Virology Journal

RESEARCH Open Access

Prediction of recent HIV-1 infections using Shannon entropy analysis of HIV-1 group-specific antigen protein sequence

Tumelo L. Fortuin[1], Paballo Nkone[1], Shayne Loubser[2], Caroline T. Tiemessen[2] and Simnikiwe H. Mayaphi[1,3*]

Abstract

Background Avidity assays often misclassify chronic HIV-1 infection as recent HIV-1 infection (false recency rate), especially in participants on antiretroviral therapy. The aim of this study was to use Shannon entropy to evaluate HIV-1 group-specific antigen (Gag) sequence diversity for the prediction of recent HIV-1 infections.

Methods This was a retrospective study that characterised the complete HIV-1 Gag using Sanger sequences obtained from participants with confirmed recent or chronic HIV-1 infection. Shannon entropy was calculated for the entire HIV-1 Gag amino acid (aa) sequence (501aa) and sliding window analysis was computed at intervals of 100aa each. This was followed by searching for aa sites that exhibited a different distribution of mutations between recent and chronic HIV-1 infection stages. Reference sequences were obtained from GenBank and the Los Alamos HIV database to verify the findings obtained from study sequences.

Results Forty-seven participants with a mean age of 28.7 years (18 – 44) were enrolled, and fourteen (30%) of them had recent HIV-1 infection. Shannon entropy analysis showed a significantly higher aa diversity in chronic HIV-1 infection compared to recent HIV-1 infection ($p = 0.0003$). Analysis of sliding windows led to identification of four aa positions; S54, E55, I256, and S451; with different pattern of distribution between recent and chronic HIV-1 infection stages; however statistical significance was only observed for three of these aa, p values = 0.094, 0.027, 0.027 and 0.045, respectively. The performance of these informative sites for detection of recent HIV-1 infection in study sequences ranged from 71—86%, however, they had a high false recency rate (FRR) ranging from 39%—52%. Similar performance was observed in reference sequences. The combination of some informative aa sites reduced FRR in study sequences to below 24%.

Conclusions Our data show that a Gag-based molecular strategy can be used to detect recent HIV-1 infections where Gag sequences are available. However, the results would have to be interpreted with caution due to an association with a high FRR. Further studies are needed to develop a molecular-based strategy with better performance for detection of recent HIV-1 infections.

Keywords HIV, 1 Gag diversity, Shannon entropy, Recent HIV, 1 infection, False recency rate, Chronic HIV, 1 infection

*Correspondence:
Simnikiwe H. Mayaphi
sim.mayaphi@up.ac.za
[1]Department of Medical Virology, University of Pretoria, Private Bag X323, Gezina 0002, South Africa
[2]National Institute for Communicable Diseases and Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[3]National Health Laboratory Service-Tshwane Academic Division (NHLS-TAD), Tshwane, South Africa

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

The scale-up of antiretroviral therapy (ART) along with various global efforts to eliminate HIV/AIDS have led to a gradual decline in the number of new HIV-1 infections. However, there still is a high number of new HIV-1 infections reported globally with 1.3 million new HIV-1 infections reported in 2023 [1–4].

HIV incidence refers to the rate of new HIV infections which occur in a specific population over a defined period. It is an important metric which is used to evaluate the dynamics of the HIV-1 epidemic, as well as assessing the effectiveness of HIV-1 prevention strategies [5, 6]. HIV incidence reports aid in identifying populations at higher risk of new infections and thus guiding public health interventions [6–8].

Longitudinal cohort studies are the current gold standard for estimating HIV incidence, however these are costly, require extended sampling periods, and have other challenges such as retention of participants and sampling bias [9, 10]. Cross-sectional studies are an alternative to longitudinal cohort studies, and these studies only sample once-off, and use biomarkers to distinguish between participants who are likely to have either recent or chronic HIV-1 infection [10, 11]. Recent HIV-1 is defined as the period of infection prior to the viral load (VL) set point, which is usually achieved at 6 months after viral acquisition. In untreated individuals, chronic HIV-1 infection follows soon after the VL set point and ends before the AIDS stage [12–14].

Cross-sectional studies often use HIV-1 avidity assays such as limiting antigen avidity assay (LAg-Avidity assay) to characterise participants as having recent infection [15, 16]. LAg-Avidity assays measure the binding capacity of immunoglobulin G (IgG) antibodies against HIV-1 antigen, wherein a low IgG avidity would be classified as a recent infection and a high IgG avidity would be classified as a chronic infection [11, 17–19]. These assays can be used on their own or as part of multi-assay algorithms (MAAs) for measuring HIV incidence. MAAs may include HIV-1 VL and/or CD4 T-cell count in addition to the avidity assay [11, 20]. Avidity assays often misclassify chronic HIV-1 infection as recent HIV-1 infection (false recency rate [FRR]), especially in participants on ART, as these individuals would have a suppressed HIV-1 VL and less stimulation of antibody responses [17, 20–22].

Previous studies have developed molecular-based strategies which use Sanger and next generation sequencing (NGS) methods to assess viral sequence diversity at the nucleotide and amino acid (aa) level as an alternative to predict recency. These strategies have shown good performances; however, they often have a high FRR [7, 23, 24]. Our group previously designed a *pol*-based molecular strategy which showed good performance for prediction of recent HIV-1 infection, however, this also had a high FRR [14].

The Gag protein encodes immunodominant CTL epitopes which are targeted by CTL responses in recent HIV-1 infection to reduce viraemia to a low set-point. Gag CTL responses also play an important role in the control of viraemia during chronic HIV-1 infection stage. Thus, diversity within Gag is mainly driven by immune selection pressure from CTL responses, and this could lead to mutations which could be used to discriminate between recent and chronic HIV-1 infection stages [25–27]. The aim of this study was to use Shannon entropy to evaluate HIV-1 group-specific antigen (Gag) sequence diversity for the prediction of recent HIV-1 infections.

Materials and methods

Study population

This was a retrospective study that characterised complete HIV-1 Gag protein sequences from stored plasma sample pairs obtained at two different time points from a previous study, with a median sampling interval of 8 weeks (interquartile range: 2–8 weeks) between baseline and follow-up samples. HIV-1 nucleic acid amplification test (NAAT) was used to screen for recent HIV-1 infections in participants who had negative rapid tests at five HIV testing and counselling clinics between 2012 – 2016, in the Tshwane district of South Africa. HIV-1 infections identified through NAAT were further characterised and confirmed as recent or chronic using the following tests: p24 antigen, Western blot, third generation enzyme-linked immunoassay, and LAg avidity assay. CD4 count data was also available for most participants [13]. Thus, in this study, we amplified and sequenced the HIV-1 *gag* gene from samples known to have recent or chronic HIV-1 infection and looked for amino acid mutations that can potentially be used to differentiate chronic from recent HIV-1 infection.

Total nucleic acid extraction and amplification of HIV-1 gag

The QIAamp UltraSens Virus Kit (Qiagen, Hilden, Germany) was used to extract total nucleic acids from plasma samples. A published nested PCR was used to amplify the complete HIV-1 *gag* gene in all samples using the SuperScript™ III One-Step RT-PCR System in the first round and the Platinum™ *Taq* High Fidelity DNA Polymerase (Invitrogen, Carlsbad, CA, USA) in the second round [28]. Agarose gel electrophoresis was used to confirm successful amplification of a 1.6 kb target fragment.

HIV-1 gag Sanger sequencing and sequence data analysis

Sanger sequencing was performed in five overlapping regions which cover the entire HIV-1 *gag* open reading frame (ORF) (Inqaba Biotechnical Industries, Tshwane, South Africa). CLC Main Workbench software was used

to generate consensus sequences which were analysed on BioEdit 7.7.1 and MEGA 11 software. Consensus sequences were aligned on MAFFT, and MEGA 11 software was used to compute phylogenetic trees. The HXB2 (K03455.1) HIV-1 reference strain was used for numbering. HIV-1 subtype C study sequences were included in the final analysis and, non-subtype C sequences were excluded.

Phylogenetic analysis showed that the majority of study sequences (47/50, 94%) clustered with subtype C references, and a minority (3/50, 6%) clustered with non-subtype C (Supplementary Fig. 1). Final analysis excluded non-subtype C sequences.

Evaluation of sequence diversity with Shannon entropy analysis

Shannon entropy measures the variability within nucleotide or aa sequences and assigns a low score to sites with low variability and a high score to sites with high variability, thus assessing the conservation or variability across different positions of the sequence. Consensus sequences of the complete HIV-1 *gag* gene were grouped into recent and chronic HIV-1 and aligned to an HIV-1 subtype B reference sequence (HXB2, K03455.1), and were then translated to aa sequences which were submitted to the Los Alamos HIV database (https://www.hiv.lanl.gov/content/sequence/ENTROPY/entropy_one.html) to calcula te Shannon entropy using the entropy one feature. Data containing the entropy values for the different aa positions was then exported to an excel spreadsheet, which was later used as an input file for further analysis on GraphPad Prism (see statistical analysis section below).

Initially, Shannon entropy was calculated for the entire HIV-1 Gag aa sequence (501aa) for both recent HIV-1 and chronic HIV-1, this was followed by a comparison of median entropy scores of these two HIV-1 stages. Thereafter, sliding window analysis was computed for HIV-1 Gag at intervals of 100aa each as this interval allowed for coverage of the entire protein except the last aa, then the median entropy scores for the different sliding windows were compared between the two HIV-1 infection stages. Following sliding window analysis, we screened for specific aa sites or informative sites that could be used to discriminate between recent and chronic HIV-1 disease stages, starting with the sliding window(s) that had a statistically significant median entropy score, and then expanding screening to other windows. These informative sites are aa that showed a different pattern of distribution between recent and chronic HIV-1 disease stages such as being more conserved in one stage. They were evaluated individually for differentiating between recent and chronic HIV-1 disease stages and were also evaluated in aa groups of two. The FRR of these sites or groups was assessed in chronic sequences, and this was expressed as proportion of these sequences that resembled the pattern of the informative sites seen in recent sequences. Analysis was performed at the aa level as this allowed for optimal coverage of the Gag polyprotein while working with shorter sequence length (501 aa vs 1503 nucleotides). As such, data was easily visualised and interpreted. The aa level also allowed for an easier selection of variable informative sites which could be used to discriminate between the two HIV-1 infection stages.

Comparison of study findings with reference sequence data

Reference sequences were obtained from GenBank and the LANL (Los Alamos National Laboratory) HIV sequence database (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html). HIV-1 subtype C reference sequences were obtained from previous studies in which the participants were well characterised to have either recent (HIV-1 infection < 6 months or Fiebig stage ≤ 6, [29]) or chronic infection (HIV-1 infection > 6 months). Following screening of articles to gather reference sequence data, a total of eight studies were assessed for recent infection [26, 30–36], and four studies were assessed for chronic infection [37–39]. From these studies, five were excluded for recent infection, and all four were excluded for chronic infection as these did not meet the criteria for inclusion into final analysis. Three studies with a total of 176 reference sequences were included for recent infection [31, 34, 35] (Supplementary Fig. 2). A separate analysis was done for assessing the performance of identified informative sites for prediction of recent infection in non-subtype C reference sequences.

Statistical analysis

Descriptive statistics was used to present mean and range values for age, HIV-1 VL, CD4 T-cell count; and for proportion of participants with recent or chronic HIV-1 infection. Normal distribution of variables was evaluated before deciding on reporting mean values. Prior to analysis, absolute HIV-1 VL values were log-transformed and comparison of mean log-transformed values between recent and chronic HIV-1 stages was done. Recent and chronic HIV-1 VL $\log_{10}$ and absolute CD4 T-cell count values were compared using a two-sample t-test. CD4 T-cell count values were unavailable for four participants with recent HIV-1 infection and for eleven participants with chronic HIV infection. The distribution of informative aa sites between recent and chronic HIV-1 was assessed and one-sided Fischer's exact test was used to determine if there was an association between this distribution and HIV-1 infection stages. The above-mentioned statistical tests were performed on STATA 16.0 software package (StataCorp LP, College Station, TX, USA). GraphPad Prism 10.3.0 was used to compute graphical

comparisons of the median Shannon entropy scores between recent and chronic HIV-1 infection using the unpaired t-test. Statistical significance was considered when the *p*-value was ≤ 0.05.

Results

Demographics

A total of forty-seven participants, with a mean age of 28.7 years (range: 18 – 44 years) were enrolled in this study. The majority (94%) of participants were female. Fourteen (30%) participants had recent HIV-1 infection while thirty-three (70%) participants had chronic HIV-1 infection. The mean HIV-1 VL $\log_{10}$ was significantly higher in participants with recent infection (5.2 $\log_{10}$ [range: 3.2 – 8.5 $\log_{10}$] compared to those with chronic infection (4.2 $\log_{10}$ [range: 2.6 – 5.3 $\log_{10}$], $p = 0.005$. Similarly, the mean $CD4^+$ T-cell count was significantly higher in participants with recent infection (555 cells/µl [range: 215 – 964 cells/µl] compared to those with chronic infection (343 cells/µl [range: 72 – 675 cells/µl], $p = 0.006$ (Table 1).

Shannon entropy for the entire HIV-1 gag aa sequence

All the enrolled participants had HIV-1 subtype C strains (Supplementary Fig. 1). Shannon entropy analysis revealed that the aa diversity observed in chronic HIV-1 infection was significantly higher than in recent HIV-1 infection ($p = 0.0003$) (Fig. 1).

HIV-1 gag sliding window analysis and amino acid diversity

Sliding window analysis showed that one window (201—300) had a significantly higher median entropy score in chronic HIV-1 infection stage compared to recent HIV-1 infection stage, ($p = 0.017$) and another window (401 – 500) showed a similar trend but had marginal significance, $p = 0.059$ (Fig. 2).

An analysis was performed in all sliding windows to assess individual aa sites that can discriminate between recent and chronic HIV-1 infection stages. Four aa positions; S54, E55, I256, and S451, were identified to have a different pattern of distribution between recent and chronic HIV-1 infection stages, however statistical significance was only observed for three of these aa; p values of 0.094, 0.027, 0.027 and 0.045, respectively. Diversity of some of these aa is summarised on a WebLogo file on **Supplementary Fig. 3.**

Performance of an HIV-1 gag-based molecular technique for prediction of recent infection

The performance to predict recent infection varied across the individual informative aa sites. S54 aa had a performance of 71% in study sequences for prediction of recent HIV infection, and this ranged from 74 to 90% in reference sequences; while E55 and I256 each had a performance of 86% in study sequences, and this ranged from 76 to 85%, and 60% to 100% in reference sequences, respectively. S451 had a performance of 71% in study sequences and its performance ranged from 30 to 79% in reference sequences. Combinations of the individual informative aa sites were assessed for the ability to predict recent infection, and the performance for S54 and E55 was 71% in study sequences and ranged from 60 to 80% in reference sequences. Combinations such as S54 and S451, E55 and S451, along with I256 and S451 were also evaluated, however gave a lower performance which ranged from 50 to 57% in study sequences and ranged from 20 to 79% in reference sequences (Table 2). Performance of the identified informative sites was also assessed in non-subtype C reference sequences, mainly subtypes A and B. However, this yielded different results compared to what was observed in subtype C sequences. For instance, S54 aa position in subtype A recent sequences had a lot of mutations (Supplementary Fig. 4).

The individual aa sites showed a high FRR in study sequences. S54, E55, I256 and S451 showed FRRs of 45%, 52%, 52%, and 39%, in study sequences, respectively. The combination of some informative aa sites reduced FRR in study sequences to below 24% (Table 2). Further analysis showed that the identified informative aa positions of interest were located in recognised cytotoxic T lymphocyte (CTL) epitopes within the matrix (p17), the capsid (p24), and Gag polyprotein (p2p7p1p6) (Table 3**).**

Discussion

This study evaluated the performance of an HIV-1 Gag-based molecular technique to predict recent HIV-1 infection in ART-naïve participants confirmed to have either recent or chronic HIV-1 infection [13]. The significance of ART-naïve participants is that performance can be assessed in a population wherein the genetic diversity is not driven by drug pressure [40–43]. It is not surprising that the mean VL and $CD4^+$ T-cell count were significantly higher in recent HIV-1 infection as this stage is known to be associated with higher VL and $CD4^+$ T cell count compared to chronic infection [44, 45].

The median Shannon entropy scores were significantly higher in chronic infection compared to recent infection, which highlight that there was more genetic diversity associated with the chronic infection stage. Varying factors such as virus-host interactions, viral replication kinetics and viral adaptions would have influenced viral evolution dynamics allowing for more genetic variation to occur hence over time, the chronic HIV-1 infection stage exhibited increased diversity [46, 47]. All the informative sites identified in this study were found to be located within known CTL epitopes (Table 3). This is not surprising as Gag is known to harbour immunodominant epitopes which induce virus-specific CTL responses

Table 1 Summary of the participants demographics and distribution of amino acid mutations between recent and chronic HIV-1 infection for the study sequences

Pt ID	HIV stage	Sex	Age (years)	HIV VL (copies/ml)	CD4 count	S54	E55	S54 + E55	I256	S451	S54 + S451	E55 + S451	I256 + S451
9498	R	F	34	500,000	964	-	-	-	-	X	X	X	X
8575	R	F	27	93,000	668	X	X	X	-	-	X	X	-
5041	R	M	23	22,000,000	n/a	-	-	-	-	X	X	X	X
8047	R	M	31	1,200,000	n/a	X	-	X	-	-	X	-	-
9049	R	F	20	16,000	n/a	-	-	-	X	-	-	-	X
6638	R	F	28	190,000	457	-	-	-	-	-	-	-	-
261	R	M	33	84,000,000	386	X	X	X	-	-	X	X	-
6512	R	F	23	1700	215	X	-	X	-	X	X	X	X
6743	R	F	26	27,000	638	-	-	-	-	-	-	-	-
6582	R	F	24	6200	818	-	-	-	-	-	-	-	-
6727	R	F	28	4800	706	-	-	-	X	-	-	-	X
6737	R	F	24	2200	n/a	-	-	-	-	X	X	X	X
7084	R	F	28	330,000,000	411	-	-	-	-	-	-	-	-
2504	R	F	24	37,000	287	-	-	-	-	-	-	-	-
5054	C	F	26	27,000	72	X	X	X	X	X	X	X	X
9915	C	F	30	14,000	382	-	-	-	-	X	X	X	X
4351	C	F	35	2600	n/a	-	X	X	X	X	X	X	X
7959	C	F	40	140,000	343	X	X	X	X	X	X	X	X
9895	C	F	31	4800	607	-	X	X	-	X	X	X	X
9986	C	F	28	97,000	160	-	-	-	-	-	-	-	-
843	C	F	21	29,000	348	-	-	-	X	X	X	X	X
6990	C	F	26	17,000	269	X	-	X	-	-	X	-	-
2340	C	F	22	14,000	n/a	X	-	X	X	X	X	X	X
6671	C	F	35	14,000	n/a	-	X	X	-	X	X	X	X
6380	C	F	25	11,000	385	-	X	X	X	-	-	X	X
6565	C	F	28	5600	371	-	X	X	X	-	-	X	X
6640	C	F	37	3000	407	X	-	X	X	X	X	X	X
6649	C	F	32	21,000	164	-	-	-	-	X	X	X	X
3869	C	F	32	210,000	n/a	-	-	-	-	X	X	X	X
3912	C	F	27	32,000	287	X	X	X	X	X	X	X	X
3920	C	F	20	66,000	306	-	-	-	-	-	-	-	-
3880	C	F	30	7500	575	X	X	X	-	X	X	X	X
1121	C	F	27	80,000	127	X	X	X	-	-	X	X	-
3474	C	F	21	16,000	675	-	-	-	-	X	X	X	X
1475	C	F	32	44,000	255	X	-	X	-	X	X	X	X
3387	C	F	37	79,000	n/a	-	X	X	-	-	-	X	-
3253	C	F	28	89,000	n/a	X	X	X	X	X	X	X	X
3606	C	F	24	15,000	529	X	X	X	X	-	X	X	X
1213	C	F	36	32,000	n/a	-	-	-	-	X	X	X	X
3910	C	F	33	28,000	n/a	X	X	X	-	X	X	X	X
8828	C	F	27	41,500	230	X	X	X	X	X	X	X	X
641	C	F	30	70,569	228	X	-	X	-	-	X	-	-
6748	C	F	29	932	n/a	X	-	X	X	-	X	-	X
921	C	F	44	9781	469	X	-	X	-	-	X	-	-
1692	C	F	18	396	n/a	X	-	X	X	-	X	-	X
2866	C	F	35	3352	n/a	X	-	X	X	-	X	-	X
6557	C	F	32	614	353	-	X	X	X	X	X	X	X

HIV-1 – Human Immunodeficiency virus type 1; Pt ID – participant identification; VL – Viral load; R – recent; C – chronic; ml – millilitre; M – Male; F – Female; n/a – not available; X – presence of a mutation; dash (-) – absence of a mutation

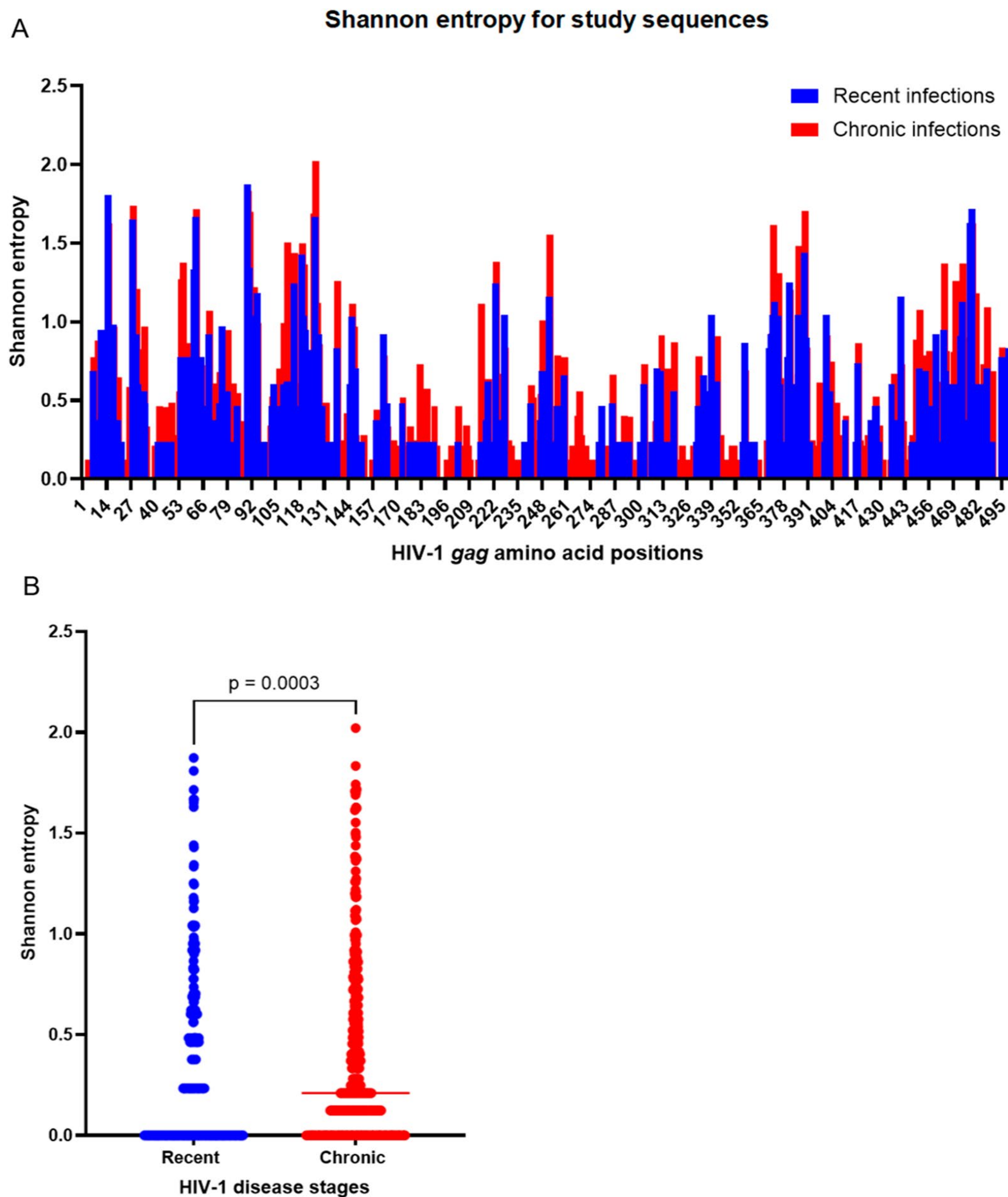

Fig. 1 Shannon entropy scores for the entire HIV-1 *gag* aa gene sequence. A. Median Shannon entropy scores were higher for chronic HIV-1 infection when compared to recent HIV-1 infection, with the highest variability observed between aa 170 and aa 300. B. Scatterplot indicating that the variability observed between chronic and recent HIV-1 was statistically significant with $p = 0.0003$. HIV-1—Human Immunodeficiency Virus type 1; gag – group-specific antigen; aa – amino acid

effective in the control of HIV-1 infection, attenuation of disease progression, and/or lower VL set point [41, 43, 48]. This emphasizes that the host CTL responses are the main driving force for HIV-1 evolution during the course of an HIV-1 infection that leads to higher genetic diversity between early and chronic infection stages [49, 50].

This study identified informative aa sites that could be used to discriminate between the two HIV-1 infection stages. The identified individual informative sites showed good performance (71—86%) for prediction of recent HIV-1 infection in study sequences, and this was similar to the reference sequences. These findings are comparable or better than what was observed during the early development stages of the BioRad Avidity assay, LAg Avidity assay, and BED-capture enzyme immunoassay (BED-CEIA) which had performance of 88%, 60%, and 48% for the detection of recent HIV-1 infection, respectively [51]. Notably, the performance of these avidity assays has improved over time with studies reporting performances ranging from 92 to 96% [19, 52, 53]. Additionally, the current study findings are also comparable to data previously published by our group that showed good

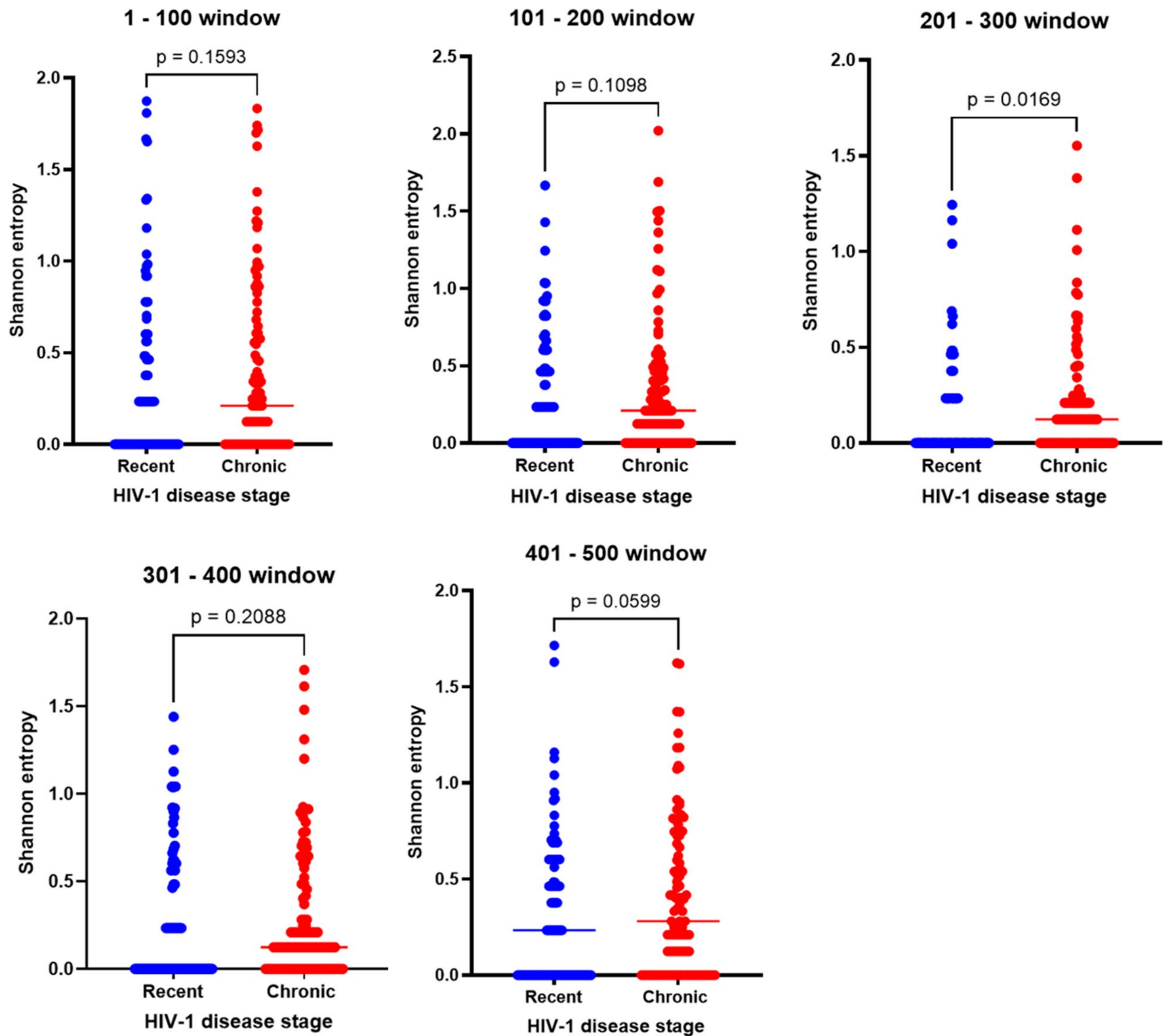

Fig. 2 Shannon entropy scores for recent compared to chronic HIV-1 infection for the entire HIV-1 *gag* aa sequence following sliding window analysis. One out five of the windows showed statistical significance (window 201 – 300, with $p = 0.0169$), while window 401 – 500 showed marginal significance. HIV-1 – Human Immunodeficiency virus type 1; gag – group-specific antigen; aa – amino acid

Table 2 Performance of study and HIV-1 subtype C reference sequences in the prediction of recent infection

Study sequences	n	S54	E55	S54 + E55	I256	S451	S54 + S451	E55 + S451	I256 + S451
Recent HIV-1 infection	14	10 (71%)	12 (86%)	10 (71%)	12 (86%)	10 (71%)	7 (50%)	8 (57%)	8 (57%)
Chronic HIV-1 infection (false recency rate)	33	15 (45%)	17 (52%)	8 (24%)	17 (52%)	13 (39%)	5 (15%)	8 (24%)	7 (21%)
Recent HIV-1 infection reference sequences									
Claiborne et al. (ZM, 2014)	127	94 (74%)	97 (76%)	76 (60%)	81 (64%)	76 (60%)	59 (46%)	60 (47%)	49 (39%)
Salazar-Gonzalez et al. (ZM, 2009)	29	23 (79%)	23 (79%)	23 (79%)	29 (100%)	23 (79%)	23 (79%)	23 (79%)	23 (79%)
Treurnicht et al. (ZA, 2010)	20	18 (90%)	17 (85%)	16 (80%)	12 (60%)	6 (30%)	5 (25%)	5 (25%)	4 (20%)

HIV-1 – Human Immunodeficiency virus type 1; n – number; ZM – Zambia; ZA – South Africa

performance (64.3 – 100%) for prediction of recent infection using a *pol*-based molecular strategy [14]. The varying distribution patterns observed with the informative aa sites on the non-subtype C reference sequences can be attributed to differences in consensus sequences between the HIV-1 subtypes [54, 55].

A high FRR (39%—52%) was observed with the individual informative sites in the study sequences, however, this was reduced to 24% with the combination of S54 and

Table 3 Location of informative amino acid mutation sites within HIV-1 Gag CTL epitopes

Amino acid site	CTL epitope mapped to	HIV-1 gag position	Subtype identified for
S54	GLLET**S**EGCKQIMKQL	p17(49–64)	C
E55	GLLETS**E**GCKQIMKQL	p17(49–64)	C
I256	WMTNNPP**I**PV	p24(117–126)	B, C
S451	GKIWPSHKGRPGNFLQ**S**R	p2p7p1p6(72–89)	C

HIV-1 – Human Immunodeficiency Virus type 1; Gag – group-specific antigen gene; CTL – cytotoxic T lymphocytes

E55. Interestingly, these study findings are also comparable with early avidity assays where FRRs ranged from 12 to 46% for the BED-CEIA, and 43% for the LAg-Avidity assay, while the BioRad Avidity assay had a FRR as high as 20% [20, 51, 56]. Although the performance and FRRs of these avidity assays have improved over time, participants that are on ART are still misclassified as recent, as they would have a lower or undetectable viremia and lower HIV-specific antibody stimulation [52, 57–59]. The FRRs observed with the Gag-based molecular strategy are comparable to the *pol*-based molecular strategy which also showed a high FRR (35.5—48.4%) [14].

The performance of a Gag-based molecular strategy observed in our study is similar to what has been reported by previous studies that have evaluated Sanger-based methods for prediction of recent HIV-1 infections [60–63]. Ragonnet-Cronin et al. and Andersson et al., previously described the use of ambiguous bases in HIV-1 *pol* gene for the prediction of recent HIV-1 infection, and reported sensitivities of 82.7% and 88.8%, and specificities of 78.8% and 74.6%, respectively [60, 61]. Other studies have also described molecular-based methods which showed good performances for the prediction of recent HIV-1 infections [62, 63]. Wu et al. was able to demonstrate use of Gag sequence diversity measured by Shannon entropy in an HIV-1 subtype C cohort as a method to improve the prediction of recent HIV-1 infection. They observed that Shannon entropy data from highly informative sites within Gag showed better performance (75 – 86%) for the prediction of recency [64].

Other previous studies have described the use of viral diversity obtained through NGS technologies as an alternative method to discriminate recent from chronic HIV-1 infection [24, 65–69]. Faraci et al. recently showed the use of micro drop sequencing and genome similarity index (GSI) for the prediction of HIV-1 incidence and observed a sensitivity of 81% for participants recently infected within 8 months and sensitivity of 92% for participants recently infected within 6 months [68]. Carlisle et al. showed the use of average pairwise diversity derived from NGS data had a sensitivity of 88% and a specificity of 85% for the prediction of recent HIV-1 infection [69]. Other NGS-based techniques have also been described and have shown better performance in estimating recent infections compared to the use Sanger-based methods [24, 65–67].

The main challenge with developing a Gag-based molecular strategy to detect recent infection is the high FRR. The possible reasons for this are that *gag* is more conserved, and that different participants express different human leukocyte antigen (HLA) profiles which affect the extent of immune selection pressure and evolution of the virus over time [70–72]. Additionally, Gag epitopes containing the informative as sites identified in this study are less likely to contribute to the control of viremia during early HIV-1 infection as they seem to be targeted later during the course of infection, which may explain why more mutations were observed during the chronic HIV-1 infection stage [26, 73, 74].

The limitations of this study include small sample size, and that its findings are specific to HIV-1 subtype C. However, to address the small sample size issue, performance was also evaluated in 176 reference sequences classified as recent HIV-1 infection [31, 34, 35], and similar performances were observed between study and reference sequences. Unfortunately, analysis was not performed in chronic reference sequences as we could not find studies that had well characterised chronic HIV-1 infection stage, as an infection carried for longer than 6 months. One of the limitations is that this study did not include participants on ART with lower HIV-1 viral load. Thus, we do not know how the Gag-based molecular strategy would perform in this group of people.

Conclusion

This study has developed a Gag-based molecular strategy that could be used to predict recent HIV-1 infection. This strategy showed good performance in study sequences and in subtype C reference sequences, however, it is associated with a high FRR. Our data show that a Gag-based molecular strategy can be used to detect recent HIV-1 infections where *gag* sequences are available. However, the results would have to be interpreted with caution due to a high FRR. Further studies are needed to develop a molecular-based strategy with better performance for detection of recent HIV-1 infections. These future studies should also evaluate the performance of the molecular-based strategy in participants on ART with lower viraemia, as these are likely to be misclassified as recent by avidity assays.

Abbreviations

Aa	Amino acid
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral Therapy
ARV	Antiretroviral
BED-CEIA	HIV Subtypes B, E, D Capture Enzyme Immunoassay
BR	Brazil
CA	California
c/mL	Copies/Milliliter
CDC	Centre for Disease Control Prevention
CD4	Cluster of Differentiation 4
cDNA	Complementary Deoxyribonucleic Acid
CTL	Cytotoxic T Lymphocytes
DNA	Deoxyribonucleic Acid
DRC	Democratic Republic of the Congo
EIA	Enzyme Immunoassay
FRR	False Recency Rate
Gag	Group-Specific Antigen
GSI	Genome Similarity Index
HIV-1	Human immunodeficiency virus type-1
HLA	Human Leukocyte Antigen
IgG	Immunoglobulin G
Kb	Kilobase
KE	Kenya
LAg	Limiting antigen avidity
LANL	Los Alamos National Laboratory
Ltd	Limited
LTR	Long Terminal Repeats
MA (p17)	Matrix Protein
MAA	Multi Assay Algorithm
MAFFT	Multiple Alignment Using Fast Fourier Transform
MBC	Mixed Base Classifier
MEGA	Molecular Evolutionary Genetics Analysis
MW	Malawi
NAAT	Nucleic Acid Amplification Test
N/A	Not Available
NC	Nucleocapsid Protein
NHLS	National Health Laboratory Service
NRF	National Research Foundation
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PLWH	People Living with Human Immunodeficiency Virus
Pol	Polymerase Gene
PRF	Poliomyelitis Research Foundation
Pt	Participant
p2p7p1p6	Gag polyprotein
p24	Capsid protein
RNA	Ribonucleic Acid
RT	Reverse Transcriptase
TX	Texas
TZ	Tanzania
USA	United Sates of America
VL	Viral Load
WHO	World Health Organization
ZA	South Africa
ZM	Zambia
%	Percent
℃	Degree Celsius
µl	Microliter

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12985-026-03080-x.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements
Not applicable.

Author contributions
Conceptualization: SHM. Methodology: TL, PN, SHM. Funding acquisition: TL, SHM. Investigation: TL, PN, SHM. Supervision: PN, CTT, SHM. Project administration: TL, PN, SHM. Data analysis: TL, PN, SL, CTT, SHM. Writing – original draft: TL. Writing – review & editing: TL, PN, SL, CTT, SHM.

Funding
TL was funded by the National Research Foundation (NRF) and Poliomyelitis Research Foundation (# 24/30). SHM received funding from Poliomyelitis Research Foundation (#23/69), National Research Foundation and University of Pretoria Research Development Fund. CTT receives funding through the South African Research Chairs Initiative of the Department of Science and Innovation and National Research Foundation of South Africa (84177).

Data availability
All data generated or analysed during this study are included in this manuscript.

Declarations

Ethics approval and consent to participate
Ethics approval (reference number – 165/2024) was obtained for this study from the Faculty of Health Sciences Research Ethics Committee, University of Pretoria, South Africa. Written consent was obtained from all participants in the parent study (Ethics reference: 295/2015) [13].

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 12 June 2025 / Accepted: 13 January 2026
Published online: 02 February 2026

References

1. UNAIDS. Fact sheet 2025. 2025 [accessed: 01 September 2025]. https://www.unaids.org/sites/default/files/2025-07/2025_Global_HIV_Factsheet_en.pdf.
2. Trickey A, McGinnis K, Gill MJ, Abgrall S, Berenguer J, Wyen C, et al. Longitudinal trends in causes of death among adults with HIV on antiretroviral therapy in Europe and North America from 1996 to 2020: a collaboration of cohort studies. Lancet HIV. 2024;11:e176–85. https://doi.org/10.1016/s2352-3018(23)00272-2.
3. Nicol E, Jama NA, Mehlomakulu V, Hlongwa M, Pass D, Basera W, et al. Enhancing linkage to HIV care in the "Universal Test and Treat" era: barriers and enablers to HIV care among adults in a high HIV burdened district in KwaZulu-Natal, South Africa. BMC Public Health. 2023;23:1756. https://doi.org/10.1186/s12889-023-16576-w.
4. Frescura L, Godfrey-Faussett P, Feizzadeh AA, El-Sadr W, Syarif O, Ghys PD. Achieving the 95 95 95 targets for all: a pathway to ending AIDS. PLoS ONE. 2022;17:e0272405. https://doi.org/10.1371/journal.pone.0272405.
5. Hallett TB. Estimating the HIV incidence rate: recent and future developments. Curr Opin HIV AIDS. 2011;6:102–7. https://doi.org/10.1097/COH.0b013e328343bfdb.
6. Zuma K, Simbayi L, Zungu N, Moyo S, Marinda E, Jooste S, et al. The HIV Epidemic in South Africa Key Findings from 2017 National Population Based Survey. Int J Environ Res Public Health. 2022. https://doi.org/10.3390/ijerph19138125.
7. Park SY, Love TMT, Reynell L, Yu C, Kang TM, Anastos K, et al. The HIV genomic incidence assay meets false recency rate and mean duration of recency infection performance standards. Sci Rep. 2017;7:7480. https://doi.org/10.1038/s41598-017-07490-4.
8. Klock E, Wilson E, Fernandez RE, Piwowar-Manning E, Moore A, Kosloff B, et al. Validation of population-level HIV-1 incidence estimation by cross-sectional incidence assays in the HPTN 071 (PopART) trial. J Int AIDS Soc. 2021;24:e25830. https://doi.org/10.1002/jia2.25830.

9. Mustanski B, Ryan DT, Newcomb ME, D'Aquila RT, Matson M. Very high HIV incidence and associated risk factors in a longitudinal cohort study of diverse adolescent and young adult men who have sex with men and transgender women. AIDS Behav. 2020;24:1966–75. https://doi.org/10.1007/s10461-019-02766-4.
10. Laeyendecker O, Konikoff J, Morrison DE, Brookmeyer R, Wang J, Celum C, et al. Identification and validation of a multi-assay algorithm for cross-sectional HIV incidence estimation in populations with subtype C infection. J Int AIDS Soc. 2018. https://doi.org/10.1002/jia2.25082.
11. Grant-McAuley W, Klock E, Laeyendecker O, Piwowar-Manning E, Wilson E, Clarke W, et al. Evaluation of multi-assay algorithms for identifying individuals with recent HIV infection: HPTN 071 (PopART). PLoS ONE. 2021;16:e0258644. https://doi.org/10.1371/journal.pone.0258644.
12. Taiwo BO, Hicks CB. Primary human immunodeficiency virus. South Med J. 2002;95:1312–7.
13. Mayaphi SH, Martin DJ, Quinn TC, Laeyendecker O, Olorunju SA, Tintinger GR, et al. Detection of acute and early HIV-1 infections in an HIV hyper-endemic area with limited resources. PLoS ONE. 2016;11:e0164943. https://doi.org/10.1371/journal.pone.0164943.
14. Nkone P, Loubser S, Quinn TC, Redd AD, Laeyendecker O, Tiemessen CT, et al. Evaluation of the HIV-1 polymerase gene sequence diversity for prediction of recent HIV-1 infections using Shannon entropy analysis. Viruses. 2022. https://doi.org/10.3390/v14071587.
15. Kin-On Lau J, Murdock N, Murray J, Justman J, Parkin N, Miller V. A systematic review of limiting antigen avidity enzyme immunoassay for detection of recent HIV-1 infection to expand supported applications. J Virus Erad. 2022;8:100085. https://doi.org/10.1016/j.jve.2022.100085.
16. Shepherd SJ, McAllister G, Kean J, Wallace LA, Templeton KE, Goldberg DJ, et al. Development of an avidity assay for detection of recent HIV infections. J Virol Methods. 2015;217:42–9. https://doi.org/10.1016/j.jviromet.2015.02.016.
17. Serhir B, Hamel D, Doualla-Bell F, Routy JP, Beaulac SN, Legault M, et al. Performance of Bio-Rad and limiting antigen avidity assays in detecting recent HIV infections using the Quebec primary HIV-1 infection cohort. PLoS ONE. 2016;11:e0156023. https://doi.org/10.1371/journal.pone.0156023.
18. Wei X, Liu X, Dobbs T, Kuehl D, Nkengasong JN, Hu DJ, et al. Development of two avidity-based assays to detect recent HIV type 1 seroconversion using a multisubtype gp41 recombinant protein. AIDS Res Hum Retroviruses. 2010;26:61–71. https://doi.org/10.1089/aid.2009.0133.
19. Shah NS, Duong YT, Le LV, Tuan NA, Parekh BS, Ha HTT, et al. Estimating false-recent classification for the limiting-antigen avidity EIA and BED-capture enzyme immunoassay in Vietnam: implications for HIV-1 incidence estimates. AIDS Res Hum Retroviruses. 2017;33:546–54. https://doi.org/10.1089/aid.2016.0203.
20. Konikoff J, Brookmeyer R, Longosz AF, Cousins MM, Celum C, Buchbinder SP, et al. Performance of a limiting-antigen avidity enzyme immunoassay for cross-sectional estimation of HIV incidence in the United States. PLoS ONE. 2013;8:e82772. https://doi.org/10.1371/journal.pone.0082772.
21. Grant-McAuley W, Laeyendecker O, Monaco D, Chen A, Hudelson SE, Klock E, et al. Evaluation of multi-assay algorithms for cross-sectional HIV incidence estimation in settings with universal antiretroviral treatment. BMC Infect Dis. 2022;22:838. https://doi.org/10.1186/s12879-022-07850-0.
22. Torres TS, Teixeira SLM, Hoagland B, Konda KA, Derrico M, Moreira RI, et al. Recent HIV infection and annualized HIV incidence rates among sexual and gender minorities in Brazil and Peru (ImPrEP seroincidence study): a cross-sectional, multicenter study. Lancet Reg Health Am. 2023;28:100642. https://doi.org/10.1016/j.lana.2023.100642.
23. Kafando A, Fournier E, Serhir B, Martineau C, Doualla-Bell F, Sangaré MN, et al. HIV-1 envelope sequence-based diversity measures for identifying recent infections. PLoS ONE. 2017;12:e0189999. https://doi.org/10.1371/journal.pone.0189999.
24. Puller V, Neher R, Albert J. Estimating time of HIV-1 infection from next-generation sequence diversity. PLoS Comput Biol. 2017;13:e1005775. https://doi.org/10.1371/journal.pcbi.1005775.
25. Leslie AJ, Pfafferott KJ, Chetty P, Draenert R, Addo MM, Feeney M, et al. HIV evolution: CTL escape mutation and reversion after transmission. Nat Med. 2004;10:282–9. https://doi.org/10.1038/nm992.
26. Liu MK, Hawkins N, Ritchie AJ, Ganusov VV, Whale V, Brackenridge S, et al. Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. J Clin Invest. 2013;123:380–93. https://doi.org/10.1172/jci65330.
27. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature. 2004;432:769–75. https://doi.org/10.1038/nature03113.
28. Zhang Y, Peng Y, Yan H, Xu K, Saito M, Wu H, et al. Multilayered defense in HLA-B51-associated HIV viral control. J Immunol. 2011;187:684–91. https://doi.org/10.4049/jimmunol.1100316.
29. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, et al. Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection. AIDS. 2003;17:1871–9. https://doi.org/10.1097/00002030-200309050-00005.
30. Ritchie AJ, Cai F, Smith NM, Chen S, Song H, Brackenridge S, et al. Recombination-mediated escape from primary CD8+ T cells in acute HIV-1 infection. Retrovirology. 2014;11:69. https://doi.org/10.1186/s12977-014-0069-9.
31. Claiborne DT, Prince JL, Scully E, Macharia G, Micci L, Lawson B, et al. Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4+ T cells, and disease progression. Proc Natl Acad Sci U S A. 2015;112:E1480–9. https://doi.org/10.1073/pnas.1421607112.
32. Iyer SS, Bibollet-Ruche F, Sherrill-Mix S, Learn GH, Plenderleith L, Smith AG, et al. Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness. Proc Natl Acad Sci U S A. 2017;114:E590–9. https://doi.org/10.1073/pnas.1620144114.
33. Kijak GH, Sanders-Buell E, Chenine AL, Eller MA, Goonetilleke N, Thomas R, et al. Rare HIV-1 transmitted/founder lineages identified by deep viral sequencing contribute to rapid shifts in dominant quasispecies during acute and early infection. PLoS Pathog. 2017;13:e1006510. https://doi.org/10.1371/journal.ppat.1006510.
34. Salazar-Gonzalez JF, Salazar MG, Keele BF, Learn GH, Giorgi EE, Li H, et al. Genetic identity, biological phenotype, and evolutionary pathways of transmitted/founder viruses in acute and early HIV-1 infection. J Exp Med. 2009;206:1273–89. https://doi.org/10.1084/jem.20090378.
35. Treurnicht FK, Seoighe C, Martin DP, Wood N, Abrahams MR, Rosa Dde A, et al. Adaptive changes in HIV-1 subtype C proteins during early infection are driven by changes in HLA-associated immune pressure. Virology. 2010;396:213–25. https://doi.org/10.1016/j.virol.2009.10.002.
36. Permar SR, Salazar MG, Gao F, Cai F, Learn GH, Kalilani L, et al. Clonal amplification and maternal-infant transmission of nevirapine-resistant HIV-1 variants in breast milk following single-dose nevirapine prophylaxis. Retrovirology. 2013;10:88. https://doi.org/10.1186/1742-4690-10-88.
37. Wright JK, Brumme ZL, Carlson JM, Heckerman D, Kadie CM, Brumme CJ, et al. Gag-protease-mediated replication capacity in HIV-1 subtype C chronic infection: associations with HLA type and clinical parameters. J Virol. 2010;84:10820–31. https://doi.org/10.1128/jvi.01084-10.
38. Sharma S, Balakrishnan P, Solomon S, Ranga U. A prospective analysis of viral immune escape in the chronic phase of the subtype C HIV-1 infections of India.
39. Parrish NF, Wilen CB, Banks LB, Iyer SS, Pfaff JM, Salazar-Gonzalez JF, et al. Transmitted/founder and chronic subtype C HIV-1 use CD4 and CCR5 receptors with equal efficiency and are not inhibited by blocking the integrin α4β7. PLoS Pathog. 2012;8:e1002686. https://doi.org/10.1371/journal.ppat.1002686.
40. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. AIDS Rev. 2006;8:125–40.
41. Espinosa E. Finding the traces of the HIV selection. Rev Latinoam Microbiol. 2006;48:84–90.
42. Avila-Ríos S, Reyes-Terán G, Espinosa E. Cornering HIV: taking advantage of interactions between selective pressures. Med Hypotheses. 2007;69:422–31. https://doi.org/10.1016/j.mehy.2006.12.012.
43. Berger CT, Carlson JM, Brumme CJ, Hartman KL, Brumme ZL, Henry LM, et al. Viral adaptation to immune selection pressure by HLA class I-restricted CTL responses targeting epitopes in HIV frameshift sequences. J Exp Med. 2010;207:61–75. https://doi.org/10.1084/jem.20091808.
44. Miller WC, Rosenberg NE, Rutstein SE, Powers KA. Role of acute and early HIV infection in the sexual transmission of HIV. Curr Opin HIV AIDS. 2010;5:277–82. https://doi.org/10.1097/COH.0b013e32833a0d3a.
45. Gbadamosi SO, Trepka MJ, Dawit R, Jebai R, Sheehan DM. A systematic review and meta-analysis to estimate the time from HIV infection to diagnosis for people with HIV. AIDS Rev. 2022;24:32–40. https://doi.org/10.24875/AIDSRev.21000007.
46. Monaco DC, Ende Z, Hunter E. Virus-host gene interactions define HIV-1 disease progression. Curr Top Microbiol Immunol. 2017;407:31–63. https://doi.org/10.1007/82_2017_33.
47. Boutwell CL, Rolland MM, Herbeck JT, Mullins JI, Allen TM. Viral evolution and escape during acute HIV-1 infection. J Infect Dis. 2010;202(2):S309–14. https://doi.org/10.1086/655653.

48. Gounder K, Padayachi N, Mann JK, Radebe M, Mokgoro M, van der Stok M, et al. High frequency of transmitted HIV-1 Gag HLA class i-driven immune escape variants but minimal immune selection over the first year of clade C infection. PLoS ONE. 2015;10:e0119886. https://doi.org/10.1371/journal.pone.0119886.
49. Chopera DR, Ntale R, Ndabambi N, Garrett N, Gray CM, Matten D, et al. Early evolution of human leucocyte antigen-associated escape mutations in variable Gag proteins predicts CD4+ decline in HIV-1 subtype C-infected women. AIDS. 2017;31:191–7. https://doi.org/10.1097/qad.0000000000001298.
50. Avila-Rios S, Carlson JM, John M, Mallal S, Brumme ZL. Clinical and evolutionary consequences of HIV adaptation to HLA: implications for vaccine and cure. Curr Opin HIV AIDS. 2019;14:194–204. https://doi.org/10.1097/coh.0000000000000541.
51. Hauser A, Santos-Hoevener C, Meixenberger K, Zimmermann R, Somogyi S, Fiedler S, et al. 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. 2014;9:e98038. https://doi.org/10.1371/journal.pone.0098038.
52. Hauser A, Heiden MA, Meixenberger K, Han O, Fiedler S, Hanke K, et al. Evaluation of a BioRad avidity assay for identification of recent HIV-1 infections using dried serum or plasma spots. J Virol Methods. 2019;266:114–20. https://doi.org/10.1016/j.jviromet.2019.02.002.
53. Moyo S, LeCuyer T, Wang R, Gaseitsiwe S, Weng J, Musonda R, et al. Evaluation of the false recent classification rates of multiassay algorithms in estimating HIV type 1 subtype C incidence. AIDS Res Hum Retroviruses. 2014;30:29–36. https://doi.org/10.1089/aid.2013.0055.
54. Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. Curr Opin HIV AIDS. 2019;14:153–60. https://doi.org/10.1097/coh.0000000000000534.
55. Buonaguro L, Tornesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. J Virol. 2007;81:10209–19. https://doi.org/10.1128/jvi.00872-07.
56. Bärnighausen T, McWalter TA, Rosner Z, Newell ML, Welte A. HIV incidence estimation using the BED capture enzyme immunoassay: systematic review and sensitivity analysis. Epidemiology. 2010;21:685–97. https://doi.org/10.1097/EDE.0b013e3181e9e978.
57. Schlusser KE, Konikoff J, Kirkpatrick AR, Morrison C, Chipato T, Chen PL, et al. Short communication: comparison of Maxim and Sedia limiting antigen assay performance for measuring HIV incidence. AIDS Res Hum Retroviruses. 2017;33:555–7. https://doi.org/10.1089/aid.2016.0245.
58. Chen A, Laeyendecker O, Eshleman SH, Monaco DR, Kammers K, Larman HB, et al. A top scoring pairs classifier for recent HIV infections. Stat Med. 2021;40:2604–12. https://doi.org/10.1002/sim.8920.
59. Wendel SK, Longosz AF, Eshleman SH, Blankson JN, Moore RD, Keruly JC, et al. Short communication: the impact of viral suppression and viral breakthrough on limited-antigen avidity assay results in individuals with clade B HIV infection. AIDS Res Hum Retroviruses. 2017;33:325–7. https://doi.org/10.1089/aid.2016.0105.
60. Ragonnet-Cronin M, Aris-Brosou S, Joanisse I, Merks H, Vallée D, Caminiti K, et al. 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. 2012;206:756–64. https://doi.org/10.1093/infdis/jis411.
61. Andersson E, Shao W, Bontell I, Cham F, Cuong do D, Wondwossen A, et al. 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. 2013;18:125–31. https://doi.org/10.1016/j.meegid.2013.03.050.
62. Kouyos RD, von Wyl V, Yerly S, Böni J, Rieder P, Joos B, et al. Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. Clin Infect Dis. 2011;52:532–9. https://doi.org/10.1093/cid/ciq164.
63. Zheng DP, Rodrigues M, Bile E, Nguyen DB, Diallo K, DeVos JR, et al. Molecular characterization of ambiguous mutations in HIV-1 polymerase gene: implications for monitoring HIV infection status and drug resistance. PLoS ONE. 2013;8:e77649. https://doi.org/10.1371/journal.pone.0077649.
64. Wu JW, Patterson-Lomba O, Novitsky V, Pagano M. A generalized entropy measure of within-host viral diversity for identifying recent HIV-1 infections. Medicine Baltimore. 2015;94:e1865. https://doi.org/10.1097/md.0000000000001865.
65. Alves BM, Siqueira JD, Prellwitz IM, Botelho OM, Da Hora VP, Sanabani S, et al. Estimating HIV-1 genetic diversity in Brazil through next-generation sequencing. Front Microbiol. 2019;10:749. https://doi.org/10.3389/fmicb.2019.00749.
66. Russell ML, Fish CS, Drescher S, Cassidy NAJ, Chanana P, Benki-Nugent S, et al. Using viral sequence diversity to estimate time of HIV infection in infants. PLoS Pathog. 2023;19:e1011861. https://doi.org/10.1371/journal.ppat.1011861.
67. Xi X, Spencer SEF, Hall M, Grabowski MK, Kagaayi J, Ratmann O. Inferring the sources of HIV infection in Africa from deep-sequence data with semi-parametric Bayesian Poisson flow models. J R Stat Soc Ser C Appl Stat. 2022;71:517–40. https://doi.org/10.1111/rssc.12544.
68. Faraci G, Park SY, Love TMT, Dubé MP, Lee HY. Precision detection of recent HIV infections using high-throughput genomic incidence assay. Microbiol Spectr. 2023;11:e0228523. https://doi.org/10.1128/spectrum.02285-23.
69. Carlisle LA, Turk T, Kusejko K, Metzner KJ, Leemann C, Schenkel CD, et al. Viral diversity based on next-generation sequencing of HIV-1 provides precise estimates of infection recency and time since infection. J Infect Dis. 2019;220:254–65. https://doi.org/10.1093/infdis/jiz094.
70. Tarosso LF, Vieira VA, Sauer MM, Tomiyama HI, Kalil J, Kallas EG. Conserved HIV-1 gag p24 epitopes elicit cellular immune responses that impact disease outcome. AIDS Res Hum Retroviruses. 2017;33:832–42. https://doi.org/10.1089/aid.2016.0168.
71. Masemola AM, Mashishi TN, Khoury G, Bredell H, Paximadis M, Mathebula T, et al. Novel and promiscuous CTL epitopes in conserved regions of gag targeted by individuals with early subtype C HIV type 1 infection from southern Africa. J Immunol. 2004;173:4607–17. https://doi.org/10.4049/jimmunol.173.7.4607.
72. Goulder PJ, Walker BD. HIV and HLA class I: an evolving relationship. Immunity. 2012;37:426–40. https://doi.org/10.1016/j.immuni.2012.09.005.
73. Jiang F, Han X, Zhang H, Zhao B, An M, Xu J, et al. Multi-layered gag-specific immunodominant responses contribute to improved viral control in the CRF01_AE subtype of HIV-1-infected MSM subjects. BMC Immunol. 2016;17:28. https://doi.org/10.1186/s12865-016-0166-8.
74. Lichterfeld M, Yu XG, Le Gall S, Altfeld M. Immunodominance of HIV-1-specific CD8(+) T-cell responses in acute HIV-1 infection: at the crossroads of viral and host genetics. Trends Immunol. 2005;26:166–71. https://doi.org/10.1016/j.it.2005.01.003.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.