Infection, Genetics and Evolution 69 (2019) 146–152

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Sequence ambiguity determined from routine *pol* sequencing is a reliable tool for real-time surveillance of HIV incidence trends

Maja M. Lunar[a], Snježana Židovec Lepej[b], Mario Poljak[a,*]

[a] Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Zaloška 4, Ljubljana 1105, Slovenia
[b] Dr. Fran Mihaljevič University Hospital for Infectious Diseases, Mirogojska 8, Zagreb 10000, Croatia

ARTICLE INFO

Keywords:
HIV-1
Incidence
Sequence ambiguity
Mixed base calls
Surveillance

ABSTRACT

Identifying individuals recently infected with HIV has been of great significance for close monitoring of HIV epidemic dynamics. Low HIV sequence ambiguity (SA) has been described as a promising marker of recent infection in previous studies. This study explores the utility of SA defined as a proportion of ambiguous nucleotides detected in baseline *pol* sequences as a tool for routine real-time surveillance of HIV incidence trends at a national level.

A total of 353 partial HIV-1 *pol* sequences obtained from persons diagnosed with HIV infection in Slovenia from 2000 to 2012 were studied, and SA was reported as a percentage of ambiguous base calls. Patients were characterized as recently infected by examining anti-HIV serological patterns and/or using commercial HIV-1 incidence assays (BED and/or LAg-Avidity assay). A mean SA of 0.29%, 0.14%, and 0.19% was observed for infections classified as recent by BED, LAg, or anti-HIV serological results, respectively. Welch's *t*-test showed a significant difference in the SA of recent versus long-standing infections ($p < 0.001$). CD4+ T-cell counts ≤250 cells/mm^3 significantly correlated with higher SA ($p < 0.001$), whereas the homo/bisexual transmission route significantly correlated with lower SA ($p = 0.005$). When the LAg-assay was used as an indicator of recent infection, a receiver operating characteristic curve with the largest area under the curve (0.896) was observed for SA (sensitivity and specificity of 79%), indicating the best correlation of the data. A reliable estimation of the trends of HIV incident infection could be inferred from measuring SA irrespective of the cutoff used; however, in Slovenia it seems that lower cutoffs are more appropriate.

Our data suggest that SA could be used as a real-time surveillance tool for close monitoring of HIV incidence trends, especially in countries where baseline HIV resistance genotyping is performed routinely, rendering this approach cost-effective.

1. Introduction

Surveillance of the HIV epidemic and monitoring its dynamics requires reliable differentiation of individuals infected with HIV recently (RI) from individuals with a long-standing infection (LSI). This can best be achieved in prospective cohort studies, in which HIV-negative persons are monitored longitudinally for HIV seroconversion, or less ideally retrospectively in a cross-sectional design by using assays able to reliably discriminate RI from LSI in a single specimen as combined under the generic term Serological Testing Algorithm for Recent HIV Seroconversion (STARHS) (Murphy and Parry, 2008). Two of the more widely used assays are the BED test, which quantifies normalized optical density (ODn) resulting from the proportion of HIV-specific IgG antibodies (Calypte Biomedical Corporation, 2008; Parekh et al., 2002), and the LAG assay, which measures the avidity of HIV antibodies and expresses it as ODn (Duong et al., 2012; Kassanjee et al., 2016; Sedia Biosciences Corporation, 2016). However, the STARHS assays are known to have several limitations: the HIV-1 subtype and even clade can affect the window period, and low CD4+ T-cell count or antiretroviral therapy has been shown to yield false-recent

Abbreviations: AUC, area under the curve; BED, Aware BED EIA HIV-1 incidence test (Calypte Biomedical Corporation, Portland, Oregon); DR, drug-resistant; HIV, human immunodeficiency virus; IDU, injecting drug use; LAG, LAg-Avidity assay (Sedia Biosciences Corporation, Portland, Oregon); LSI, long-standing infection; MDRI, mean duration of recent infection; MSM, men who have sex with men; NGS, next-generation sequencing; ODn, normalized optical density; *pol*, polymerase gene; RI, recent infection; ROC, receiver operating characteristic; SA, sequence ambiguity; SER, serological pattern supporting HIV seroconversion; SD, standard deviation; SE, standard error; STARHS, Serological Testing Algorithm for Recent HIV Seroconversion; TDR, transmitted drug resistance

* Corresponding author at: Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia.
E-mail address: mario.poljak@mf.uni-lj.si (M. Poljak).

https://doi.org/10.1016/j.meegid.2019.01.015
Received 31 October 2018; Received in revised form 9 January 2019; Accepted 13 January 2019
Available online 22 January 2019
1567-1348/

Table 1
Baseline characteristics of patients included in the evaluation of SA as a method of differentiating RI from LSI when BED, LAG, or SER were used as the gold standard.

Characteristic		BED *n* (%)		LAG *n* (%)		SER *n* (%)	
Sex							
	Male	233	(89.6)	102	(89.5)	314	(89.2)
	Female	27	(10.4)	12	(10.5)	38	(10.8)
HIV transmission							
	MSM	192	(73.8)	78	(68.4)	256	(72.7)
	Heterosexual/IDU	53	(20.4)	30	(26.3)	76	(21.6)
	Unknown	15	(5.8)	6	(5.3)	20	(5.7)
HIV subtype							
	Subtype B	227	(87.3)	92	(80.7)	301	(85.5)
	Other	33	(12.7)	22	(19.3)	51	(14.5)
HIV viral load (log copies/ml ± SD)		4.97 ± 0.94		4.93 ± 1.07		4.97 ± 0.96	
CD4 cell count (mm^3 ± SD)		309 ± 221		312 ± 247		313 ± 228	

SA, sequence ambiguity method; RI, recent HIV infection; LSI, long-standing HIV infection; BED, the BED test; LAG, the LAg-Avidity assay; SER, anti-HIV serological patterns; MSM, men having sex with men; IDU, injecting drug use; SD, standard deviation.

misclassifications (Murphy and Parry, 2008). Several accuracy improvements have been proposed in the last decade, the most significant being the inclusion of low viral load criteria for the characterization of LSI (Kassanjee et al., 2016).

In addition to the STARHS assays, HIV diversity has been proposed as a promising indicator of RI; it is determined either by sequencing (Sanger or next-generation) or by assays such as high-resolution melting assay and heteroduplex mobility assay (Cousins et al., 2011; Delwart et al., 1997). When analyzing sequences obtained by Sanger population sequencing, the chromatograms can exhibit two or more different peaks on one nucleotide position. This can be due to sequencing error or is caused by the presence of more than one HIV variant. If these mixed (ambiguous) nucleotide codons are confirmed in the chromatograms generated in the forward and reverse manner, these are typically true indicators of sequence ambiguity (SA) and thus genetic diversity of the virus in question. SA is typically expressed as a proportion of ambiguous base calls in the sequence assessed and as such has been examined as a marker of RI (Kouyos et al., 2011). This approach is based on the fact that HIV is normally transmitted by a single viral variant (founder virus) because of the genetic bottleneck (Carlson et al., 2014; Keele et al., 2008). Soon after infection, due to the high mutation rate of the HIV-1 reverse transcriptase (3.4×10^{-5} mutations per base pair per generation), the HIV population gradually diversifies in the form of genetic variants, known as quasispecies (Domingo and Holland, 1997; Mansky and Temin, 1995). It has been shown that SA increases linearly by approximately 0.2% each year in the first 8 years of HIV infection, followed by a plateau, and then decreases in the terminal phase of the disease, thus representing a promising marker for estimating the duration of HIV infection (Kouyos et al., 2011).

HIV diversity was estimated in previous studies from the proportion of ambiguous base calls determined in routinely genotyped HIV sequences of the *pol* region (encompassing protease and reverse transcriptase) with SA cutoffs of < 0.45% to 0.50% to distinguish RI from LSI (Andersson et al., 2013; Kouyos et al., 2011; Ragonnet-Cronin et al., 2012). Andersson et al. (2013) proposed a cutoff of < 0.47% ambiguous nucleotides to characterize RI within 1 year of HIV diagnosis with a sensitivity and specificity of 89% and 75%, respectively (Andersson et al., 2013). On the other hand, Ragonnet-Cronin et al. (2012) aimed to distinguish RI that occurred within 6 months prior to diagnosis via the mean duration of recent infection (MDRI) commonly used in HIV incidence studies and measured by BED test as the preferred HIV incidence assay (MDRI = 155 days). In this study, the SA cutoff of 0.45% best predicted RI (Ragonnet-Cronin et al., 2012).

The prevalence of HIV-transmitted drug resistance (TDR) in Slovenia, a country with centralized HIV care and 843 cases diagnosed by the end of 2017, is well below 5%, and thus baseline HIV genotypic drug resistance testing of newly diagnosed HIV individuals is not performed routinely, with few exceptions: for pregnant women and individuals suspected of acquiring HIV infection abroad. Even so, partial *pol* sequences have been collected in over half of newly diagnosed persons since 2000 as part of the SPREAD study of TDR in Europe (Babič et al., 2006b; Lunar et al., 2013; Lunar et al., 2018).

This study explores the utility of HIV diversity observed in baseline *pol* sequences for estimating the duration of patients' HIV infection and thereby HIV incidence at a national level. We also tested whether the SA method could be used as a real-time surveillance tool for close monitoring of HIV incidence trends.

2. Material and methods

2.1. Study samples

Slovenia is a country with a centralized system of HIV diagnostics, monitoring, and clinical care. Confirmation of HIV infection and monitoring of all HIV patients in the country are performed exclusively at the Slovenian HIV Reference Center at the Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana. Clinical care for all HIV-positive persons is provided at the Department of Infectious Diseases, Ljubljana University Medical Center. Therefore, the clinical and laboratory data collected for this study by the two institutions mentioned above are representative of the entire national HIV cohort. A total of 353 partial HIV-1 *pol* sequences were included in this study. Sequences were obtained from the same number of treatment-naive persons newly diagnosed with HIV from 2000 to 2016 in Slovenia, representing 56% (353/628) of all individuals diagnosed during those 17 years. The individuals included were mostly men who have sex with men (MSM) infected with HIV-1 subtype B (Table 1). Sequences were obtained as a part of the European Commission–sponsored SPREAD program for the study of TDR in Europe, and this study was approved by the national Medical Ethics Committee at the Slovenian Ministry of Health (approval ref. no. 126/12/03; Babič et al., 2006a; Lunar et al., 2013). Patients' samples were anonymized and the study was conducted in accordance with the code of Ethics of the World Medical Association (Declaration of Helsinki). In addition, a total of 1210 results obtained using four different HIV incidence methods on 484 patients' samples (see details below) were included in the time trend analysis of recent HIV-1 infections in Slovenia, representing 77% of HIV persons diagnosed in Slovenia from 2000 to 2016.

2.2. Serology-based incidence estimation methods

Altogether 353 persons with a corresponding baseline partial HIV-1

Fig. 1. Venn diagram classifying the 353 samples included in this study according to i) anti-HIV serological patterns (SER), ii) the BED assay (BED), and iii) the LAg-Avidity assay (LAG). A total of 352 were characterized according to SER, 260 were tested with the BED test, and 114 were tested with the LAG assay.

pol sequence available were classified as either RI or LSI based on the evidence of HIV seroconversion (SER) and/or by using one or two commercial HIV-1 incidence assays on baseline samples (Fig. 1, Table 1). Samples analyzed, either using the incidence assays or sequencing, were obtained within the 3 months following HIV diagnosis. The same sample that was tested with the incidence assays was used to generate *pol* sequences. The baseline characteristics of the patients included in this study are summarized in Table 1.

A total of 352 serum/plasma samples were classified based on the anti-HIV serological patterns observed (Fig. 1). Individuals were determined as RI when at least one of the serological assays for detecting HIV infection (screening or confirmatory) returned an indeterminate or negative result in the baseline sample and the HIV diagnosis was confirmed with either the detection of HIV-1 RNA in the baseline sample or by full seroconversion in the subsequent blood sample(s).

In addition, 348 serum/plasma samples were tested with one or two commercial HIV-1 incidence assays. Samples from 234 patients were tested using the Aware BED EIA HIV-1 Incidence Test (Calypte Biomedical Corporation, Portland, Oregon) (BED), 88 samples with the LAg-Avidity assay (Sedia Biosciences Corporation, Portland, Oregon) (LAG), and 26 samples with both incidence assays, following manufacturers' instructions (Fig. 1). ODn of ≤ 0.8 and ≤ 1.5 was characteristic of RI, corresponding to the MDRI of 155 days and 130 days when using the BED and LAG tests, respectively. An additional 131 samples were tested with the BED test, but baseline *pol* sequences were not obtained. Therefore, the results from these patients could subsequently be included only in the time trend analysis of HIV incident infections, adding up to 391 BED assay results altogether.

RI was detected in 11% (38/352), 30% (119/391), and 21% (24/114) of persons based on SER, BED, and LAG assay results, respectively. An incidence estimation of the complete 2000–2016 period is not feasible, and the three estimates obtained are not comparable because the three methods employ different MDRIs and different patient groups were tested.

2.3. Sequence ambiguity method for determining RI/LSI

Partial *pol* sequences were obtained from treatment-naive persons, as reported previously (Lunar et al., 2013). Briefly, between 2000 and 2006 sequences were obtained using the commercially available kit ViroSeq HIV-1 Genotyping System (Celera Diagnostics, Alameda, CA), which had included sequence assembly software, and between 2007 and 2012 using an in-house sequencing method and Vector NTI Advance v11.0 (Invitrogen, Carlsbad, CA) for sequence assembly. Both protocols included manual inspection of chromatograms, upon which a decision was made on the presence of ambiguous nucleotides. Ambiguous nucleotides were assigned if present in both forward and reverse sequence orientation and relative to the sequence's baseline noise. Sequence diversity was measured by counting ambiguous base calls using the nucleotide composition tool in the Bioedit package (Hall, 1999) and a Python script to filter out and count the ambiguous calls. The observed number of mixed codons was divided by the length of the sequence and reported as a percentage of ambiguous nucleotides. The length of the sequences was on average 1000.75 ± 4.77 nucleotides (standard deviation) with a range of 967 to 1002 base pairs. No low-quality sequences were included because sufficient quality had to be achieved for identification of transmitted drug resistance mutations. The HIV-1 subtype of the *pol* sequences analyzed was determined using the REGA HIV-1 Subtyping Tool, version 3.0 (Pineda-Peña et al., 2013).

The GenBank accession numbers of the sequences included are: AJ971091, AJ971093–AJ971119, AJ971121–AJ971123, AJ971125–AJ971130, AJ971133–AJ971135, AJ971137, AJ971139–AJ971144, AM113750, GQ398934, GQ399003, GQ399157, GQ399167, GQ399210, GQ399318, GQ399406, GQ399433, GQ399494, GQ399553, GQ399574, GQ399677, GQ399709, GQ399721, GQ399731, GQ399787, GQ399882, GQ399950, GQ399979, GQ400015, GQ400033, GQ400039, GQ400057, GQ400283, GQ400355, GQ400410, GQ400411, GQ400442, GQ400452, GQ400472, JX028303–JX028406, JX046402–JX046417, KF753699–KF753751, KP013639–KP013677, and KY656612–KY656674.

2.4. Statistical analysis

A *t*-test was performed using JASP version 0.8.0.0 and receiver operating characteristic (ROC) curve analysis using MedCalc Statistical Software version 16.8.4 (JASP Team, https://jasp-stats.org/, 2016; MedCalc Software bvba, Ostend, Belgium, https://www.medcalc.org, 2016), and $p < 0.05$ was considered significant.

3. Results

The following HIV-1 subtypes were assigned to the 353 *pol* sequences included in this study: subtype B in 83% ($n = 292$), subtype A in 7% ($n = 24$), CRF01_AE in 2% ($n = 6$), CRF02_AG in 2% ($n = 6$), and other subtypes or recombinant forms in 7% ($n = 25$).

3.1. Sequence ambiguity as a tool to discriminate between recent and long-standing infections

First, we determined whether SA is discriminatory enough to differentiate RI from LSI individuals. In the group of RI, a mean SA of 0.29%, 0.14%, and 0.19% was observed when BED, LAG, or SER results were used as the gold standard for RI, respectively (Table 2, Fig. 2). Levene's test of equality of variances was significant, suggesting unequal variance, and so instead of Student's *t*-test Welch's *t*-test was used to compare the groups of RI and LSI. As shown in Table 3, the *t*-test determined that the sequence ambiguity of RI versus LSI differed

Table 2
Characteristics of samples included in the evaluation of SA as a method of differentiating RI from LSI when BED, LAG, or SER were used as the gold standard.

	Group	*n*	Mean SA (%)	SD (%)	SE
BED					
	LSI	177	0.839	0.813	0.061
	RI	83	0.292	0.433	0.047
LAG					
	LSI	90	1.042	0.860	0.091
	RI	24	0.137	0.249	0.051
SER					
	LSI	314	0.789	0.825	0.047
	RI	38	0.192	0.369	0.060

SA, sequence ambiguity method; BED, the BED test; LAG, the LAg-Avidity assay; SER, anti-HIV serological patterns; SD, standard deviation; SE, standard error; RI, recent HIV infection; LSI, long-standing HIV infection.

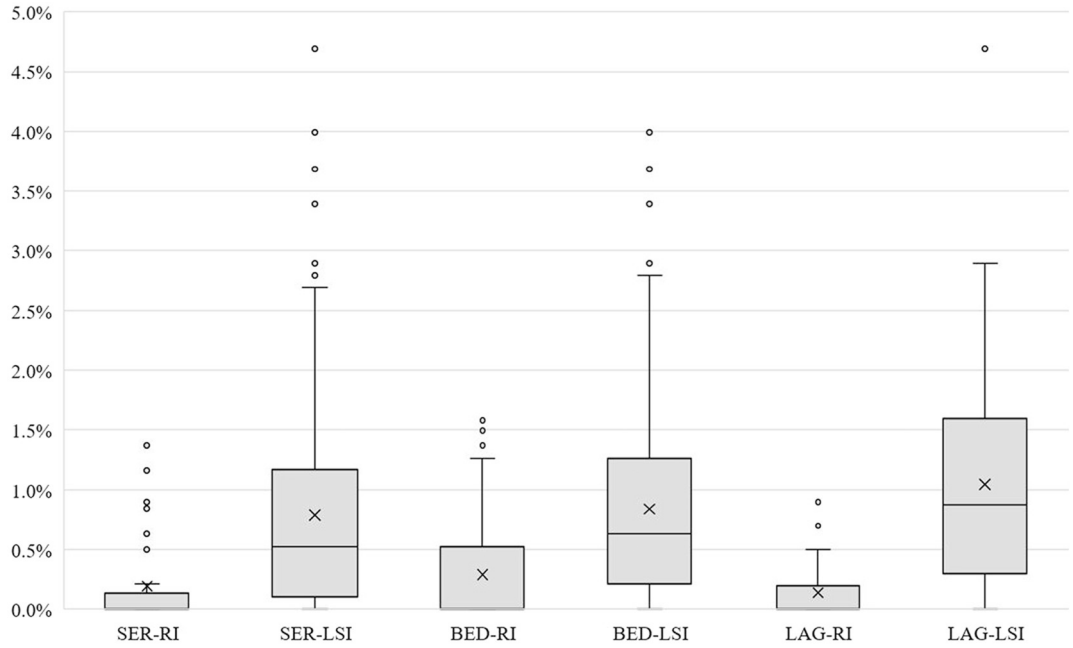

Fig. 2. Box and whiskers plot of the proportion of ambiguous base calls in baseline HIV *pol* sequences for sets of samples analyzed with a duration of HIV infection determined by anti-HIV serological patterns (SER), or using the incidence assays (BED and/or LAG).
BED-RI, recent infection (RI) determined using the BED assay; BED-LSI, long-standing infection (LSI) determined using the BED assay; LAG-RI, RI determined using the LAg-Avidity assay; LAG-LSI, LSI determined using the LAg-Avidity assay; SER-RI, RI determined by anti-HIV serological patterns; SER-LSI, LSI determined by anti-HIV serological patterns.

Table 3
Independent samples *t*-test (Welch's) for the sequence ambiguity method when BED, LAG, or SER were used as the gold standard.

							95% Confidence Interval	
	t	*df*	*p*	Mean difference	SE difference	Cohen's *d*	Lower	Upper
BED	7.077	254.0	< 0.001	0.548	0.077	0.767	0.395	0.700
LAG	8.707	111.2	< 0.001	0.905	0.104	1.168	0.699	1.111
SER	7.874	91.27	< 0.001	0.597	0.076	0.757	0.447	0.748

BED, the BED test; LAG, the LAg-Avidity assay; SER, anti-HIV serological patterns; SD, standard deviation; SE, standard error.

significantly, irrespective of the incidence assay used ($p < 0.001$). Cohen's *d* was observed to be relatively high, indicating that a large sample size was used to test the data. This clear difference in the SA of RI versus LSI can be observed in the box and whiskers plot in Fig. 2. We also tested whether CD4+ T-cell count, HIV-1 viral load, and HIV transmission route influence SA. Welch's independent samples *t*-test showed a significant correlation of CD4+ T-cell count ≤250 cells/mm^3 with a higher SA ($p < 0.001$; mean SA 1.11% and 0.45% for CD4 ≤ 250 cells/mm^3 and CD4 > 250 cells/mm^3, respectively) and MSM transmission route with a lower SA ($p = 0.005$; mean SA 0.64% and 0.94% for MSM and other HIV acquisition, respectively). No significant correlation was found with HIV-1 viral load.

3.2. Performance of sequence ambiguity at different thresholds

We also tested how SA performs at various thresholds when BED, LAG, or SER results were used as the gold standards of RI. ROC curves of each method were obtained as presented in Fig. 3 and in Table 4. The largest area under the curve (AUC) was observed when the LAG test was used as an indicator of RI, with a relatively high AUC of 0.896 (Table 4). According to ROC, optimal SA cutoffs exhibiting the best sensitivity/specificity were set at 0.30%, 0.20%, and 0.11%, when BED, LAG, and SER were used, respectively (Table 4). As already mentioned, SA correlated best with the LAG data and demonstrated a sensitivity of 79% and specificity of 79% when the SA cutoff of 0.2% was applied to discriminate RI from LSI (Table 4). SA ≥ 1% was found among 9/83 (11%), 0/24 (0%), and 2/38 (5%) samples characterized as RI using BED, LAG, or SER, respectively. SA ≥ 0.5% was found among 22/83 (27%), 3/24 (13%), and 8/38 (21%) samples characterized as RI using BED, LAG, or SER, respectively.

3.3. Sequence ambiguity performance in a real-time nationwide setting

In the third part of the study, we tested how SA performs as a discriminator of RI from LSI in a nationwide setting, preferably in real time. The share of RI among persons newly diagnosed with HIV from 2000 to 2016 in Slovenia, as determined by four different methods, is presented in Fig. 4. The BED data were available for 2000–2012 for 391/428 (91.4%) persons altogether and are therefore the most representative of the Slovenian HIV epidemic. The BED assay was replaced with the LAG assay in 2012, and so the LAG data were available for 2012–2016 only. SA results with the optimal cutoffs of 0.11%, 0.2%, and 0.3%, as determined in this study, are shown back-to-back with those using the previously suggested 0.45% cutoff based on BED testing (Ragonnet-Cronin et al., 2012). A reliable estimation of the trends of HIV incident infection could be inferred from measuring SA irrespective of the cutoff used; however, for the Slovenian situation it seemed that lower cutoffs are more appropriate (Fig. 4).

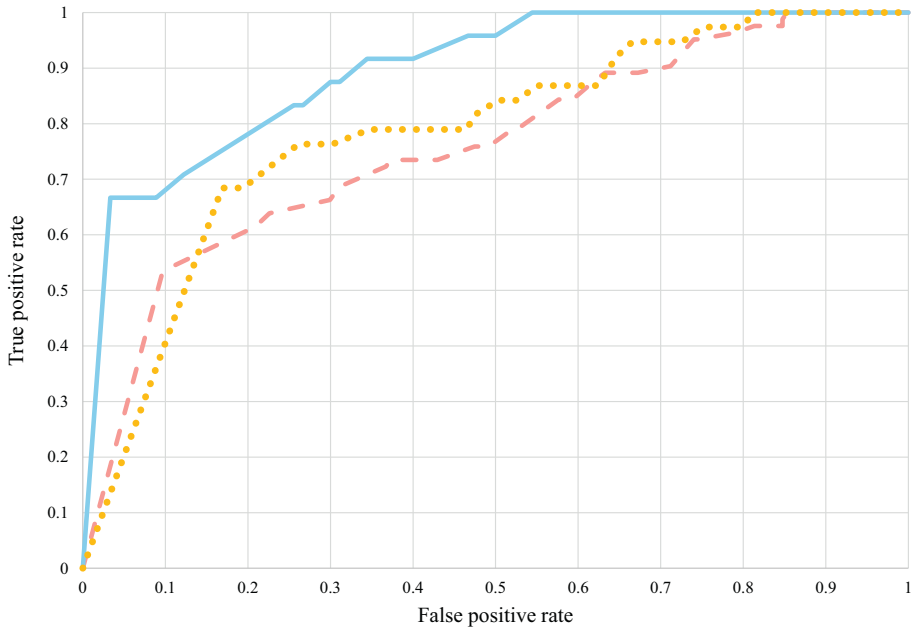

Fig. 3. ROC curves of the sequence ambiguity method using three different estimates of recent HIV-1 infection. The LAG assay is depicted with a solid line, the BED assay with a dashed line, and the estimation based on anti-HIV serological patterns with a dotted line.

Table 4
ROC curve analyses of the sequence ambiguity method when BED, LAG, or SER were used as the gold standard to discriminate RI from LSI.

Measure	BED	LAG	SER
Area under the ROC curve (AUC)	0.758	0.896	0.785
Standard error[a]	0.0326	0.0344	0.0385
95% confidence interval[b]	0.701–0.809	0.824–0.945	0.738–0.827
z statistic	7.92	11.5	7.393
Significance level *p* (area = 0.5)	< 0.0001	< 0.0001	< 0.0001
Youden index *J*	0.4348	0.6333	0.5154
Optimal SA cutoff (%)	0.30	0.20	0.11
Sensitivity (%)	68.7	79.2	76.3
Specificity (%)	68.9	78.9	73.9

ROC, receiver operating characteristic; BED, the BED test; LAG, the LAg-Avidity assay; SER, anti-HIV serological patterns; RI, recent HIV infection; LSI, long-standing HIV infection; SA, sequence ambiguity method.
[a] DeLong et al., 1988.
[b] Binomial exact.

4. Discussion

The results of this study support previous findings that SA has sufficient discriminatory power for differentiation of RI and LSI at a population level. Although all the standard methods used to characterize RI in this study and similar previous studies (BED, LAG, SER) face obstacles regarding both sensitivity and specificity (Hauser et al., 2014; Longosz et al., 2014), SA was not perceivably inferior to them. The observed sensitivity and specificity of SA for RI of about 80% may be real but may also result from misclassification of RI by the reference incidence method(s). Several studies reported that serology-based incidence methods tend to misclassify samples as RI when individuals presented with a low CD4+ T-cell count or were receiving antiretroviral treatment (low viral load) (Chaillon et al., 2012; Murphy and Parry, 2008). Thus, use of multi-assay algorithms was proposed, encompassing additional criteria, or at least the HIV viral load, to exclude false RI (Kassanjee et al., 2016; Laeyendecker et al., 2013; Lunar et al., 2015). Low viral load was not included as a criterion of LSI in this study for two reasons: to study the characteristics of incidence assays solely in relation to SA and because *pol* sequencing was not performed if the HIV viral load was below 1000 copies/ml.

Based on our results, we propose the use of a marginally lower SA cutoff of 0.11% to 0.30% compared to the previously suggested cutoff of 0.45% to 0.50% (Andersson et al., 2013; Kouyos et al., 2011; Ragonnet-Cronin et al., 2012). The rationale is that in our study persons were characterized as RI by assays using a relatively short time window for MDRI; namely, 155 and 130 days for the BED and LAG tests, respectively. MDRI is even shorter for SER because it is determined by the short serological window of HIV infection. In contrast, the two large previous studies characterized RI and generated sequences within 1 year of observed seroconversion, resulting in a higher cutoff (Andersson et al., 2013; Kouyos et al., 2011). The study by Ragonnet-Cronin et al. (2012) proposed a MDRI of 6 months and a cutoff of 0.45% based on their BED data (Ragonnet-Cronin et al., 2012). In our study, most of the BED data correlated with a cutoff of 0.30% SA for a MDRI of approximately 5 months. Our data fit well with the observation of the 0.2% increase in SA that HIV gains per year and also with the 0.47% to 0.50% SA cutoff employed to characterize individuals infected within 1 year before diagnosis (Andersson et al., 2013; Kouyos et al., 2011). An explanation for the observed cutoff discrepancy is that persons infected with more than one founder virus contributed to the SA observed within the group of RI and/or that differences in SA could be simply attributed to inter-laboratory variability, as shown by Kouyos et al. (2011). However, one needs to bear in mind that, if only the routinely sequenced *pol* region is used for analysis and thresholds lower than 0.3% are applied, the ambiguity of only three nucleotide positions is used in calculations, making misclassification more probable. Nevertheless, our study showed that the time trends of RI could be reliably determined regardless of the cutoff used. Although no difference in the time trends was observed in Slovenia, depending on the cohort tested and the definition of RI, it is probable that in different settings the optimal cutoff would change for time.

The BED data were the most representative for the entire HIV-infected population in Slovenia due to wide inclusion of patients, and SA corresponded well with it. Some differences observed, especially for 2011, were probably due to sampling error because not all newly diagnosed individuals were routinely screened for the presence of TDR at baseline, but usually only half of them are included.

The SA method has several limitations. First, if assembly of

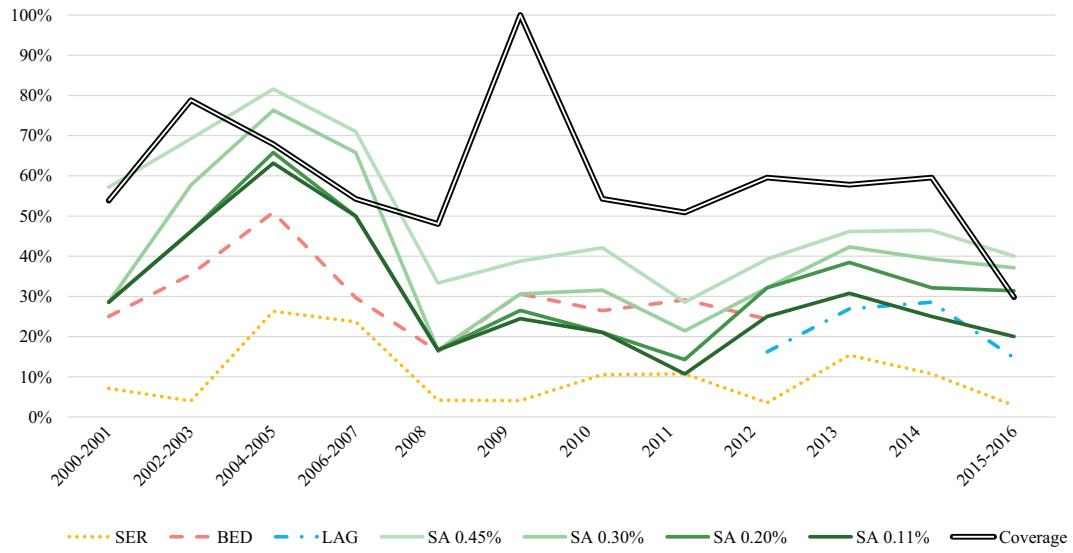

Fig. 4. Annual dynamic of shares of recent HIV-1 infections among newly diagnosed individuals in Slovenia over a 16-year period estimated using four different methods: BED, the BED assay; LAG, the LAg-Avidity assay; SER, evaluation based on anti-HIV serological patterns; SA, sequence ambiguity of the *pol* region with different cutoffs employed for recent infection (< 0.11%, < 0.2%, < 0.3%, and < 0.45%). Added is also the annual coverage of SA results among newly diagnosed individuals.

sequences and editing is performed manually, the procedure is subject to interpretation bias due to the operator's individual judgment whether mixed codons observed at a certain nucleotide position in chromatograms represent a true mixed variant or only baseline noise. This bias could potentially lead to lower intra-laboratory reproducibility and inter-laboratory agreement. Ragonnet-Cronin et al. (2012) studied this in detail by analyzing the performance of SA at different detection thresholds of mixed base calls. They concluded that thresholds between 15% and 25% perform equally well, given that this is actually the perceived general sensitivity of ambiguous nucleotide detection, whether analysis was automatic or manual (Ragonnet-Cronin et al., 2012). The estimate of the virus diversity can be improved considerably by increasing sequencing depth using next-generation sequencing (NGS), as recently shown by Puller et al. (2017). They employed a continuous approach as opposed to a binominal one to provide estimates of the time since infection and showed best results when based on the average pairwise distance on 3rd codon positions in the *pol* region and selection of the lowest possible cutoff for minority variants (in their case 0.3%), which proved superior over Sanger sequencing. In addition, this approach provided an adequate estimate of the time since infection for a prolonged period; for example, an absolute error of 1.1 years was observed at 6 years after infection. However, NGS is still not the method of choice for routine HIV genotyping and proves to be financially challenging, especially in small laboratories and in resource-limited settings (Puller et al., 2017). Second, even though it is generally believed that HIV is transmitted via a single genetic variant (genetic bottleneck), in certain occasions more than one variant can be transmitted or superinfection can occur shortly after initial infection, leading to increased SA (Keele et al., 2008). For example, in the Swiss cohort 18% of individuals that were infected within 1 month prior to sampling had a SA of > 0.68% (Kouyos et al., 2011). A higher SA proportion was also found among persons that acquired HIV through injecting drug use, a group that represented 26% of the Swiss cohort (Kouyos et al., 2011). Because injecting drug use is seldom found to be related to HIV transmission in Slovenia (only two out of the 353 persons included in this study, data not shown), this should not be a concern in our setting. However, our study found 21% of persons characterized as RI using SER with a SA of > 0.5%, suggesting other transmission groups such as MSM to be equally able to acquire HIV infection with several founder viruses. Although Keele et al. (2008) found that half of MSM were initially infected with more than one founder virus, their findings were not confirmed in a recent study by Tully et al. (2016), showing an equivalent number of founder viruses among MSM and heterosexually infected persons (Keele et al., 2008; Tully et al., 2016). Interestingly, the MSM transmission route was significantly correlated with lower SA in this study. This is probably due to the fact that MSM in Slovenia tend to test for HIV more frequently than other risk groups and heterosexually infected persons presented with the more advanced disease at diagnosis (lower CD4+ T-cell count) (data not shown). A CD4+ T-cell count of fewer than 250 cells/mm^3 was expectedly associated with higher SA, but, on the other hand, the baseline viral load had no effect on SA. This finding is reassuring because it would be possible that, in cases of low viral load, the detection of HIV quasispecies is limited and we would expect that low viral load would be associated with low SA. Although this was not the case, it is still possible that the detection limitations of SA in persons with low viral loads balanced with the high SA observed in LSI, where the viral load tends to be lower compared to recently infected patients. Third, the rate of HIV diversification is affected by the state of the patient's immune system and can vary considerably among individuals (Delwart et al., 1997; Allen et al., 2005).

Despite all limitations, an accurate estimation of the HIV epidemic and its dynamics can be gleaned from SA if SA is used at a population level only and not at the individual patient level, as is true for other serology-based incidence assays.

In conclusion, our data suggest that SA could be used as a real-time surveillance tool for close monitoring of HIV incidence trends, especially in countries where baseline HIV resistance genotyping is performed routinely, rendering this approach cost-effective.

Conflict of interest

The authors report no conflict of interest connected with this study.

Funding

This research received no specific funding from any public, commercial, or not-for-profit agency/institution.

Acknowledgements

The authors would like to thank all patients included in the study, the HIV clinicians at the Department of Infectious Diseases, Ljubljana University Medical Center, for sample and data collection, and Miha Lunar for writing the Python script.

References

Allen, T.M., Altfeld, M., Geer, S.C., Kalife, E.T., Moore, C., O'Sullivan, K.M., Desouza, I., Feeney, M.E., Eldridge, R.L., Maier, E.L., Kaufmann, D.E., Lahaie, M.P., Reyor, L., Tanzi, G., Johnston, M.N., Brander, C., Draenert, R., Rockstroh, J.K., Jessen, H., Rosenberg, E.S., Mallal, S.A., Walker, B.D., 2005. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. J. Virol. 79, 13239–13249.

Andersson, E., Shao, W., Bontell, I., Cham, F., Cuong do, D., Wondwossen, A., Morris, L., Hunt, G., Sönnerborg, A., Bertagnolio, S., Maldarelli, F., Jordan, M.R., 2013. Evaluation of sequence ambiguities of the HIV-1 pol gene as a method to identify recent HIV-1 infection in transmitted drug resistance surveys. Infect. Genet. Evol. 18, 125–131. https://doi.org/10.1016/j.meegid.2013.03.050.

Babič, D.Z., Poljak, M., Seme, K., Tomažič, J., Vidmar, L., 2006a. Molecular epidemiology of HIV-1 subtypes based on analysis of pol sequences in Slovenia, 1996–2005. J. Med. Virol. 78, 997–1002.

Babič, D.Z., Zelnikar, M., Seme, K., Vandamme, A.M., Snoeck, J., Tomažič, J., Vidmar, L., Karner, P., Poljak, M., 2006b. Prevalence of antiretroviral drug resistance mutations and HIV-1 non-B subtypes in newly diagnosed drug-naïve patients in Slovenia, 2000–2004. Virus Res. 118, 156–163.

Calypte Biomedical Corporation, 2008. Aware™ BED™ EIA HIV-1 Incidence Test (IgG-Capture HIV-EIA) Package Insert. http://www.calypte.com/pdf/LN-20597%20A_%20Aware%20BED%20EIA%20HIV-1_2%20Incidence%20Test%20Package%20Insert.pdf (accessed 9 May 2017).

Carlson, J.M., Schaefer, M., Monaco, D.C., Batorsky, R., Claiborne, D.T., Prince, J., Deymier, M.J., Ende, Z.S., Klatt, N.R., DeZiel, C.E., Lin, T.H., Peng, J., Seese, A.M., Shapiro, R., Frater, J., Ndung'u, T., Tang, J., Goepfert, P., Gilmour, J., Price, M.A., Kilembe, W., Heckerman, D., Goulder, P.J., Allen, T.M., Allen, S., Hunter, E., 2014. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science 345, 1254031. https://doi.org/10.1126/science.1254031.

Chaillon, A., Le Vu, S., Brunet, S., Gras, G., Bastides, F., Bernard, L., Meyer, L., Barin, F., 2012. Decreased specificity of an assay for recent infection in HIV-1-infected patients on highly active antiretroviral treatment: implications for incidence estimates. Clin. Vaccine Immunol. 19, 1248–1253. https://doi.org/10.1128/CVI.00120-12.

Corporation, Sedia Biosciences, 2016. Sedia HIV-1 LAg-Avidity EIA: Single Well Avidity Enzyme Immunoassay for Detection of Recent HIV-1 Infection Using Liquid Serum or Plasma, Cat. No. 1002. http://www.sediabio.com/LiteratureRetrieve.aspx?ID=134692, Accessed date: 9 May 2017.

Cousins, M.M., Laeyendecker, O., Beauchamp, G., Brookmeyer, R., Towler, W.I., Hudelson, S.E., Khaki, L., Koblin, B., Chesney, M., Moore, R.D., Kelen, G.D., Coates, T., Celum, C., Buchbinder, S.P., Seage 3rd, G.R., Quinn, T.C., Donnell, D., Eshleman, S.H., 2011. Use of a high resolution melting (HRM) assay to compare gag, pol, and env diversity in adults with different stages of HIV infection. PLoS One 6, e27211. https://doi.org/10.1371/journal.pone.0027211.

Delwart, E.L., Pan, H., Sheppard, H.W., Wolpert, D., Neumann, A.U., Korber, B., Mullins, J.I., 1997. Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. J. Virol. 71, 7498–7508.

Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51, 151–178.

Duong, Y.T., Qiu, M., De, A.K., Jackson, K., Dobbs, T., Kim, A.A., Nkengasong, J.N., Parekh, B.S., 2012. Detection of recent HIV-1 infection using a new limiting-antigen avidity assay: potential for HIV-1 incidence estimates and avidity maturation studies. PLoS One 7, e33328. https://doi.org/10.1371/journal.pone.0033328.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. 41, 95–98.

Hauser, A., Santos-Hoevener, C., Meixenberger, K., Zimmermann, R., Somogyi, S., Fiedler, S., Hofmann, A., Bartmeyer, B., Jansen, K., Hamouda, O., Bannert, N., Kuecherer, C., 2014. Improved testing of recent HIV-1 infections with the BioRad avidity assay compared to the limiting antigen avidity assay and BED Capture enzyme immunoassay: evaluation using reference sample panels from the German Seroconverter Cohort. PLoS One 9, e98038. https://doi.org/10.1371/journal.pone.0098038.

Kassanjee, R., Pilcher, C.D., Busch, M.P., Murphy, G., Facente, S.N., Keating, S.M., Mckinney, E., Marson, K., Price, M.A., Martin, J.N., Little, S.J., Hecht, F.M., Kallas, E.G., Welte, A., Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA), 2016. Viral load criteria and threshold optimization to improve HIV incidence assay characteristics. AIDS 30, 2361–2371.

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., Wei, X., Jiang, C., Kirchherr, J.L., Gao, F., Anderson, J.A., Ping, L.H., Swanstrom, R., Tomaras, G.D., Blattner, W.A., Goepfert, P.A., Kilby, J.M., Saag, M.S., Delwart, E.L., Busch, M.P., Cohen, M.S., Montefiori, D.C., Haynes, B.F., Gaschen, B., Athreya, G.S., Lee, H.Y., Wood, N., Seoighe, C., Perelson, A.S., Bhattacharya, T., Korber, B.T., Hahn, B.H., Shaw, G.M., 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. 105, 7552–7557. https://doi.org/10.1073/pnas.0802203105.

Kouyos, R.D., von Wyl, V., Yerly, S., Böni, J., Rieder, P., Joos, B., Taffé, P., Shah, C., Bürgisser, P., Klimkait, T., Weber, R., Hirschel, B., Cavassini, M., Rauch, A., Battegay, M., Vernazza, P.L., Bernasconi, E., Ledergerber, B., Bonhoeffer, S., Günthard, H.F., Study, Swiss H.I.V. Cohort, 2011. Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. Clin. Infect. Dis. 52, 532–539. https://doi.org/10.1093/cid/ciq164.

Laeyendecker, O., Brookmeyer, R., Cousins, M.M., Mullis, C.E., Konikoff, J., Donnell, D., Celum, C., Buchbinder, S.P., Seage 3rd, G.R., Kirk, G.D., Mehta, S.H., Astemborski, J., Jacobson, L.P., Margolick, J.B., Brown, J., Quinn, T.C., Eshleman, S.H., 2013. HIV incidence determination in the United States: a multiassay approach. J. Infect. Dis. 207, 232–239. https://doi.org/10.1093/infdis/jis659.

Longosz, A.F., Serwadda, D., Nalugoda, F., Kigozi, G., Franco, V., Gray, R.H., Quinn, T.C., Eshleman, S.H., Laeyendecker, O., 2014. Impact of HIV subtype on performance of the limiting antigen-avidity enzyme immunoassay, the bio-rad avidity assay, and the BED capture immunoassay in Rakai, Uganda. AIDS Res. Hum. Retrovir. 30, 339–344. https://doi.org/10.1089/AID.2013.0169.

Lunar, M.M., Židovec Lepej, S., Abecasis, A.B., Tomažič, J., Vidmar, L., Karner, P., Vovko, T.D., Pečavar, B., Maver, P.J., Seme, K., Poljak, M., 2013. Short communication: prevalence of HIV type 1 transmitted drug resistance in Slovenia: 2005–2010. AIDS Res. Hum. Retrovir. 29, 343–349. https://doi.org/10.1089/AID.2012.0152.

Lunar, M.M., Matković, I., Tomažič, J., Vovko, T.D., Pečavar, B., Poljak, M., 2015. Longitudinal trends of recent HIV-1 infections in Slovenia (1986–2012) determined using an incidence algorithm. J. Med. Virol. 87, 1510–1516. https://doi.org/10.1002/jmv.24209.

Lunar, M.M., Židovec Lepej, S., Tomažič, J., Vovko, T.D., Pečavar, B., Turel, G., Maver, M., Poljak, M., 2018. HIV-1 transmitted drug resistance in Slovenia and its impact on predicted treatment effectiveness: 2011–2016 update. PLoS One 13, e0196670. https://doi.org/10.1371/journal.pone.0196670.

Mansky, L.M., Temin, H.M., 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J. Virol. 69, 5087–5094.

Murphy, G., Parry, J.V., 2008. Assays for the detection of recent infections with human immunodeficiency virus type 1. Euro Surveill. 13, 314–320.

Parekh, B.S., Kennedy, M.S., Dobbs, T., Pau, C.P., Byers, R., Green, T., Hu, D.J., Vanichseni, S., Young, N.L., Choopanya, K., Mastro, T.D., McDougal, J.S., 2002. Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. AIDS Res. Hum. Retrovir. 18, 295–307.

Pineda-Peña, A.C., Faria, N.R., Imbrechts, S., Libin, P., Abecasis, A.B., Deforche, K., Gómez-López, A., Camacho, R.J., de Oliveira, T., Vandamme, A.M., 2013. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. Infect. Genet. Evol. 19, 337–348. https://doi.org/10.1016/j.meegid.2013.04.032.

Puller, V., Neher, R., Albert, J., 2017. Estimating time of HIV-1 infection from next-generation sequence diversity. PLoS Comput. Biol. 13, e1005775. https://doi.org/10.1371/journal.pcbi.1005775.

Ragonnet-Cronin, M., Aris-Brosou, S., Joanisse, I., Merks, H., Vallée, D., Caminiti, K., Rekart, M., Krajden, M., Cook, D., Kim, J., Malloch, L., Sandstrom, P., Brooks, J., 2012. Genetic diversity as a marker for timing infection in HIV-infected patients: evaluation of a 6-month window and comparison with BED. J. Infect. Dis. 206, 756–764. https://doi.org/10.1093/infdis/jis411.

Tully, D.C., Ogilvie, C.B., Batorsky, R.E., Bean, D.J., Power, K.A., Ghebremichael, M., Bedard, H.E., Gladden, A.D., Seese, A.M., Amero, M.A., Lane, K., McGrath, G., Bazner, S.B., Tinsley, J., Lennon, N.J., Henn, M.R., Brumme, Z.L., Norris, P.J., Rosenberg, E.S., Mayer, K.H., Jessen, H., Kosakovsky Pond, S.L., Walker, B.D., Altfeld, M., Carlson, J.M., Allen, T.M., 2016. Differences in the selection bottleneck between modes of sexual transmission influence the genetic composition of the HIV-1 founder virus. PLoS Pathog. 12, e1005619. https://doi.org/10.1371/journal.ppat.1005619.