD or PBS	Seroconverter
	AAGGCGAGTAC ACGGTGAATAC ATGGTGAGTAC CTAGGTAGGTGC CTAGGTGGGTAC CTGGTAGGTGC GAGGTGAGTAC	LSC
TGGCGCCCGAACAGGGAC TGGCGCCCGAACAGAGAC TGGCGCC-GAACAGGGAC TGGCGCCCGAACGGGGAC TGGC-CCCGAACAGGGAC TGGCGCCCGAACAGGGAC TG-CGCCCGAACAGGGAC TGGCGCCCGAACTGGGAC (PBS-3)	ACGGTGAGTGC ACGGTAAGTAC CAGGTGAGTAC CAGGTGAGTGC CCGGTGAGCAC CGGGGGAGTAC CGGGTGAGTAC CTGGTGAGTGT CTGGGGAGTAC CTGGGTGAGTAC CTGG-TGAGTGC CTGGTGAATAC GTGGGTGAGTAC TTGGTGAGTAC	EC
	AAGGCGAGTAC ACGGTGAATAC ACGGTGTGTAC ACGGTGTTTAC ATGGTGAGTAC ?	LSC
Subtype D	ACGGTAAGTAC ACGGTGAGTGC CCGGTGAGTAC CTAGTGAGTAC CTGGTAAGTAC CTGGTGAATAC CTGGTGAGCAC CTGGCGAGTAC	EC
	TGGCGCCCGAACAGGGTC TGGCGCCCGAACGGGGAC TGGCGCCCGAACAGGAAC TGG-GCCCGAACAGGGAC TGGCGCCCGAACAGGGAC TG-CGCCCGAACAGGGAC TGGCGCCCGAACAGGGAC (D-PBS-1)	LSC
Subtype D	TGGCGCCCGAACAGGGAC TGGCGCCCGAACAGGGTAC TGGCGCCCGACGTGGGGC TGGCGACCGAACAGGGAC TGGCGCCCGAACCGGGAC TGGCGCCCGTACAGGGAC TGGC-CCCGAACAGGGAC TGGCGCCCGATCAGGGAC TG-CGCC-CGAACAGGGAC TGGCGCCCGAACAGGGAC (D-PBS-3)	EC
A1.KE	CTAGGTGAGTAC CTAGGTGGGTAC CTAGTGAGTAC CTGGTAGGTGC (A1.KE-SD-1)	LSC

Table 3. Cont.

Subtypes or PBS	SD or PBS	Seroconverter
A1.KE	ACGGTGAGTAC	EC
	ACGGTGAGTGC	
	CAGGTGAGTAC	
	CAGGTGAGTGC	
	CCGGTGAGTAC	
	CGGGGGAGTAC	
	CTGGGTGAGTAC	
	CTGGTGAGTGT	
	CTGG-TGAGTGC	
	TTGGTGAGTAC	
	?	
	(A1.KE-SD-4)	

Note: '?' denotes lack of sequence.

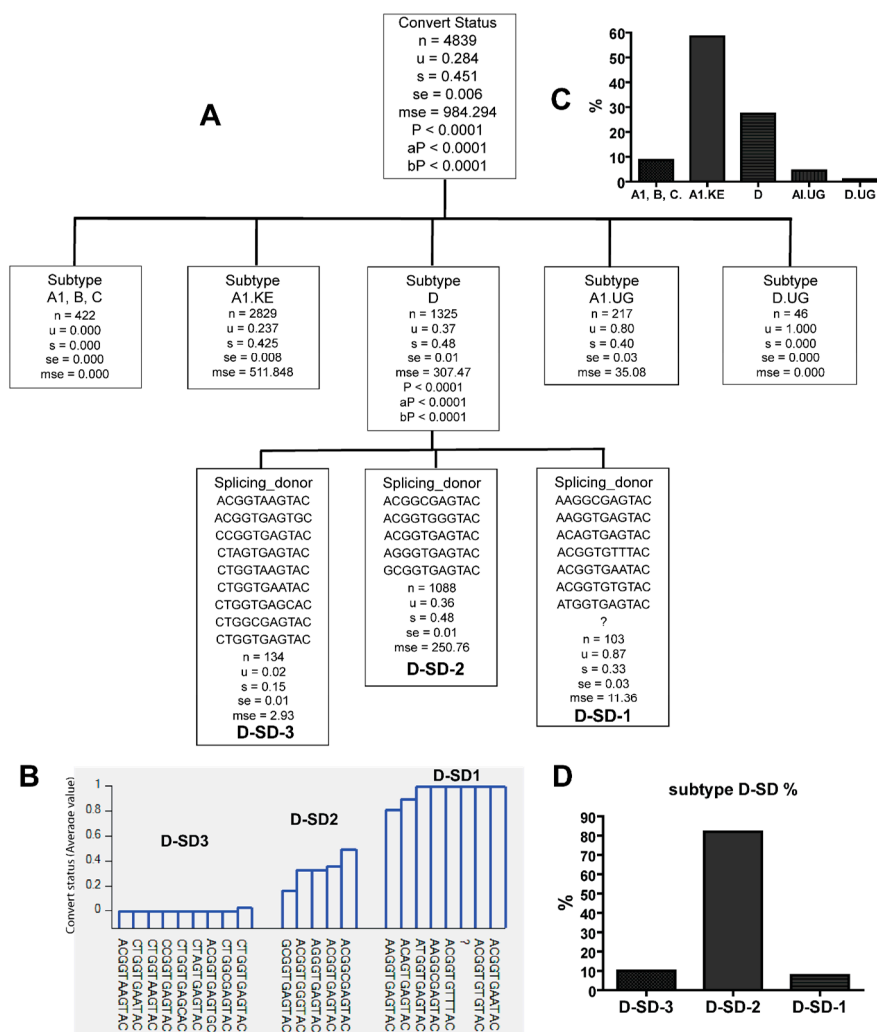


**Figure 6.** The combination of PBS variants (PBS-2) and splicing donor sequences identified only in either late ( $u = 1$ ) or early seroconverters ( $u = 0$ ). (A) PBS sequence variants in tree node of PBS-2 were further classified with SD sequence variants based on whether they were identified from late or early seroconverters; (B) the figure shows  $u$  value of sequence combinations based on whether the PBS-SD combinations were identified from late (1) or early (0) seroconverters. Note:  $n$ —counts (clone);  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value. “?” denotes lack of sequences”.

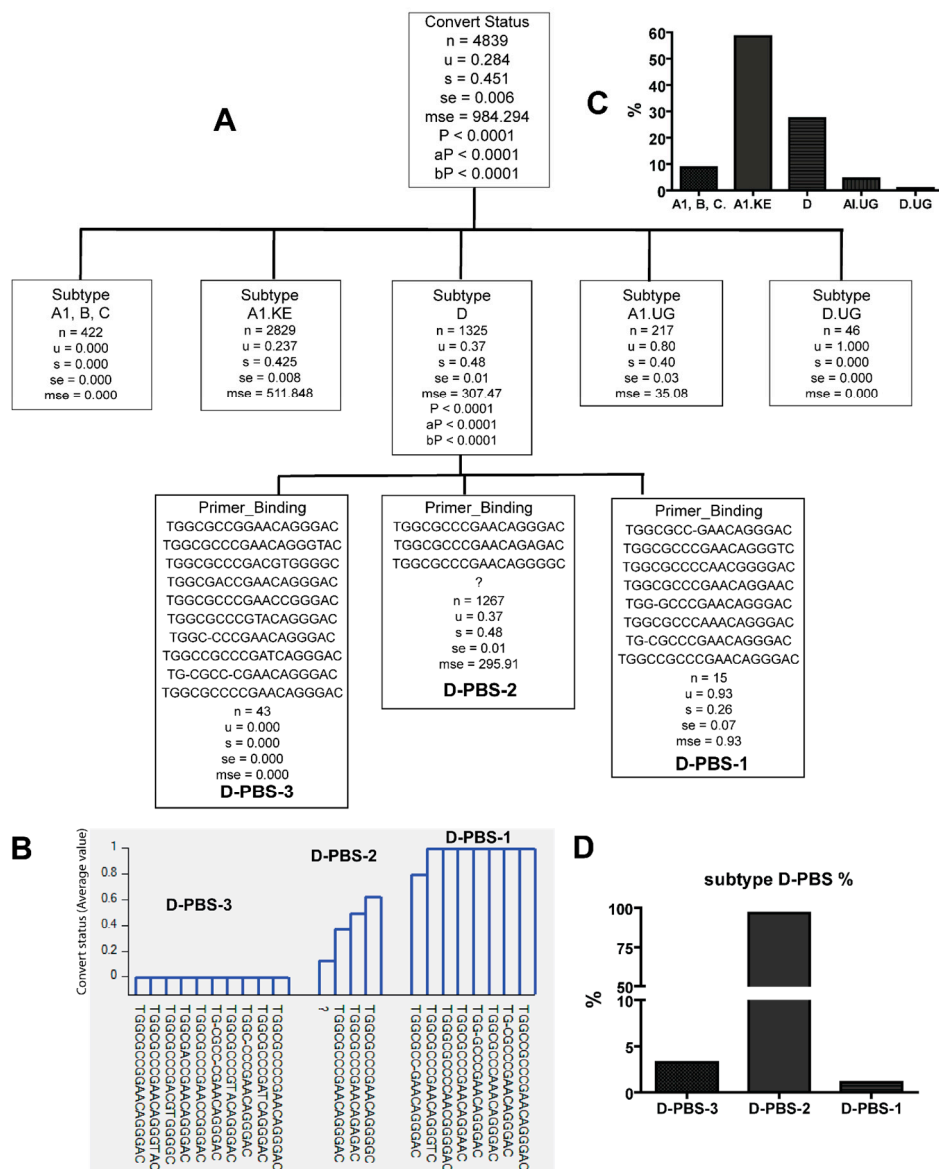


3.3. Combinations of Subtype A1.KE or D with Unique PBS and SD Sequence Variants in Late Seroconverters

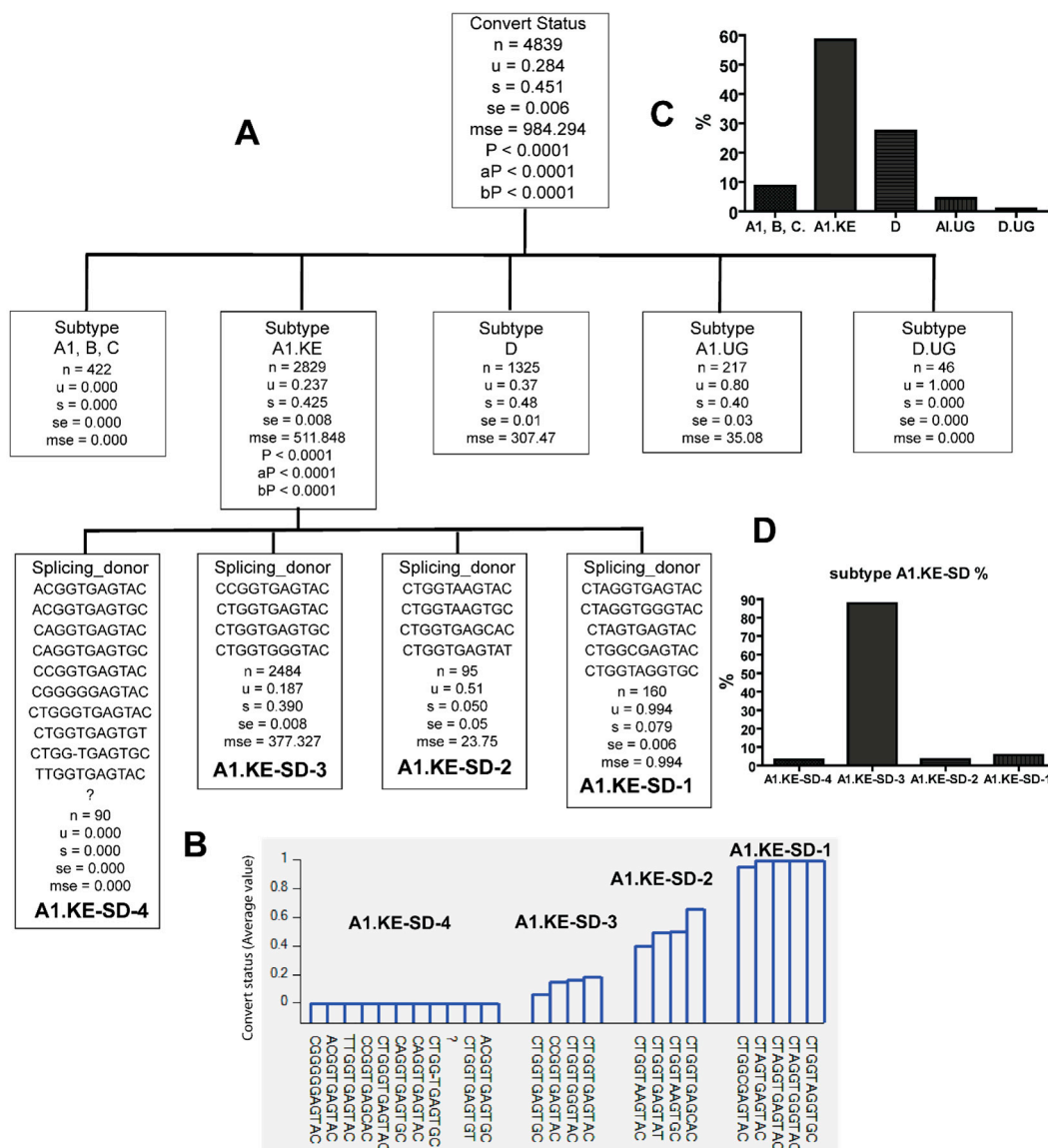
The late seroconverters are most likely to be infected with HIV variants with 5’LTR sequences belonging to A1.UG and D.UG, and specific PBS, SD, and PS variants are only identified in the viral population infecting the late seroconverters. However, late seroconverters were also infected with A1.KE and D, the two major HIV subtypes circulating in Kenya. Are there unique PBS, SD, and PS sequence variants in A1.KE and D infecting late seroconverters? The recursive analysis showed that specific SD variants or PBS variants in subtype D were identified only in late seroconverters or early seroconverters (Figures 8 and 9, and Table 3). Specific SD variants in A1.KE were only identified in late seroconverters or early seroconverters (Figure 10, and Table 3). Thus, A1.KE or D with specific PBS and SD variants infect late seroconverters.



**Figure 8.** The combination of HIV subtype D and splicing donor sequences (SD) identified only in either late ( $u = 1$ ) or early seroconverters ( $u = 0$ ). (A) HIV subtype D in tree node of subtype D was further classified with SD sequence variants based on whether they were identified from late ( $u = 1$ ) or early seroconverters ( $u = 0$ ); (B) the figure shows  $u$  value of sequence combinations based on whether the subtype D–SD combinations were identified from late (1) or early (0) seroconverters. Note:  $n$ —counts (clone);  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value; (C) HIV subtype frequency; (D) Subtype D–SD frequencies in subtype D.



**Figure 9.** The combination of HIV subtype D and primer binding sequences (PBS) identified only in either late ( $u = 1$ ) or early seroconverters ( $u = 0$ ). (A) HIV subtype D in tree node of subtype D was further classified with PBS sequence variants, based on whether they were identified from late or early seroconverters. (B) The figure shows  $u$  value of sequence combinations based on whether the subtype D–PBS combinations were identified from late (1) or early (0) seroconverters. Note:  $n$ —counts (clone);  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value; “?” denotes the lack of sequence. (C) HIV subtype frequency. (D) Subtype D–PBS frequencies in subtype D.



**Figure 10.** The combination of HIV subtype A1.KE and splicing donor sequences (SD) identified SD sequences only in either late ( $u = 1$ ) or early seroconverters ( $u = 0$ ). (A) HIV subtype A1.KE in tree node of subtype A1.KE was further classified with SD sequence variants based on whether they were identified from late or early seroconverters. (B) The figure shows  $u$  value of sequence combinations based on whether the subtype A1.KE-SD combinations were identified from late (1) or early (0) seroconverters. Note:  $n$ —counts (clone);  $u$ —mean value;  $s$ —standard deviation;  $se$ —standard error;  $mse$ —mean square error;  $p$ — $p$  value;  $aP$ —adjusted  $p$  value;  $bP$ —Bonferroni corrected  $p$  value; “?”: denotes lack of sequence. (C) HIV subtype frequency. (D) Subtype A1.KE-SD frequencies in subtype A1.KE.

Our study showed that late seroconverters are more likely to be infected with A1 and D from Uganda, and specific PBS, SD, and PS sequences were only identified in the late seroconverters. Also, A1.KE and D with specific PBS and/or SD variants are also likely to infect late seroconverters. Table 4 summarized the identified 5’LTR subtypes, PBS, SD, PS variants, and the combinations identified and enriched in the 20 late seroconverters. These identified 5’LTR subtypes, PBS, SD, PS, and their combinations were identified and enriched in 16 out 20 late seroconverters (Table 4). The subtype classification of 5’LTR-leader sequence of viruses infecting late seroconverters is shown in Table 5.



**Table 4.** Later seroconverters with HIV subtypes, PBS, SD, PS variants that are enriched or only identified in LSC.

mlno	PBS1	SD1	PS1	PBS-2/SD-1	PBS2/A1.UG	PBS-3/SD-1	PBS-3/SD-2	SD2	A1-UG/D.UG	A1.KE/SD1	D-SD-1	D-PBS-1
37			+									
58	+	+				+		+	+			
290				+	+				+			
452				+			+	+				
546		+				+				+		
768	+		+	+					+			
814		+				+				+		
825												
888	+	+	+	+		+					+	+
890	+		+	+			+	+				
1072				+	+				+			
1102		+	+	+		+						
1232							+	+				
1248												
1250	+			+				+				+
1287		+		+		+		+			+	
1430				+								
1626			+				+	+				
1707												
1730												

Note: '+' denotes presence of indicated sequences or subtypes, or their combinations.

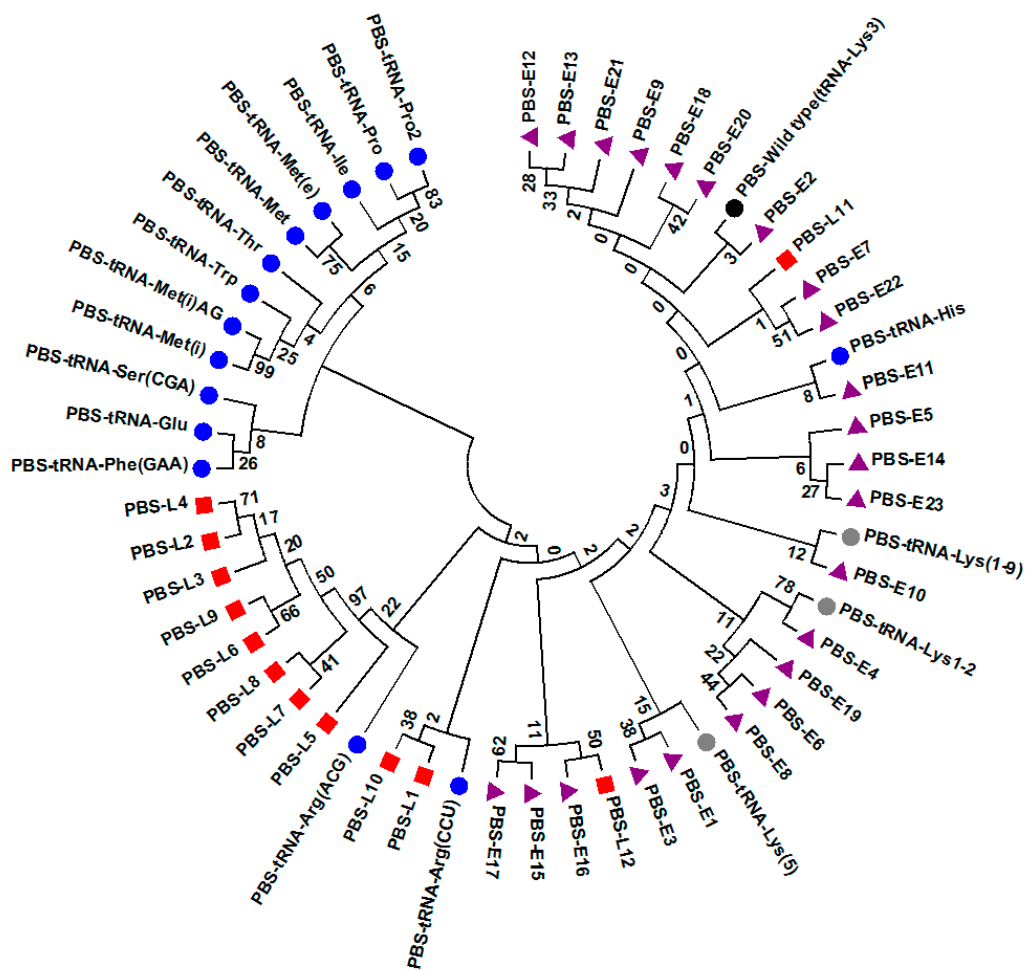
**Table 5.** HIV-1 subtype classification of 5'LTR-leader variants of viruses infecting late seroconverters.

mlno	A1.KE	A1.UG	D	D.UG
37	*			
58				*
290		*		
452	*			
546	*			
768	*	*		
814	*			
825	*		*	
888			*	
890	*			
1072	*	*		
1102			*	
1232	*			
1248			*	
1250			*	
1287			*	
1430	*		*	
1626	*			
1707	*			
1730	*			

Note: '\*' denotes presence of indicated subtypes.

### 3.4. Potential Functional Differences among PBS Variants in Late Seroconverters

Among the three sites studied, only PBS had sufficient supporting literature available to permit analysis for their potential functional significance. A phylogenetic tree was constructed containing PBS variant sequences only identified in late or early seroconverters, together with 19 PBS reference sequences that have been studied for their function (Figure 11). With the exception of tRNA<sup>Lys,3</sup> and tRNA<sup>Lys,5a</sup>, none of the other tRNA molecules have been reported to be used as primers in naturally occurring HIV-1. Phylogenetic analysis showed that majority of the PBS sequence variants identified only in late seroconverters (PBS-1) co-clustered with PBS reference sequences utilizing tRNA<sup>Arg</sup> molecules. Whereas, the PBS sequence variants identified only in early seroconverters (PBS-4) co-clustered with PBS wild type references PBS-tRNA<sup>Lys,3</sup> and its variants PBS-tRNA<sup>Lys1-9</sup>, PBS-tRNA<sup>Lys1,2</sup>, PBS-tRNA<sup>Lys(5)</sup>, and PBS-tRNA<sup>His</sup> (Figure 11).



**Figure 11.** Molecular phylogenetic analysis of PBS sequence variants identified only in late (PBS-1 as PBS-L) or early (PBS-4 as PBS-E) seroconverters by maximum likelihood method with reference sequences.

The evolutionary history was inferred by using the maximum likelihood method based on the Tamura–Nei model [1]. The bootstrap consensus tree inferred from 1000 replicates [2] is taken to represent the evolutionary history of the taxa analyzed [2]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. Initial tree(s) for the heuristic search were obtained automatically by applying neighbor-join and BioNJ algorithms to a matrix of pairwise distances estimated using the maximum composite likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The analysis involved 54 nucleotide sequences. There was a total of 22 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

Note: reference sequences are marked with colored filled circles. The PBS sequences identified only from late seroconverters (PBS-1 as PBS-L) are marked with red filled square. The PBS sequences identified only from early seroconverters (PBS-4 as PBS-E) are marked with purple filled triangle.

#### 4. Discussion

The outcome of exposure to HIV-1 is influenced by both host as well as pathogen derived genetic factors. HIV-1 late seroconversion has been observed in Pumwani sex worker cohort. Here, we investigate whether the late seroconversion is associated with specific subtypes and 5′LTR-leader sequence variants in this epidemiologically well-characterized cohort. We showed that the 5′LTR-leader sequence variants are dominated by clade A1 and D viruses in this cohort, and this is consistent with previous studies of Kenyan HIV infected patients [20–22]. We observed a significant difference in HIV-1

subtype distribution between late seroconverters and the early seroconverters. A significantly higher proportion of late seroconverters were infected by subtype A1 and D from Uganda. Two possibilities may explain this observation. One, viral subtypes from Uganda may differ in its ability to cause infection and exhibit superior replicative properties. Two, the late seroconverters may be infected while they were back in their home village during a break from sex work [17]. As none of the late seroconverters were from Uganda, it is possible that the migration of their clientele between Uganda and Kenya was responsible for the transmission of subtype A1.UG and D.UG. The predominance of the Uganda subtype in the late seroconverter population suggests a relationship between Ugandan viral origin and late seroconversion. HIV-1 subtypes originating in Uganda may be more infectious than their Kenyan counterparts, and comparative infectivity studies will need to be carried out to confirm this possibility. Moreover, the rates of disease progression of patients infected with Ugandan A and D subtypes could be examined and compared with that of patients infected with Kenyan subtypes A and D. In addition, other genetic factors unique to subtypes A1.UG and D.UG might play an important role in HIV-1 late seroconversion.

We also showed that unique sequence variants of PBS, SD, and PS exist in viruses infecting late seroconverters. Specific SD sequences were identified only in viruses from late seroconverters or early seroconverters. SD is essential to all splicing events in HIV-1 [36], and as such, the association of specific SD sequence variants with late seroconverters deserves specific attention. Functional studies, currently lacking, could address whether these specific SD sequence variants exhibit more efficient splicing activity. SD, PBS, and PS each have different roles in HIV-1 replication [28,29]. Our study showed that combinations of sequence variants from these sites associated significantly with late seroconverters or with the early seroconverters, suggesting a synergistic effect between these three functional sites. This appears also true in the combination of A1.KE or D with specific PBS and SD sequence variants infecting late or early seroconverters. Thus, both viral subtypes and PBS, SD, and PS sequence variants, play a role in late seroconversion. The interplay between the sequence variants of these sites and their effect on HIV-1 exposure outcome is not clear, and warrants further functional investigations.

Studies have shown that most of HIV viruses, including proviral sequences and virions in plasma samples, were defective. Our study is limited to the analysis of 5'LTR leader sequences; these diverse sequences may be associated with defective or non-defective HIV viruses. The identification of the specific PBS, SD, or PS variants, that exist only in LSC or EC, may provide a reasonable base to further investigate whether these specific sequence variants actually play a more important role in viral pathogenesis than the ones indicated by their population frequencies. In addition, studies have shown that defective viruses are known to drive HIV infection, persistence, and pathogenesis [58], and the data from our study provide another aspect of HIV pathogenesis.

Earlier studies done in our cohort suggested viral cytotoxic T lymphocyte escape variants were not likely to be the primary factors influencing HIV-1 late seroconversion, and pointed out potential links between loss or waning of HIV-1 epitope-specific responses after a break from sex work and late seroconversion [17]. The present study explored the phenomenon of late seroconversion further, and suggests that the process need not purely be immunological; virological factors, viz, PBS, SD1, PS variants and subtypes, could play important roles.

Analysis of potential functional implications of the PBS variant that were only identified in late or early seroconverters, based on the published data, showed that most of the PBS variants identified only in late seroconverters co-clustered with PBS sequence variants using tRNA<sup>Arg</sup> as a primer for reverse transcription, whereas the PBS variants identified only in early seroconverters were co-clustered with the wild type PBS sequences using tRNA<sup>Lys,3</sup>, tRNA<sup>Lys</sup> variants, or tRNA<sup>His</sup> as a primer for reverse transcription. Studies have shown that HIV can replicate using either tRNA<sup>His</sup> or tRNA<sup>Lys1,2</sup> as primers [59–64], however, HIV mutants that use reverse transcription primers other than tRNA<sup>Lys,3</sup> have reduced replication [65]. The only retrovirus that has been reported to use tRNA<sup>Arg</sup> as a primer for reverse transcription is MuLV [66,67]. Analysis of the replication and stability of MuLVs with alternative PBSs revealed a preference for a PBS complementary to tRNA<sup>Pro</sup>, tRNA<sup>Gly</sup>, or tRNA<sup>Arg</sup> [67].

The selection of tRNA<sup>Arg</sup> for MuLV was probably facilitated, in part, by the multiple isoacceptors for tRNA<sup>Arg</sup> [67]. Our study is the first to report that HIV PBS sequence variants identified only in late seroconverters, co-cluster with PBS sequences utilizing tRNA<sup>Arg</sup> as a primer for reverse transcription. The PBS variants do not appear to belong to one specific subtype by interaction analysis (data not shown). Studies have shown that primer selection and viral translation, in particular, the synthesis of Gag-Pol, are linked [66,67]. How these specific HIV PBS variants, clustering with PBS sequences using tRNA<sup>Arg</sup> as a primer for reverse transcription, contribute to the infection of women who were relatively resistant to HIV-1 infection, needs to be investigated.

The current study intends to investigate viral factors influencing HIV-1 late seroconversion observed in the Pumwani cohort. It is clear that the viral subtypes, as well as PBS, SD, and PS variants within the 5' leader sequence, are associated with this clinical outcome, underscoring the importance of viral factors in the late seroconversion. Viral genotypes have been shown to exert profound influence over HIV-1 viral load [68]. Understanding why viruses of certain clades exhibit seemingly more infectiousness and pathogenicity will provide us with valuable information that could be used to help prevent HIV-1 infection. There is also a potential application for this knowledge to be used as clinical predictors that can serve to guide treatment decisions for patients. Successful inhibition of HIV-1 replication through small interfering RNA targeted to the PBS has been reported [69]. RNA transcripts containing HIV-1 PS sequences as HIV-1 antivirals have been explored [70]. To our knowledge, this is the first report of association of 5' LTR-leader sequence variation with HIV-1 late seroconversion, in addition to reporting the specific sequence variations in 5' leader sequence region. The association of PBS, SD, and PS variants with LSC or EC identified in this study may help to find additional pharmaceutical targets, aiding the development of new anti-HIV therapeutics and HIV/AIDS prevention strategies.

**Supplementary Materials:** The following are available online at [www.mdpi.com/1999-4915/10/1/4/s1](http://www.mdpi.com/1999-4915/10/1/4/s1).

**Acknowledgments:** This study was supported by funds from National Microbiology Laboratory, Public Health Agency of Canada. We thank Ian McArthur for his contribution to cloning and sequencing the 5' LTR-LS and data analysis. He received B.Sc Medical Research support from University of Manitoba. We are thankful to Tony Kariri for maintaining the database of the cohort at the University of Nairobi. We thank the nurses and staff working for the Pumwani sex worker cohort (Jane Njoki, Jane Kamene, Elizabeth Bwibo, Edith Amatiwa). We are grateful to women enrolled in Pumwani sex worker cohort for their dedication and continued participation.

**Author Contributions:** Raghavan Sampathkumar Conducted data analysis, drafted and revised the manuscript; Joel Scott-Herridge conducted viral sequencing and data analysis; Binhua Liang conducted data analysis, revised the manuscript; Joshua Kimani maintenance of study cohort, sample and biological data collection; Francis A. Plummer establishment and maintenance of study cohort, supported the study, reviewed and revised the manuscript; Ma Luo designed the study, conducted data analysis, drafted and revised the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. UNAIDS Data 2017. Available online: [http://www.unaids.org/sites/default/files/media\\_asset/20170720\\_Data\\_book\\_2017\\_en.pdf](http://www.unaids.org/sites/default/files/media_asset/20170720_Data_book_2017_en.pdf) (accessed on 5 October 2017).
2. Cohen, J. The emerging race to cure HIV infections. *Science* **2011**, *332*, 784–785, 787–789. [[CrossRef](#)] [[PubMed](#)]
3. Lakhashe, S.K.; Silvestri, G.; Ruprecht, R.M. No acquisition: A new ambition for HIV vaccine development? *Curr. Opin. Virol.* **2011**, *1*, 246–253. [[CrossRef](#)] [[PubMed](#)]
4. Talbot, J.R. Size matters: The number of prostitutes and the global HIV/AIDS pandemic. *PLoS ONE* **2007**, *2*, e543. [[CrossRef](#)] [[PubMed](#)]
5. Baral, S.; Beyrer, C.; Muessig, K.; Poteat, T.; Wirtz, A.L.; Decker, M.R.; Sherman, S.G.; Kerrigan, D. Burden of HIV among female sex workers in low-income and middle-income countries: A systematic review and meta-analysis. *Lancet Infect. Dis.* **2012**, *12*, 538–549. [[CrossRef](#)]
6. Fowke, K.R.; Nagelkerke, N.J.; Kimani, J.; Simonsen, J.N.; Anzala, A.O.; Bwayo, J.J.; MacDonald, K.S.; Ngugi, E.N.; Plummer, F.A. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* **1996**, *348*, 1347–1351. [[CrossRef](#)]

7. Luo, M.; McLaren, P.J.; Plummer, F.A. (Eds.) Host Genetics and Resistance to HIV-1 infection. In *Models of Protection against HIV/SIV*; Academic Press, Elsevier Inc.: Amsterdam, The Netherlands, 2012.
8. Plummer, F.A.; Ball, T.B.; Kimani, J.; Fowke, K.R. Resistance to HIV-1 infection among highly exposed sex workers in Nairobi: What mediates protection and why does it develop? *Immunol. Lett.* **1999**, *66*, 27–34. [[CrossRef](#)]
9. Hardie, R.A.; Knight, E.; Bruneau, B.; Semeniuk, C.; Gill, K.; Nagelkerke, N.; Kimani, J.; Wachihi, C.; Ngugi, E.; Luo, M.; et al. A common human leucocyte antigen-DP genotype is associated with resistance to HIV-1 infection in Kenyan sex workers. *AIDS* **2008**, *22*, 2038–2042. [[CrossRef](#)] [[PubMed](#)]
10. Hardie, R.A.; Luo, M.; Bruneau, B.; Knight, E.; Nagelkerke, N.J.; Kimani, J.; Wachihi, C.; Ngugi, E.N.; Plummer, F.A. Human leukocyte antigen-DQ alleles and haplotypes and their associations with resistance and susceptibility to HIV-1 infection. *AIDS* **2008**, *22*, 807–816. [[CrossRef](#)] [[PubMed](#)]
11. Lacap, P.A.; Huntington, J.D.; Luo, M.; Nagelkerke, N.J.; Bielawny, T.; Kimani, J.; Wachihi, C.; Ngugi, E.N.; Plummer, F.A. Associations of human leukocyte antigen DRB with resistance or susceptibility to HIV-1 infection in the Pumwani sex worker cohort. *AIDS* **2008**, *22*, 1029–1038. [[CrossRef](#)] [[PubMed](#)]
12. Luo, M.; Daniuk, C.A.; Diallo, T.O.; Capina, R.E.; Kimani, J.; Wachihi, C.; Kimani, M.; Bielawny, T.; Peterson, T.; Mendoza, M.G.; et al. For protection from HIV-1 infection, more might not be better: A systematic analysis of HIV Gag epitopes of two alleles associated with different outcomes of HIV-1 infection. *J. Virol.* **2012**, *86*, 1166–1180. [[CrossRef](#)] [[PubMed](#)]
13. Luo, M.; Sainsbury, J.; Tuff, J.; Lacap, P.A.; Yuan, X.Y.; Hirbod, T.; Kimani, J.; Wachihi, C.; Ramdahn, S.; Bielawny, T.; et al. A genetic polymorphism of FREM1 is associated with resistance against HIV infection in the Pumwani sex worker cohort. *J. Virol.* **2012**, *86*, 11899–11905. [[CrossRef](#)] [[PubMed](#)]
14. Peterson, T.A.; Kimani, J.; Wachihi, C.; Bielawny, T.; Mendoza, L.; Thavaneswaran, S.; Narayansingh, M.J.; Kariri, T.; Liang, B.; Ball, T.B.; et al. HLA class I associations with rates of HIV-1 seroconversion and disease progression in the Pumwani sex worker cohort. *Tissue Antigens* **2013**, *81*, 93–107. [[CrossRef](#)] [[PubMed](#)]
15. Price, H.; Lacap, P.; Tuff, J.; Wachihi, C.; Kimani, J.; Ball, T.B.; Luo, M.; Plummer, F.A. A TRIM5 $\alpha$  exon 2 polymorphism is associated with protection from HIV-1 infection in the Pumwani sex worker cohort. *AIDS* **2010**, *24*, 1813–1821. [[CrossRef](#)] [[PubMed](#)]
16. Shea, P.R.; Shianna, K.V.; Carrington, M.; Goldstein, D.B. Host genetics of HIV acquisition and viral control. *Annu. Rev. Med.* **2013**, *64*, 203–217. [[CrossRef](#)] [[PubMed](#)]
17. Kaul, R.; Rowland-Jones, S.L.; Kimani, J.; Dong, T.; Yang, H.B.; Kiama, P.; Rostron, T.; Njagi, E.; Bwayo, J.J.; MacDonald, K.S.; et al. Late seroconversion in HIV-resistant Nairobi prostitutes despite pre-existing HIV-specific CD8<sup>+</sup> responses. *J. Clin. Investig.* **2001**, *107*, 341–349. [[CrossRef](#)] [[PubMed](#)]
18. Sharp, P.M.; Hahn, B.H. Origins of HIV and the AIDS pandemic. *Cold Spring Harb. Perspect. Med.* **2011**, *1*, a006841. [[CrossRef](#)] [[PubMed](#)]
19. HIV Database. Available online: <http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp> (accessed on 5 October 2017).
20. Dowling, W.E.; Kim, B.; Mason, C.J.; Wasunna, K.M.; Alam, U.; Elson, L.; Birx, D.L.; Robb, M.L.; McCutchan, F.E.; Carr, J.K. Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. *AIDS* **2002**, *16*, 1809–1820. [[CrossRef](#)] [[PubMed](#)]
21. Nyamache, A.K.; Muigai, A.W.; Khamadi, S.A. Circulating trends of non-B HIV type 1 subtypes among Kenyan individuals. *AIDS Res. Hum. Retrovir.* **2013**, *29*, 400–403. [[CrossRef](#)] [[PubMed](#)]
22. Peters, H.O.; Mendoza, M.G.; Capina, R.E.; Luo, M.; Mao, X.; Gubbins, M.; Nagelkerke, N.J.; Macarthur, I.; Sheardown, B.B.; Kimani, J.; et al. An integrative bioinformatic approach for studying escape mutations in human immunodeficiency virus type 1 gag in the Pumwani sex worker cohort. *J. Virol.* **2008**, *82*, 1980–1992. [[CrossRef](#)] [[PubMed](#)]
23. Arien, K.K.; Vanham, G.; Arts, E.J. Is HIV-1 evolving to a less virulent form in humans? *Nat. Rev. Microbiol.* **2007**, *5*, 141–151. [[CrossRef](#)] [[PubMed](#)]
24. Brenner, B.G. Resistance and viral subtypes: How important are the differences and why do they occur? *Curr. Opin. HIV AIDS* **2007**, *2*, 94–102. [[CrossRef](#)] [[PubMed](#)]
25. Kanki, P.J.; Hamel, D.J.; Sankale, J.L.; Hsieh, C.; Thior, I.; Barin, F.; Woodcock, S.A.; Gueye-Ndiaye, A.; Zhang, E.; Montano, M.; et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J. Infect. Dis.* **1999**, *179*, 68–73. [[CrossRef](#)] [[PubMed](#)]

26. Vasan, A.; Renjifo, B.; Hertzmark, E.; Chaplin, B.; Msamanga, G.; Essex, M.; Fawzi, W.; Hunter, D. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin. Infect. Dis.* **2006**, *42*, 843–852. [[CrossRef](#)] [[PubMed](#)]
27. Yuste, E.; Borderia, A.V.; Domingo, E.; Lopez-Galindez, C. Few mutations in the 5' leader region mediate fitness recovery of debilitated human immunodeficiency type 1 viruses. *J. Virol.* **2005**, *79*, 5421–5427. [[CrossRef](#)] [[PubMed](#)]
28. Berkhout, B. Structure and function of the human immunodeficiency virus leader RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **1996**, *54*, 1–34. [[PubMed](#)]
29. Lu, K.; Heng, X.; Summers, M.F. Structural determinants and mechanism of HIV-1 genome packaging. *J. Mol. Biol.* **2011**, *410*, 609–633. [[CrossRef](#)] [[PubMed](#)]
30. Rhim, H.; Park, J.; Morrow, C.D. Deletions in the tRNA(lys) primer-binding site of human immunodeficiency virus type 1 identify essential regions for reverse transcription. *J. Virol.* **1991**, *65*, 4555–4564. [[PubMed](#)]
31. Harrich, D.; Hooker, B. Mechanistic aspects of HIV-1 reverse transcription initiation. *Rev. Med. Virol.* **2002**, *12*, 31–45. [[CrossRef](#)] [[PubMed](#)]
32. Das, A.T.; Klaver, B.; Berkhout, B. Sequence variation of the human immunodeficiency virus primer-binding site suggests the use of an alternative tRNA(lys) molecule in reverse transcription. *J. Gen. Virol.* **1997**, *78 Pt 4*, 837–840. [[CrossRef](#)] [[PubMed](#)]
33. Mak, J.; Kleiman, L. Primer tRNAs for reverse transcription. *J. Virol.* **1997**, *71*, 8087–8095. [[PubMed](#)]
34. Ni, N.; Morrow, C.D. Impact of forced selection of tRNAs on HIV-1 replication and genome stability highlight preferences for selection of certain tRNAs. *Virus Res.* **2007**, *124*, 29–37. [[CrossRef](#)] [[PubMed](#)]
35. Abbink, T.E.; Berkhout, B. RNA structure modulates splicing efficiency at the human immunodeficiency virus type 1 major splice donor. *J. Virol.* **2008**, *82*, 3090–3098. [[CrossRef](#)] [[PubMed](#)]
36. Lutzelberger, M.; Reinert, L.S.; Das, A.T.; Berkhout, B.; Kjemis, J. A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. *J. Biol. Chem.* **2006**, *281*, 18644–18651. [[CrossRef](#)] [[PubMed](#)]
37. Zeffman, A.; Hassard, S.; Varani, G.; Lever, A. The major HIV-1 packaging signal is an extended bulged stem loop whose structure is altered on interaction with the Gag polyprotein. *J. Mol. Biol.* **2000**, *297*, 877–893. [[CrossRef](#)] [[PubMed](#)]
38. Miele, G.; Mouland, A.; Harrison, G.P.; Cohen, E.; Lever, A.M. The human immunodeficiency virus type 1 5' packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J. Virol.* **1996**, *70*, 944–951. [[PubMed](#)]
39. Kumar, S.; Tamura, K.; Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* **2004**, *5*, 150–163. [[CrossRef](#)] [[PubMed](#)]
40. Abbink, T.E.; Beerens, N.; Berkhout, B. Forced selection of a human immunodeficiency virus type 1 variant that uses a non-self tRNA primer for reverse transcription: Involvement of viral RNA sequences and the reverse transcriptase enzyme. *J. Virol.* **2004**, *78*, 10706–10714. [[CrossRef](#)] [[PubMed](#)]
41. Das, A.T.; Vink, M.; Berkhout, B. Alternative tRNA priming of human immunodeficiency virus type 1 reverse transcription explains sequence variation in the primer-binding site that has been attributed to APOBEC3G activity. *J. Virol.* **2005**, *79*, 3179–3181. [[CrossRef](#)] [[PubMed](#)]
42. Kelly, N.J.; Morrow, C.D. Yeast tRNA(Phe) expressed in human cells can be selected by HIV-1 for use as a reverse transcription primer. *Virology* **2003**, *313*, 354–363. [[CrossRef](#)]
43. Moore, K.L.; Kosloff, B.R.; Kelly, N.J.; Kirkman, R.L.; Dupuy, L.C.; McPherson, S.; Morrow, C.D. HIV type 1 that select tRNA(His) or tRNA(Lys1,2) as primers for reverse transcription exhibit different infectivities in peripheral blood mononuclear cells. *AIDS Res. Hum. Retrovir.* **2004**, *20*, 373–381. [[CrossRef](#)] [[PubMed](#)]
44. Djekic, U.V.; Morrow, C.D. Analysis of the replication of HIV-1 forced to use tRNAMet(i) supports a link between primer selection, translation and encapsidation. *Retrovirology* **2007**, *4*, 10. [[CrossRef](#)] [[PubMed](#)]
45. Li, M.; Eipers, P.G.; Ni, N.; Morrow, C.D. HIV-1 designed to use different tRNAGln isoacceptors prefers to select tRNAThr for replication. *Virol. J.* **2006**, *3*, 80. [[CrossRef](#)] [[PubMed](#)]
46. McCulley, A.; Morrow, C.D. Nucleotides within the anticodon stem are important for optimal use of tRNA(Lys,3) as the primer for HIV-1 reverse transcription. *Virology* **2007**, *364*, 169–177. [[CrossRef](#)] [[PubMed](#)]
47. Moore-Rigdon, K.L.; Kosloff, B.R.; Kirkman, R.L.; Morrow, C.D. Preferences for the selection of unique tRNA primers revealed from analysis of HIV-1 replication in peripheral blood mononuclear cells. *Retrovirology* **2005**, *2*, 21. [[CrossRef](#)] [[PubMed](#)]

48. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323–348. [[CrossRef](#)] [[PubMed](#)]
49. Hawkins, D.M. *FIRM: Formal Inference-Based Recursive Modeling, Release 2*; Technical Report; School of Statistics, University of Minnesota: Saint Paul, MN, USA, 1995; p. 546.
50. Hawkins, D.M. Fitting multiple change-points to data. *Comput. Stat. Data Anal.* **2002**, *37*, 323–341. [[CrossRef](#)]
51. Hawkins, D.M.; McKenzie, D.P. A data-based comparison of some recursive partitioning procedures. In *Proceedings of the Statistical Computing Section*; American Statistical Association: Boston, MA, USA, 1995; pp. 245–252.
52. Hawkins, D.M.; Musser, B.J. One tree or a forest? Alternative dendrographic models. *Comput. Sci. Stat.* **1999**, *30*, 534–542.
53. Hawkins, D.M.; Musser, B.J. Feature selection with nondeterministic recursive partitioning. In *Proceedings of the American Statistical Association, Joint Statistical Meetings, Atlanta, GA, USA, 2–9 August 2001*.
54. Hawkins, D.M.; Young, S.S.; Rusinko, A. Analysis of a large structure-activity data set using recursive partitioning. *Quant. Struct. Act. Relat.* **1997**, *16*, 296–302. [[CrossRef](#)]
55. Young, S.S.; Hawkins, D.M. Using recursive partitioning to analyze a large SAR data set. *SAR QSAR Environ. Res.* **1998**, *8*, 183–193. [[CrossRef](#)]
56. Young, S.S.; Hawkins, D.M. Using recursive partitioning analysis to evaluate compound selection methods. *Methods Mol. Biol.* **2004**, *275*, 317–334. [[PubMed](#)]
57. Douglas, M. Hawkins LLC website. Available online: <http://www.douglashawkins.com> (accessed on 21 December 2017).
58. Finzi, D.; Plaeger, S.F.; Dieffenbach, C.W. Defective Virus Drives Human Immunodeficiency Virus Infection, Persistence, and Pathogenesis. *Clin. Vaccine Immunol.* **2006**, *13*, 715–721. [[CrossRef](#)] [[PubMed](#)]
59. Wakefield, J.K.; Kang, S.M.; Morrow, C.D. Construction of a type 1 human immunodeficiency virus that maintains a primer binding site complementary to tRNA(His). *J. Virol.* **1996**, *70*, 966–975. [[PubMed](#)]
60. Wakefield, J.K.; Morrow, C.D. Mutations within the primer binding site of the human immunodeficiency virus type 1 define sequence requirements essential for reverse transcription. *Virology* **1996**, *220*, 290–298. [[CrossRef](#)] [[PubMed](#)]
61. Wakefield, J.K.; Wolf, A.G.; Morrow, C.D. Human immunodeficiency virus type 1 can use different tRNAs as primers for reverse transcription but selectively maintains a primer binding site complementary to tRNA(3Lys). *J. Virol.* **1995**, *69*, 6021–6029. [[PubMed](#)]
62. Zhang, Z.; Kang, S.M.; LeBlanc, A.; Hajduk, S.L.; Morrow, C.D. Nucleotide sequences within the U5 region of the viral RNA genome are the major determinants for an human immunodeficiency virus type 1 to maintain a primer binding site complementary to tRNA(His). *Virology* **1996**, *226*, 306–317. [[CrossRef](#)] [[PubMed](#)]
63. Li, Y.; Zhang, Z.; Wakefield, J.K.; Kang, S.M.; Morrow, C.D. Nucleotide substitutions within U5 are critical for efficient reverse transcription of human immunodeficiency virus type 1 with a primer binding site complementary to tRNA(His). *J. Virol.* **1997**, *71*, 6315–6322. [[PubMed](#)]
64. Kang, S.M.; Zhang, Z.; Morrow, C.D. Identification of a human immunodeficiency virus type 1 that stably uses tRNA<sup>Lys1,2</sup> rather than tRNA<sup>Lys,3</sup> for initiation of reverse transcription. *Virology* **1999**, *257*, 95–105. [[CrossRef](#)] [[PubMed](#)]
65. Das, A.T.; Klaver, B.; Berkhout, B. Reduced replication of human immunodeficiency virus type 1 mutants that use reverse transcription primers other than the natural tRNA(3Lys). *J. Virol.* **1995**, *69*, 3090–3097. [[PubMed](#)]
66. Palmer, M.T.; McPherson, S.; Morrow, C.D. Murine leukemia virus with a primer-binding site complementary to tRNA<sup>Lys,3</sup> adapts to select new tRNAs for replication following extended in vitro culture. *Virology* **2005**, *333*, 337–348. [[CrossRef](#)] [[PubMed](#)]
67. Palmer, M.T.; Kirkman, R.; Kosloff, B.R.; Eipers, P.G.; Morrow, C.D. tRNA isoacceptor preference prior to retrovirus Gag-Pol junction links primer selection and viral translation. *J. Virol.* **2007**, *81*, 4397–4404. [[CrossRef](#)] [[PubMed](#)]
68. Alizon, S.; von Wyl, V.; Stadler, T.; Kouyos, R.D.; Yerly, S.; Hirschel, B.; Boni, J.; Shah, C.; Klimkait, T.; Furrer, H.; et al. Phylogenetic approach reveals that virus genotype largely determines HIV set-point viral load. *PLoS Pathog.* **2010**, *6*, e1001123. [[CrossRef](#)] [[PubMed](#)]

69. Abbink, T.E.; Berkhout, B. HIV-1 reverse transcription initiation: A potential target for novel antivirals? *Virus Res.* **2008**, *134*, 4–18. [[CrossRef](#)] [[PubMed](#)]
70. Dorman, N.M.; Lever, A.M. Investigation of RNA transcripts containing HIV-1 packaging signal sequences as HIV-1 antivirals: Generation of cell lines resistant to HIV-1. *Gene Ther.* **2001**, *8*, 157–165. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).