Title

Higher HIV-1 genetic diversity is associated with AIDS and neuropsychological impairment

Journal

Virology, 433(2)

ISSN

0042-6822

Authors

Hightower, George K

Wong, Joseph K

Letendre, Scott L

et al.

Publication Date

2012-11-01

DOI

10.1016/j.virol.2012.08.028

Peer reviewed

Published in final edited form as:
Virology. 2012 November 25; 433(2): 498–505. doi:10.1016/j.virol.2012.08.028.

Higher HIV-1 Genetic Diversity is Associated with AIDS and Neuropsychological Impairment

George K. Hightower[1], **Joseph K. Wong**[2], **Scott L. Letendre**[1], **Anya A. Umlauf**[1], **Ronald J. Ellis**[1,3], **Caroline C. Ignacio**[1], **Robert K. Heaton**[1], **Ann C. Collier**[4], **Christina M. Marra**[4], **David B. Clifford**[5], **Benjamin B. Gelman**[6], **Justin C. McArthur**[7], **Susan Morgello**[8], **David M. Simpson**[8], **JA McCutchan**[1], **Igor Grant**[1,3], **Susan J. Little**[1], **Douglas D. Richman**[1,3], **Sergei L. Kosakovsky Pond**[1], **Davey M. Smith**[1,3], and **the CHARTER Study Group***

[1]University of California San Diego, La Jolla, California, USA

[2]San Francisco Veterans Affairs Medical Center, San Francisco, California, USA

[3]Veterans Affairs San Diego Healthcare System San Diego, California, USA

[4]University of Washington, Seattle, Washington, USA

[5]Washington University, St. Louis, Missouri, USA

[6]University of Texas Medical Branch, Galveston, Texas, USA

[7]Johns Hopkins University, Baltimore, Maryland, USA

[8]Mount Sinai School of Medicine, New York, New York, USA

Abstract

Standard methods used to estimate HIV-1 population diversity are often resource intensive (e.g., single genome amplification, clonal amplification and pyrosequencing) and not well suited for large study cohorts. Additional approaches are needed to address the relationships between intraindividual HIV-1 genetic diversity and disease. With a small cohort of individuals, we validated three methods for measuring diversity: Shannon entropy and average pairwise distance (APD) using single genome sequences, and counts of mixed bases (i.e. ambiguous nucleotides) from population-based sequences. In a large cohort, we then used the mixed base approach to determine associations between measure HIV-1 diversity and HIV associated disease. Normalized counts of mixed bases correlated with Shannon Entropy at both the nucleotide (rho=0.72, p=0.002) and amino acid level (rho=0.59, p=0.015), and APD (rho=0.75, p=0.001). Among participants who underwent neuropsychological and clinical assessments (n=187), increased HIV-1 population diversity was associated with both a diagnosis of AIDS and neuropsychological impairment.

Keywords

HIV; AIDS; genetic diversity; neuropsychological impairment; viral population dynamics

Corresponding author: Davey Smith, MD, MAS, Address of Corresponding author: 9500 Gilman Drive # 0679, La Jolla, CA 92093-0679, Telephone: (619) 543-8080, Fax: (858) 552-7445, davey@ucsd.edu.
*The CNS HIV Antiretroviral Therapy Effects Research, University of California San Diego, San Diego

Introduction

HIV-1 is usually transmitted as a small number of relatively homogenous variants, but exists within an infected individual as a highly diverse viral population(Keele et al., 2008; Wolinsky et al., 1996b; Wood et al., 2009). Additionally, viral RNA model systems suggest that maintaining a highly diverse viral population may provide an adaptive advantage and can alter virulence(Domingo, Menendez-Arias, and Holland, 1997; Pfeiffer and Kirkegaard, 2005; Pfeiffer and Kirkegaard, 2006; Vignuzzi et al., 2006). Additional study is needed to better characterize both HIV-1 diversity within an individual and the mechanistic underpinnings of HIV-1 virulence.

Standard methods used to estimate HIV-1 population diversity in the setting of natural infection are resource intensive (e.g. single genome amplification, clonal amplification, pyrosequencing and heteroduplex assay), which likely explains why past studies that examined intra-individual HIV-1 population diversity and disease progression relied on relatively small cohorts (range n=6–44) (Delwart et al., 1997; Ganeshan et al., 1997; Markham et al., 1998; Poon et al., 2007; Ross and Rodrigo, 2002; Shankarappa et al., 1999a; Strunnikova et al., 1998; Wolinsky et al., 1996a). To enhance the study of HIV-1 diversity, we validated a method for estimating HIV-1 population diversity that relies on HIV-1 *pol* sequence data generated as part of standard clinical care(Cornelissen et al., 2007; Kouyos et al., 2011; Poon et al., 2007). Specifically, this study investigated how ambiguous or ‘mixed bases’ contained in population-based HIV-1 *pol* sequences were associated with viral population diversity and neurocognitive impairment. After validating this approach, we studied the relationship between intra-individual HIV-1 population diversity and disease in a large cohort of clinically well-characterized HIV-1 infected individuals who underwent neuromedical and comprehensive neuropsychological testing examinations.

Methods

Ethics Statement

The Human Research Protection Program of the University of California San Diego approved the use of all participant data and samples. In accordance with the Human Research Protection Program of the University of California, informed consistent was provided by all study participants.

Study Design

Our study utilized two different participant cohorts. From the UCSD Primary Infection Cohort, we used blood samples to investigate if mixed base counts provide reliable estimates of HIV-1 diversity. From the CNS HIV Antiretroviral Therapy Effects Research (CHARTER) Cohort, we used blood and cerebrospinal fluid (CSF) samples to investigate the relationship between intra- host HIV-1 population diversity and disease.

Participants enrolled in the UCSD Primary Infection Cohort were included in this study if they were antiretroviral naive, had two or more longitudinally collected blood samples, and single genome and population-based sequences were available from blood samples collected within a month of each other. All participants in the cohort were enrolled within 70 days of the estimated date of infection, as previously described (Morris et al., 2010). Blood samples from dates at which an individual was suspected of dual infection with previously published criteria were excluded(Cornelissen et al., 2007). Using a longitudinal cohort allowed us to screen for dual infections and decrease any irregularities in our validation studies that might be attributable to dual infection. When two or more sample dates were eligible from the same participant, the last date was used. All participants enrolled in the CHARTER Cohort were included in this study if viral loads were high enough (~500 copies/ml) to generate

population-based sequences of HIV-1 *pol*, as described below. CHARTER participants received neuromedical, neuropsychological, and laboratory examinations. Standardized neuropsychological assessments tested seven ability domains (learning, delayed recall, verbal fluency, processing speed, attention/working memory, abstraction/executive functioning, motor speed), as previously described (Woods et al., 2004). Results were summarized using global ratings that ranged from 1 (above average) to 9 (severely impaired) with a global score of 5 or higher used to categorize participants as impaired (Antinori et al., 2007; Heaton et al., 2010). Also, for each participant, the "Frascati" criteria (incidental, contributing, confounding) were used to classify non-HIV-1 related conditions that increase risk for neurocognitive impairment (Antinori et al., 2007). Infection duration in this cohort was available only as self-reported information. Among CHARTER participants, AIDS diagnosis was based on the 1993 Centers for Disease Control guidelines [www.cdc.gov/hiv/resources/guidelines].

Sample Collection and Processing

For both cohorts, blood plasma was collected by venipucture and CSF by lumbar puncture, and stored at 80 °C. For blood samples collected from participants enrolled in the UCSD Primary Infection Cohort, HIV RNA was extracted using the QIAamp ViralRNA Mini Kit (Qiagen, Hilden, Germany) per manufacturer's instructions. For blood and CSF samples collected from CHARTER Cohort participants, HIV RNA was extracted using the ViroSeq v.2.0 HIV genotyping system (Applied Biosystems, Foster City, CA, USA).

Single Genome Amplification and Sequencing

Extracted HIV RNA was used to generate cDNA using random decamers and the RETROscript kit (Ambion, Applied Biosystems, Foster City, CA, USA) per manufacturer's instructions. The first round of nested PCR was performed with 10 uL of diluted cDNA template and 40 uL of reaction mix, which included the following outer primer set (CI-Pol 5′, 3RT 3′). The second round PCR, was performed with 5 uL of the first round product and 45uL of reaction mixture which included the following inner primer set (5RT 5′, 3RT 3′). Reactions were performed in parallel on a 96-well plate, and the second-round products were visualized to ensure no more than 30% of the reaction wells were positive, as previously described(Butler et al., 2009).

Population-based Pol Sequencing

The ViroSeq HIV genotyping system (Ambion Applied Biosystems) was used for population-based HIV-1 *pol* sequencing per manufacturer's instructions. Following cDNA synthesis, polymerase chain reaction (PCR) amplification was used to generate a 1500-bp amplicon including the entire protease and first two thirds of the reverse transcriptase. The resulting PCR product was sequenced using the ABI 3100 Genetic Analyzer, sequences were manually edited and resistance associated mutations were determined using the ViroSeq genotyping software.

Evaluating and Comparing Measures of Intra-individual HIV-1 Diversity

Using HIV RNA extracted from the blood samples of participants enrolled in the UCSD Primary Infection Cohort, we evaluated HIV-1 population diversity with three different methods: 1) Shannon Entropy using single genome, 2) average pairwise distance (APD) using single genome sequences and 3) counts of mixed bases (i.e., ambiguous nucleotides) in population-based sequences. Mixed bases were defined as double peaks in the electropherograms of the HIV-1 *pol* sequences that were identified by the basecaller Viroseq program.

A batch file implemented in HyPhy (Kosakovsky Pond, Frost, and Muse, 2005) was used to provide a count of total mixed bases, synonymous mixed bases and nonsynonymous mixed bases in population-based *pol* sequences. Each count was divided by sequence length to provide a normalized index: Total Mixed Base Index (TM-Index), Synonymous Mixed Base Index (SM-Index) and Amino Acid Residue Mixed Base Index (RM-Index). In summary, the TM-Index was simply the count of all mixed nucleotide bases in each HIV-*pol* sequence divided by the sequence length. The SM-Index was the count of only those mixed bases that could not change the protein sequence, i.e. synonymous change, divided by the sequence length, while the RM-Index considered only those bases that could change the protein sequence, i.e. non-synonymous change. Shannon Entropy was measured for nucleotide sequences and amino acid sequences using the HIV Los Alamos National Laboratory (LANL) Entropy- One tool (http://www.hiv.lanl.gov, accessed April 2010). For each participant's set of single genome sequences, Entropy-One results were summarized as the number of sites with a Shannon Entropy value ≥ .3 divided by the number of positions in the alignment. APD was measured by analyzing participant specific single genome sequences with HyPHy software using the Tamura-Nei 93 model (TN93) (Tamura and Nei, 1993), as previously described (Smith et al., 2009).

Estimating Cost and Time

Estimates of the comparative costs and labor required for population-based sequencing, single genome sequencing (SGS) and ultra deep sequencing (UDS) were based on previously published analysis(Bushman et al., 2008). The estimated and anticipated sequencing costs for measuring HIV-1 population diversity were calculated from the costs of reagents, disposable materials, kits and sequencing runs. The estimated and anticipated labor was calculated as the labor time plus instrument time required to perform sequencing.

Statistical Analysis

All tests were two-sided and interpreted at 5% significance level. Tests for associations between two continuous variables were performed using Spearman rank correlation. Logistic regression analyses were used to investigate the relationships between HIV-1 diversity (measured by the TM-Index, SM-Index and RM-Index) and AIDS diagnosis and between HIV-1 diversity and neuropsychological impairment. Additional variables considered in multivariable logistic regression analyses, where appropriate, included HIV RNA levels in CSF and blood, duration of infection, current and nadir CD4, and ART history. The model investigating neuropsychological impairment also controlled for AIDS and non-HIV-1 related conditions that increase risk for neurocognitive impairment. The odds ratios were calculated for a 0.01 unit increase in each mixed base index.

Results

Study Participants

Participants from the San Diego Primary Infection cohort (n=16) selected to investigate the use of mixed base indices, were mostly white men in their early thirties. The three major reasons participants enrolled in the Primary Infection Cohort did not meet inclusion criteria were: 1) the absence of longitudinal samples to assess the potential of dual infection, 2) the absence of single genome and population-based sequences performed on blood samples collected within a month of each other, and 3) presence of HIV-1 dual infection as defined by previously published methods(Cornelissen et al., 2007).

The participants from the CHARTER cohort (n=187) selected to investigate the relationship between HIV-1 population diversity and disease were in large part black (49%) men (80%) in their early forties and 23% were receiving antiretroviral therapy at the time of sampling.

Median blood and CSF HIV RNA levels were 4.6 and 3.6 $\log_{10}$ copies/ml, 43% had an AIDS diagnosis, and the median current and nadir CD4 counts were 362 and 251 cells/ul (Table 1). Drug resistance mutations were not common; M184V and K103N were the two most prevalent. The M184V mutation was detected in 8 blood samples (5%) and 7 CSF samples (4%). The K103N mutation was detected in 13 blood samples (7%) and 14 CSF samples (8%).

Mixed Base Counts and Evaluating Intra-individual HIV-1 Diversity

Using single genome and population-based sequences generated from San Diego Primary Infection Cohort participants, we measured and compared Shannon Entropy of nucleotide sequences and APD to two mixed base indices structured to estimate nucleotide diversity (Total Mixed Base Index [TM-Index] and Synonomous Mixed Base Index [SM-Index]). Also, Shannon Entropy of amino acid sequences was compared to the mixed base index structured to measure amino acid diversity (Residue Mixed Index [RM-Index]). There was an average of 26 single genome sequences per San Diego Primary Infection participants (range 18–32). In brief, all mixed base indices strongly correlated with other measures of viral population diversity (figures 1 and 2). Specifically, the TM-Index correlated with both Shannon Entropy (nucleotide) (rho = 0.76, p = 0.001) (figure 1a) and with APD (rho = 0.80, p < 0.001) (figure 2a). The SM-Index also correlated with both Shannon Entropy (nucleotide) (rho = 0.72, p = 0.002) (figure 1b) and with APD (rho = 0.75, p = 0.001) (figure 2b). The RM-Index correlated with Shannon Entropy (amino acid) (rho = 0.59, p = 0.015) (figure 1c).

HIV-1 Population Diversity, AIDS and Neuropsychological Performance

Using population-based HIV-1 *pol* sequence data, AIDS diagnosis and standard classifications of neurocognitive status from CHARTER cohort participants, we tested a number of statistical models that, where appropriate, included HIV RNA levels in CSF and blood, duration of infection, current and nadir CD4, ART history and co-morbidity status (incidental, contributing, confounding). We found a diagnosis of AIDS was independently associated with: (i) TM-Indices in blood (blood OR = 2.0, p = 0.015) and (ii) RM-Index in blood (OR = 6.5, p = 0.018) (Table 2). A trend for an association between AIDS diagnosis with the SM-Index in blood was also observed (OR = 1.8, p = 0.067). In modeling neuropsychological impairment, nadir CD4 was initially included but eventually excluded because current CD4 was a stronger predictor of impairment. Concerning neuropsychological impairment, we found that it was independently associated with the TM-index in blood (OR = 1.9, p = 0.01) and RM-Index in blood (OR = 11.4, p = 0.001) (Table 3). There was also a trend for the CSF TM-Index (p = 0.08) to be associated with neuropsychological impairment. For all of these models, the association between viral diversity and neurocognitive impairment was independent of diagnosed comorbid dieases, as determined by “Frascati” criteria(Antinori et al., 2007).

Intra-individual HIV-1 Diversity and Viral Dynamics in Blood and CSF

We further clarified the relationship between HIV diversity and the sampled compartments. In every case, HIV-1 population diversity in blood plasma correlated with diversity in the CSF for all mixed base indices (Figures 3a–c). In addition, for most individuals, the RM-Index was higher for viral populations derived from the CSF than for blood (0.004 vs. 0.037, p=0.059), but this was not observed for the TM-Index (CSF 0.0143 vs. blood 0.0137, p=0.115) and SM-Index (CSF 0.010 vs. blood 0.010, p=0.263). These findings were not explained by pleocytosis, as defined by >5 white blood cells/ml of CSF(Smith et al., 2009). The correlation between blood TM-index and CSF TM-Index did not differ between participants with or without pleocytosis (rho=0.641 vs. rho=0.728, p= 0.321). When we examined nucleotide diversity in blood and CSF in relation to time, we found that the

duration of infection had linear relationships with both blood TM-index ($p < 0.001$) (figure 4a) and blood SM-Index ($p < 0.001$) (figure 4b). In contrast, the relationship between duration of infection and amino acid diversity in blood, as measured by the RM-Index, was best described by quadratic regression ($p = 0.07$), where the RM-Index steadily increased until approximately year 12 of infection and then steadily declined (figure 4c).

Comparing Cost and Time

Using population-based HIV-1 *pol* sequences to investigate HIV viral population diversity in both blood and CSF for the CHARTER cohort instead of SGS or UDS saved approximately $298,078 ($390,830 for SGS and $298,078 for UDS) and 2,431 hours of labor (14,586 hours for SGS and 2,431 hours for UDS). The costs saving are based on the estimated cost per sample of $278 for population-based sequencing, $1,323 for SGS, and $1,075 for UDS. The labor savings are based on the estimated labor time per sample of 3 hours for population-based sequencing, 42 hours for SGS, and 9.5 hours for UDS.

Discussion

This study investigated the use of a novel approach for measuring intra-host viral population diversity and the relationship between viral genetic diversity and HIV-1 disease state. The approach evaluated HIV-1 population diversity by providing a normalized count of mixed bases from population-based sequences. Three different measures of nucleotide variability were evaluated: Total Mixed Base Index (TM-Index), Synonymous Mixed Base Index (SM-Index) and amino acid or Residue Mixed Base Index (RM-Index). Since the mixed base methods required only one population-based sequence to evaluate population diversity, they were well suited to analyze HIV-1 RNA populations in blood and CSF from a relatively large study cohort (n=187).

The time and cost associated with performing standard methods of evaluating HIV-1 genotypic population diversity likely explain why previous studies have used relatively small sample sizes (Delwart et al., 1997; Ganeshan et al., 1997; Markham et al., 1998; Poon et al., 2007; Ross and Rodrigo, 2002; Shankarappa et al., 1999a; Strunnikova et al., 1998; Wolinsky et al., 1996a). To address this methodological limitation, we investigated the use of population-based sequences and a normalized count of mixed bases to evaluate HIV-1 population diversity. We found the mixed base method was comparable to Shannon Entropy and APD, two widely used methods for measuring HIV-1 population diversity. This approach builds on previous studies that have used the occurrence of mixed bases in population-based sequences to screen for HIV-1 dual infection(Cornelissen et al., 2007) and as a site-specific marker of variation(Kouyos et al., 2011; Poon et al., 2007). Using the normalized mixed base approach, we found that increased HIV-1 population diversity was associated with: AIDS (TM-Index measured in blood and RM-Index in blood) and neuropsychological impairment (TM-Index measured in blood and RM-Index in blood). These associations between HIV-1 population diversity and disease state, suggest that HIV-1 virulence is most likely associated with the capacity of a viral population to maintain genetic diversity and not simply a specific and predominant genotypic variant, although the exact underlying mechanism for this remains unclear. For complex phenotypes such as neuropsychological impairment, HIV-1 variants may work in a coordinated manner to cause disease. As a hypothesis, neuropathogenesis may require multiple distinct viral variants. For example, one variant with high viral replication in the CNS may trigger inflammation and prime other cells for infection, while a second variant replicates in a specific tissue that disrupts a process essential for optimal neuropsychological function. This hypothesis builds on studies of poliovirus that link viral diversity and pathogenesis(Bushell and Sarnow, 2002; Pfeiffer and Kirkegaard, 2005; Pfeiffer and Kirkegaard, 2006; Vignuzzi et al., 2006), but

these and other hypotheses need to be tested to better characterize the mechanisms of HIV-1 virulence in relation to viral population diversity.

In this study, we also observed a steady increase in genetic variability (i.e. mixed bases) at synonymous sites over time, as measured by the SM-Index, in both CSF and blood plasma. In contrast, diversity at nonsynonymous sites (i.e. amino acid residues), as measured the RM-Index, in blood plasma steadily increased until about year 12 of duration of infection and then steadily declined. The observed rise and decline in diversity at nonsynonymous sites in blood likely results from the diminishing capacity of the immune system or fixation of cytotoxic T-lymphocyte immune mediated escape mutations over time (Allen et al., 2005; Borrow et al., 1997; Friedrich et al., 2004; Price et al., 1997). While this pattern was seen in the blood, diversity at nonsynonymous sites in the CSF continued to steadily increase over time, suggesting that immune and potentially other selection pressures in the CNS differ from those in the blood. Further, these changes in mixed base indices in regards to duration of infection are likely to be more pronounced during chronic infection given the overall changes in viral population diversity over the natural history of HIV infection(Kouyos et al., 2011; Shankarappa et al., 1999b).

To further investigate intra-individual differences between HIV-1 variants circulating in the blood and CSF, we compared measurements of HIV-1 population diversity in the blood and CSF. For most participants, HIV-1 population diversity was higher in CSF than blood plasma, suggesting that HIV-1 in the CNS likely originated from multiple variants and not a consequence of monoclonal amplification. Additionally, pleocytosis, a marker of lymphocyte trafficking from the blood to the CSF, was not associated with changes in viral diversity in the CSF. This lack of difference may be related to the high proportion of cases demonstrating pleocytosis in this cohort (118/174; 68%), and is consistent with the relatively high CSF viral loads observed, which was a condition of eligibility in this study. Since it is not feasible to amplify virus at low levels, these patients were excluded from the study, and may represent a selection bias in the investigation.

In this study, neuropsychological impairment was *negatively* correlated with reported duration of infection, which was unexpected given that loss of CD4+ T lymphocytes is associated with greater duration of infection (O'Brien et al., 1997; Strain et al., 2005), and lower CD4 counts have been associated with neuropsychological impairment (Antinori et al., 2007; Lentz et al., 2009; Tozzi et al., 2007). This may point to another limitation of the study, and is most likely a bias in the study design and cohort, which concerns factors like cross-sectional design, survival bias and differences in antiretroviral therapy use, effectiveness and duration. An obvious example of survival bias for all HIV cohorts is that HIV- infected people who live longer and with less disease would be more likely to participate in study cohorts. This factor is further obscured by differential use of antiretroviral therapy and its effectiveness. A prospective longitudinal study would most likely be needed to provide additional and more conclusive insight, especially in relation to duration of infection, which was only by self-report in the CHARTER cohort.

The present study was also limited by the investigating of only the HIV-1 *pol* coding region. Viral diversity in other coding regions, like *env,* are likely to be important in neuropathogenesis (Dunfee et al., 2007; Harrington et al., 2009; Pillai et al., 2006), and this should be evaluated in future studies. However, investigating *pol* did allow us to evaluate resistance-associated mutations in the context of HIV-1 diversity and disease, but the low prevalence of these mutations in our study population, specifically M184V and K103N, prevented meaningful conclusions. An additional limitation of this study is that we did not screen for dual infections among CHARTER cohort participants included in the analysis of neuropsychological impairment. It is likely that some participants in the CHARTER cohort

were dually infected, which could influence both intra-host viral diversity measures and disease progression(Kouyos et al., 2011; Pacold et al.). Although the mixed base approach is less resource intensive than other methods used to evaluate diversity, it does not allow for the identification or quantification of specific variants and this may be necessary for characterizing dual infection or specific neurovirulent phenotypes. Future studies will be needed to evaluate these issues. Measuring mixed bases in population-based sequences can be used as less expensive and time-consuming method to evaluate HIV-1 population diversity, and is a method especially suited for evaluating intra-host diversity for large cohorts. Using this approach to investigate diversity in a relatively large well-defined cohort, we demonstrate that HIV-1 population diversity is positively associated with disease, specifically AIDS and neuropsychological impairment.

Acknowledgments

Financial support: This work was supported by the Pendleton Foundation and National Institutes of Health grants DA12065, DA034978, MH083552, AI077304, AI69432, MH62512, AI27670, AI38858, AI43638, AI43752, AI047745, NS51132, UCSD Centers for AIDS Research Translational Virology Core AI36214, AI29164, AI47745, AI64086, AI57167 and San Diego Veterans Affairs Healthcare System. S

References

Allen TM, Altfeld M, Geer SC, Kalife ET, Moore C, O'Sullivan KM, Desouza I, Feeney ME, Eldridge RL, Maier EL, Kaufmann DE, Lahaie MP, Reyor L, Tanzi G, Johnston MN, Brander C, Draenert R, Rockstroh JK, Jessen H, Rosenberg ES, Mallal SA, Walker BD. 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. 2005; 79 (21):13239–49. [PubMed: 16227247]

Antinori A, Arendt G, Becker JT, Brew BJ, Byrd DA, Cherner M, Clifford DB, Cinque P, Epstein LG, Goodkin K, Gisslen M, Grant I, Heaton RK, Joseph J, Marder K, Marra CM, McArthur JC, Nunn M, Price RW, Pulliam L, Robertson KR, Sacktor N, Valcour V, Wojna VE. Updated research nosology for HIV-associated neurocognitive disorders. Neurology. 2007; 69 (18):1789–99. [PubMed: 17914061]

Borrow P, Lewicki H, Wei X, Horwitz MS, Peffer N, Meyers H, Nelson JA, Gairin JE, Hahn BH, Oldstone MB, Shaw GM. Antiviral pressure exerted by HIV-1-specific cytotoxic T lymphocytes (CTLs) during primary infection demonstrated by rapid selection of CTL escape virus. Nature Medicine. 1997; 3(2):205–11.

Bushell M, Sarnow P. Hijacking the translation apparatus by RNA viruses. Journal of Cell Biology. 2002; 158(3):395–399. [PubMed: 12163463]

Bushman FD, Hoffmann C, Ronen K, Malani N, Minkah N, Rose HM, Tebas P, Wang GP. Massively parallel pyrosequencing in HIV research. AIDS. 2008; 22 (12):1411–5. [PubMed: 18614863]

Butler DM, Pacold ME, Jordan PS, Richman DD, Smith DM. The efficiency of single genome amplification and sequencing is improved by quantitation and use of a bioinformatics tool. Journal of Virological Methods. 2009; 162 (1–2):280–3. [PubMed: 19698751]

Cornelissen M, Jurriaans S, Kozaczynska K, Prins JM, Hamidjaja RA, Zorgdrager F, Bakker M, Back N, van der Kuyl AC. Routine HIV-1 genotyping as a tool to identify dual infections. AIDS. 2007; 21 (7):807–11. [PubMed: 17415035]

Delwart EL, Pan H, Sheppard HW, Wolpert D, Neumann AU, Korber B, Mullins JI. Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. J Virol. 1997; 71(10):7498–508. [PubMed: 9311829]

Domingo E, Menendez-Arias L, Holland JJ. RNA virus fitness. Rev Med Virol. 1997; 7(2):87–96. [PubMed: 10398474]

Dunfee RL, Thomas ER, Wang J, Kunstman K, Wolinsky SM, Gabuzda D. Loss of the N-linked glycosylation site at position 386 in the HIV envelope V4 region enhances macrophage tropism and is associated with dementia. Virology. 2007

Friedrich TC, Dodds EJ, Yant LJ, Vojnov L, Rudersdorf R, Cullen C, Evans DT, Desrosiers RC, Mothe BR, Sidney J, Sette A, Kunstman K, Wolinsky S, Piatak M, Lifson J, Hughes AL, Wilson N, O'Connor DH, Watkins DI. Reversion of CTL escape-variant immunodeficiency viruses in vivo. Nature Medicine. 2004; 10 (3):275–81.

Ganeshan S, Dickover RE, Korber BT, Bryson YJ, Wolinsky SM. Human immunodeficiency virus type 1 genetic evolution in children with different rates of development of disease. J Virol. 1997; 71 (1):663–77. [PubMed: 8985398]

Harrington PR, Schnell G, Letendre SL, Ritola K, Robertson K, Hall C, Burch CL, Jabara CB, Moore DT, Ellis RJ, Price RW, Swanstrom R. Cross-sectional characterization of HIV-1 env compartmentalization in cerebrospinal fluid over the full disease course. AIDS. 2009; 23 (8):907–15. [PubMed: 19414991]

Heaton RK, Clifford DB, Franklin DR Jr, Woods SP, Ake C, Vaida F, Ellis RJ, Letendre SL, Marcotte TD, Atkinson JH, Rivera-Mindt M, Vigil OR, Taylor MJ, Collier AC, Marra CM, Gelman BB, McArthur JC, Morgello S, Simpson DM, McCutchan JA, Abramson I, Gamst A, Fennema-Notestine C, Jernigan TL, Wong J, Grant I. HIV-associated neurocognitive disorders persist in the era of potent antiretroviral therapy: CHARTER Study. Neurology. 2010; 75 (23):2087–96. [PubMed: 21135382]

Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping LH, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A. 2008; 105(21):7552–7. [PubMed: 18490657]

Kosakovsky Pond SL, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21(5):676–679. [PubMed: 15509596]

Kouyos RD, von Wyl V, Yerly S, Boni J, Rieder P, Joos B, Taffe P, Shah C, Burgisser P, Klimkait T, Weber R, Hirschel B, Cavassini M, Rauch A, Battegay M, Vernazza PL, Bernasconi E, Ledergerber B, Bonhoeffer S, Gunthard HF. 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(4):532–9. [PubMed: 21220770]

Lentz MR, Kim WK, Lee V, Bazner S, Halpern EF, Venna N, Williams K, Rosenberg ES, Gonzalez RG. Changes in MRS neuronal markers and T cell phenotypes observed during early HIV infection. Neurology. 2009; 72 (17):1465–72. [PubMed: 19398702]

Markham RB, Wang WC, Weisstein AE, Wang Z, Munoz A, Templeton A, Margolick J, Vlahov D, Quinn T, Farzadegan H, Yu XF. Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. Proc Natl Acad Sci US A. 1998; 95 (21):12568–73.

Morris SR, Little SJ, Cunningham T, MAOM, Garfein RS, Richman DD, Smith DM. Evaluation of an HIV Nucleic Acid Testing Program with Automated Internet and Voicemail Systems to Deliver Results. Annals of Internal Medicine. 2010; 152:778–785. [PubMed: 20547906]

O'Brien WA, Hartigan PM, Daar ES, Simberkoff MS, Hamilton JD. Changes in plasma HIV RNA levels and CD4+ lymphocyte counts predict both response to antiretroviral therapy and therapeutic failure. VA Cooperative Study Group on AIDS. Annals of Internal Medicine. 1997; 126(12):939–45. [PubMed: 9182470]

Pacold ME, Pond SL, Wagner GA, Delport W, Bourque DL, Richman DD, Little SJ, Smith DM. Clinical, virologic, and immunologic correlates of HIV-1 intraclade B dual infection among men who have sex with men. AIDS. 26(2):157–65. [PubMed: 22045341]

Pfeiffer JK, Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. PLoS Pathog. 2005; 1 (2):e11. [PubMed: 16220146]

Pfeiffer JK, Kirkegaard K. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. Proc Natl Acad Sci U S A. 2006; 103(14):5520–5. [PubMed: 16567621]

Pillai SK, Pond SL, Liu Y, Good BM, Strain MC, Ellis RJ, Letendre S, Smith DM, Gunthard HF, Grant I, Marcotte TD, McCutchan JA, Richman DD, Wong JK. Genetic attributes of cerebrospinal fluid-derived HIV-1 env. Brain. 2006; 129(Pt 7):1872–83. [PubMed: 16735456]

Poon AF, Kosakovsky Pond SL, Bennett P, Richman DD, Leigh Brown AJ, Frost SD. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. PLoS Pathog. 2007; 3 (3):e45. [PubMed: 17397261]

Price DA, Goulder PJ, Klenerman P, Sewell AK, Easterbrook PJ, Troop M, Bangham CR, Phillips RE. Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc Natl Acad Sci U S A. 1997; 94 (5):1890–5. [PubMed: 9050875]

Ross HA, Rodrigo AG. Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. J Virol. 2002; 76(22):11715–20. [PubMed: 12388731]

Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol. 1999a; 73 (12): 10489–502. [PubMed: 10559367]

Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection 1. Journal of Virology. 1999b; 73 (12):10489–10502. [PubMed: 10559367]

Smith DM, Zarate S, Shao H, Pillai SK, Letendre SL, Wong JK, Richman DD, Frost SD, Ellis RJ. Pleocytosis is associated with disruption of HIV compartmentalization between blood and cerebral spinal fluid viral populations. Virology. 2009; 385(1):204–8. [PubMed: 19100592]

Strain MC, Letendre S, Pillai SK, Russell T, Ignacio CC, Gunthard HF, Good B, Smith DM, Wolinsky SM, Furtado M, Marquie-Beck J, Durelle J, Grant I, Richman DD, Marcotte T, McCutchan JA, Ellis RJ, Wong JK. Genetic composition of human immunodeficiency virus type 1 in cerebrospinal fluid and blood without treatment and during failing antiretroviral therapy. J Virol. 2005; 79(3):1772–88. [PubMed: 15650202]

Strunnikova N, Ray SC, Lancioni C, Nguyen M, Viscidi RP. Evolution of human immunodeficiency virus type 1 in relation to disease progression in children. Journal of Human Virology. 1998; 1 (3): 224–39. [PubMed: 10195246]

Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial-Dna in Humans and Chimpanzees. Molecular Biology and Evolution. 1993; 10(3): 512–526. [PubMed: 8336541]

Tozzi V, Balestra P, Bellagamba R, Corpolongo A, Salvatori MF, Visco-Comandini U, Vlassi C, Giulianelli M, Galgani S, Antinori A, Narciso P. Persistence of neuropsychologic deficits despite long-term highly active antiretroviral therapy in patients with HIV-related neurocognitive impairment: prevalence and risk factors. J Acquir Immune Defic Syndr. 2007; 45 (2):174–82. [PubMed: 17356465]

Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature. 2006; 439(7074):344–8. [PubMed: 16327776]

Wolinsky SM, Korber BT, Neumann AU, Daniels M, Kunstman KJ, Whetsell AJ, Furtado MR, Cao Y, Ho DD, Safrit JT. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science. 1996a; 272 (5261):537–42. [PubMed: 8614801]

Wolinsky SM, Korber BTM, Neumann AU, Daniels M, Kunstman KJ, Whetsell AJ, Furtado MR, Cao YZ, Ho DD, Safrit JT, Koup RA. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection 1. Science. 1996b; 272 (5261):537–542. [PubMed: 8614801]

Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, Daniels M, Ferrari G, Haynes BF, McMichael A, Shaw GM, Hahn BH, Korber B, Seoighe C. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. PLoS Pathog. 2009; 5 (5):e1000414. [PubMed: 19424423]

Woods SP, Rippeth JD, Frol AB, Levy JK, Ryan E, Soukup VM, Hinkin CH, Lazzaretto D, Cherner M, Marcotte TD, Gelman BB, Morgello S, Singer EJ, Grant I, Heaton RK. Interrater reliability of clinical ratings and neurocognitive diagnoses in HIV. Journal of Clinical and Experimental Neuropsychology. 2004; 26 (6):759–778. [PubMed: 15370374]

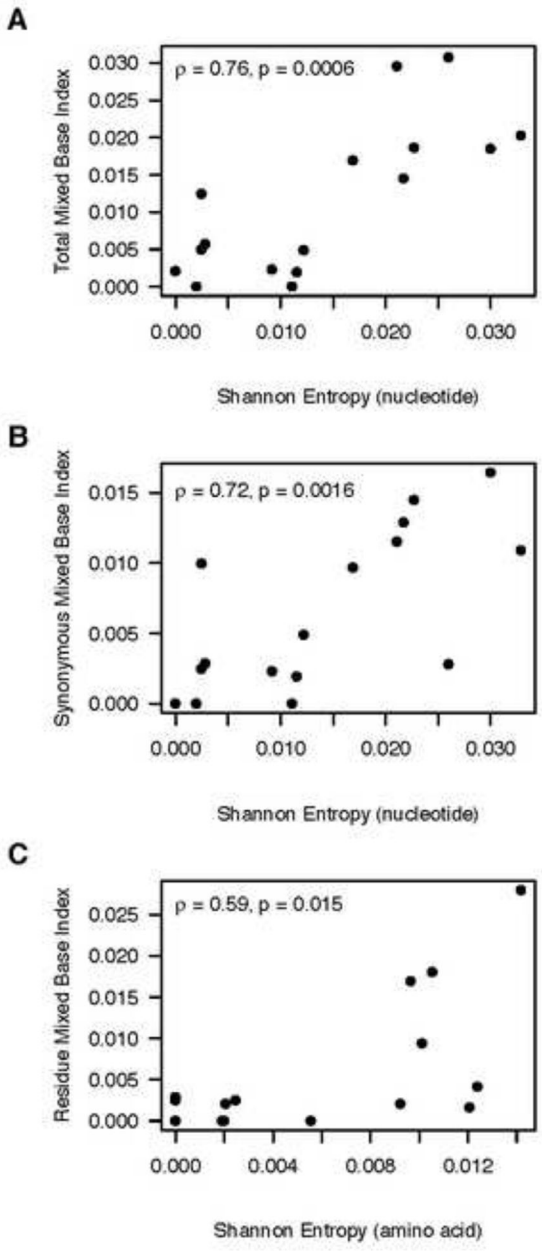

Figure 1. Comparing Shannon Entropy to mixed base indices

Legend: For participants with both single genome and population-based sequences, we measured and compared Shannon Entropy to mixed base indices structured to estimate nucleotide and amino acid diversity. Shannon Entropy of nucleotide sequences correlated with the (A) Total Mixed Base Index (rho = 0.76, p = 0.001) and (B) Synonymous Mixed Base Index (rho = 0.72, p = 0.002) and (C) Residue Mixed Index (rho = 0.59, p=0.015)

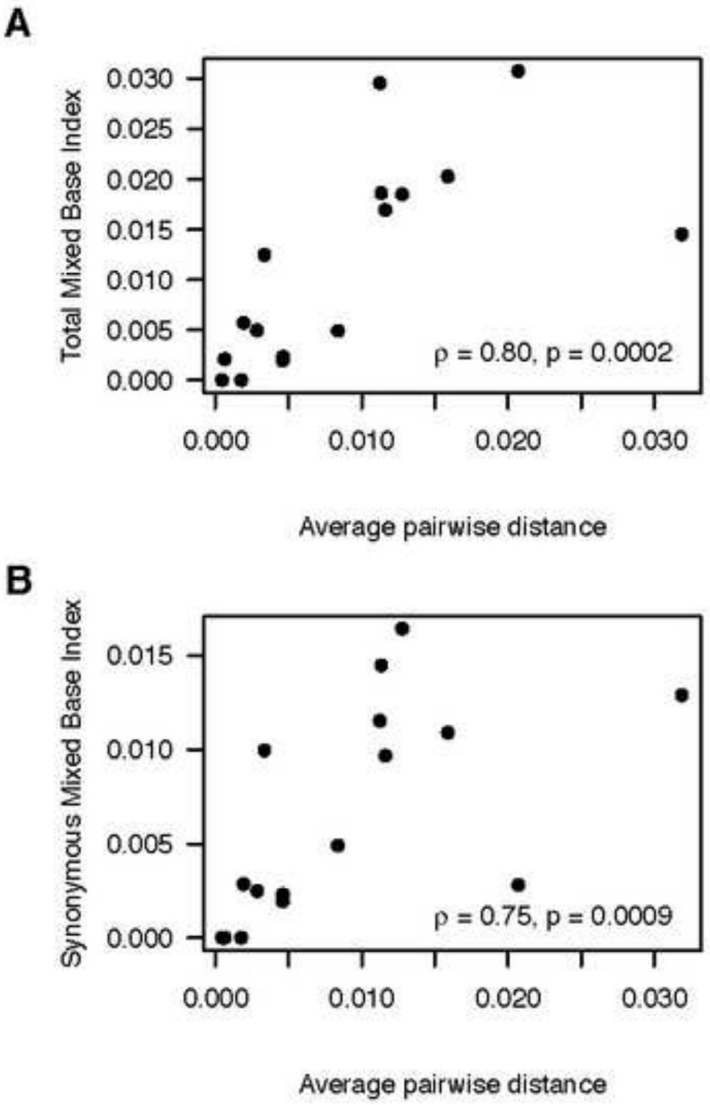

Figure 2. Comparing average pairwise distance (APD) to mixed base indices

Legend: For participants with both single genome and population-based sequences, we measured and compared APD to mixed base indices structured to estimate nucleotide diversity. APD correlated with the (A) Total Mixed Base Index (rho=0.80, p=0.002) and (B) Synonymous Mixed Base Index (rho=0.75, p<0.001).

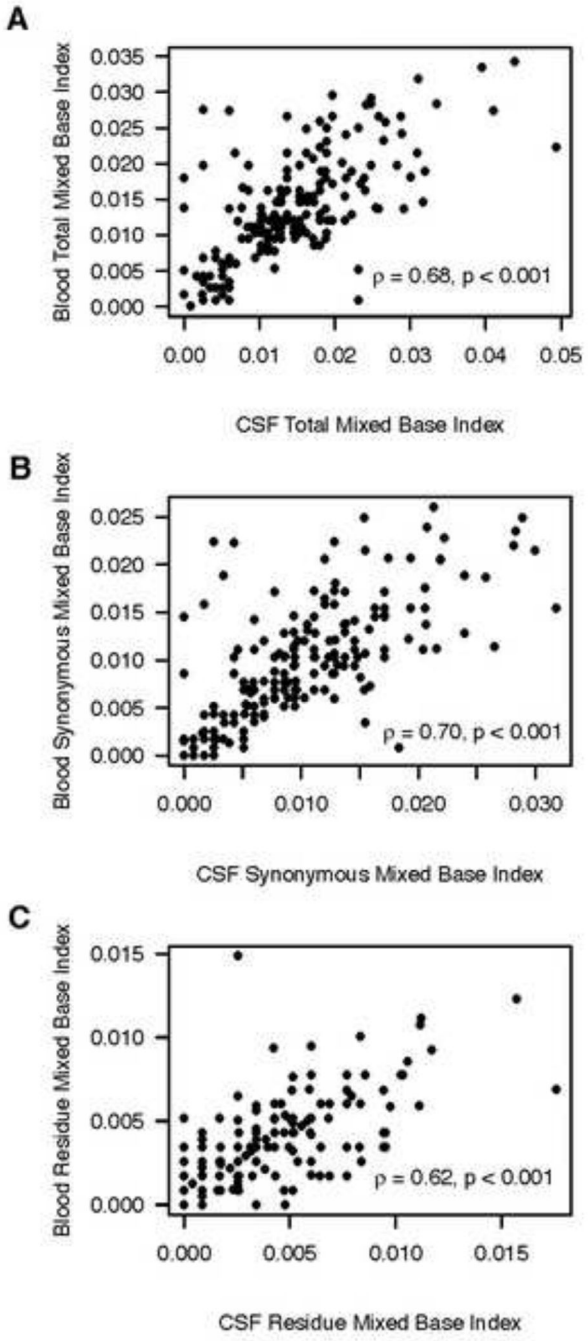

Figure 3. HIV-1 population diversity in blood and CSF

Legend: HIV-1 population diversity in blood correlated with diversity in CSF as measured by the (A) Total Mixed Base Index (spearman rho=0.68, p<0.001), (B) Synonymous Mixed Base Index (spearman rho=0.70, p<0.001) and (C) Residue Mixed Index (spearman rho=0.62, p<0.001).

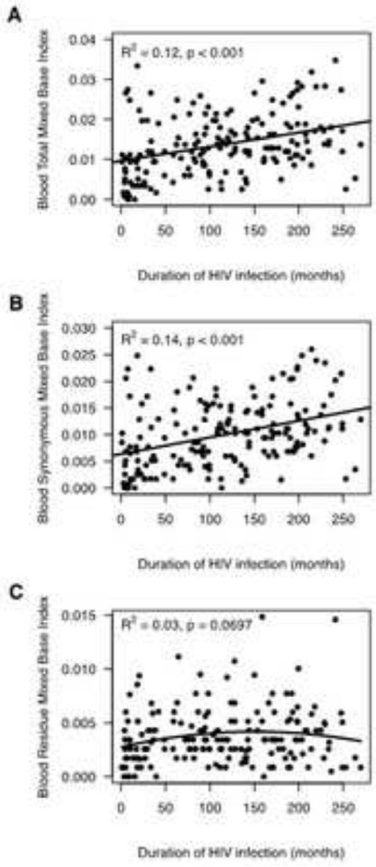

Figure 4. HIV-1 population diversity in blood with relation to time

Legend: Duration of infection had linear relationships with viral population diversity in blood as measured by the (A) Total Mixed Base Index ($p < 0.001$) and (B) Synonymous Mixed base Index ($p < 0.001$), but the trend was quadratic for the (C) Residue Mixed Index ($p = 0.07$).

Table 1

CHARTER participant demographics and clinical characteristics

Variable (N = 187)	Median (IQR) or %
Male	80%
Black	49%
Hispanic	8%
White	41%
Age (years)	42 (37 – 47)
Estimated Duration of infection (months)	113 (39 – 173)
HIV RNA CSF (log_{10}copies/ml)	3.56 (3.14 – 4.05)
HIV RNA Blood (log_{10}copies/ml)	4.55 (4.11 – 4.88)
Current CD4 (cells/ul)	362 (224 – 532)
Nadir CD4 (cells/ul)	251 (141 – 378)
AIDS	43%
Blood Total Mixed Base Index	0.013 (0.009 – 0.018)
Blood Synonymous Mixed Base Index	0.010 (0.005 – 0.014)
Blood Residue Mixed Base Index	0.003 (0.002 – 0.005)
CSF Total Mixed Base Index	0.014 (0.008 – 0.019)
CSF Synonymous Mixed Base Index	0.009 (0.005 – 0.014)
CSF Residue Mixed Base Index	0.003 (0.002 – 0.005)

Legend: Values are reported as percentages or medians with interquartile (IQR) ranges in parentheses.

Table 2

Multivariable regressions modeling AIDS

Model A: AIDS as a function of Blood Total Mixed Base Index, infection duration, plasma VL and ART history				
	Estimate	**Std Error**	**Adjusted OR**	**p-value**
Blood Total Mixed Base Index	0.666	0.275	1.95	**0.0154**
Infection Duration (years)	0.103	0.036	1.11	**0.0039**
Blood Plasma VL (log_{10})	0.669	0.267	1.95	**0.0121**
ART History (Current Use)	3.046	0.604	21.0	**<0.0001**
ART History (Past Use)	0.901	0.467	2.46	0.0539
Model B: AIDS as a function of Plasma Residue Mixed Base Index, infection duration, plasma VL and ART history				
	Estimate	Std Error	Adjusted OR	**p-value**
Blood Residue Mixed base Index	1.865	0.788	6.46	**0.0179**
Infection Duration (years)	0.123	0.035	1.13	**0.0004**
Blood Plasma VL (log_{10})	0.685	0.268	1.98	**0.0107**
ART History (Current Use)	2.886	0.594	17.9	**<0.0001**
ART History (Past Use)	0.765	0.465	2.15	0.0998

Legend: Variables that are independently significant at $p<0.05$ are in bold.

Table 3

Multivariable regressions modeling neuropsychological performance

Model A: Neurocognitive impairment as a function of Blood Total Mixed Base Index, infection duration, AIDS, current CD4, comorbidity status and ART History				
	Estimate	**Std Error**	**Adjusted OR**	**p-value**
Blood Total Mixed Base Index	0.649	0.256	1.9	**0.013**
Infection Duration (years)	−0.105	0.037	0.90	**0.004**
Current CD4 count	−0.015	0.01	0.99[a]	0.140
AIDS	−0.758	0.493	0.47	0.123
Comorbidity (Confounding)	2.36	0.541	10.6	**< 0.001**
Comorbidity (Contributing)	1.01	0.406	2.8	**0.013**
ART History (Current Use)	1.06	0.569	2.9	0.062
ART History (Past Use)	0.066	0.437	1.1	0.879
Model B: Neurocognitive impairment as a function of Blood Residue Mixed Base Index, infection duration, AIDS, current CD4, comorbidity status and ART History				
	Estimate	Std Error	Adjusted OR	p-value
Blood Residue Mixed Base Index	2.43	0.759	11.4	**0.001**
Infection Duration (years)	−0.088	0.035	0.92	**0.013**
Current CD4 count	−0.015	0.010	0.98[a]	0.136
AIDS	−0.891	0.507	0.41	0.079
Comorbidity (Confounding)	2.46	0.551	11.7	**< 0.001**
Comorbidity (Contributing)	1.08	0.413	2.9	**0.009**
ART History (Current Use)	1.03	0.585	2.8	0.077
ART History (Past Use)	0.037	0.452	1.0	0.935

Legend:

[a]Odds ratio (OR) per 10 units increase in CD4 count.

Variables that are independently significant at $p<0.05$ are in bold.