

UCLA

UCLA Previously Published Works

Title

Short Communication: HIV-1 Gag Genetic Variation in a Single Acutely Infected Participant Defined by High-Resolution Deep Sequencing

Permalink

<https://escholarship.org/uc/item/997771kh>

Journal

AIDS Research and Human Retroviruses, 30(8)

ISSN

0889-2229

Authors

Al-Mawsawi, Laith Q
Wu, Nicholas C
De La Cruz, Justin
et al.

Publication Date

2014-08-01

DOI

10.1089/aid.2014.0097

Peer reviewed

Short Communication: HIV-1 Gag Genetic Variation in a Single Acutely Infected Participant Defined by High-Resolution Deep Sequencing

Laith Q. Al-Mawsawi,^{1,2,*} Nicholas C. Wu,^{1,3,*} Justin De La Cruz,⁴ Vivian Cai Shi,¹ Ting-Ting Wu,^{1,2}
Eric S. Daar,⁵ Martha J. Lewis,^{2,6} Otto O. Yang,^{2,4,6,7} and Ren Sun¹⁻³

Abstract

Acute HIV-1 infection is characterized by the rapid generation of highly diverse genetic variants to adapt to the new host environment. Understanding the dynamics of viral genetic variation at this stage of infection is critical for vaccine design efforts and early drug treatment. Here, using a high-resolution deep sequencing approach targeting the HIV-1 *gag* region, we reveal very early immune pressure with dramatic subpopulation shifts in a single acutely infected participant providing further insight into the genetic dynamics of acute HIV-1 infection.

Introduction

ACUTE HIV-1 INFECTION represents a critical period where intervention with antiretroviral drugs or vaccine-generated immunity could control viral replication and disease progression in infected patients. In general, initial replication is localized to the mucosal epithelium and associated lymph nodes. After ≥ 10 days postinfection (dpi), the virus spreads via the blood leading to systemic dissemination and rapid genetic diversification, with exponential increases in plasma viremia typically peaking at 21–28 dpi.¹ Peak viremia is followed by a gradual decrease over a period of months to a quasistable “set-point” that is predictive of the rate of disease progression,^{2,3} and influenced by patient-specific factors, the best defined of which is human leukocyte antigen (HLA) genotype.^{4–7}

Revealing the dynamic genetic adaptations of HIV-1 under immune selection pressure in the new host environment is essential for an improved understanding of viral pathogenesis and for efforts to develop a universal vaccine. Multiple studies provide support that a single founder HIV-1 virus per patient is responsible for establishing the majority of productive infections.^{8–11} In the weeks following viral transmission the founder genotype undergoes an unprecedented

expansion in the mutational space enabled through the replicative error rate of the viral reverse transcriptase (RT) enzyme at $\sim 10^{-3,6}$, a rapid generation output of $\sim 10^{10}$ virions per patient per day, and the propensity of RT to mediate RNA recombination via template switching.^{12–17} Host CTL responses targeting HIV-1 epitopes are the major immune selection force driving the evolution of early viral sequences (other than *env*) during acute infection.^{18–22} Here we use a deep sequencing approach to define the immunodominant HIV-1 *gag* gene at serial time points around peak viremia and provide high-resolution genetic detail into the evolutionary dynamics of viral adaptation in a single case of acute HIV-1 infection.

Participant Information

Samples from an HIV-1-infected person enrolled in the NIH-funded Acute Infection Early Disease Research Program (AIEDRP) were obtained under University of California, Los Angeles, Institutional Review Board exempt protocol. The HLA genotype, viremia level, and percent CD4⁺ and CD8⁺ cell count measurements from the first six sample collection visits are shown in Fig. 1. Viremia at the first two visits (27 and 36 days post-symptoms onset) remained relatively stable

¹Department of Molecular and Medical Pharmacology, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, California.

²AIDS Institute, University of California, Los Angeles, Los Angeles, California.

³Molecular Biology Institute, University of California, Los Angeles, Los Angeles, California.

⁴Department of Microbiology, Immunology, and Medical Genetics, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, California.

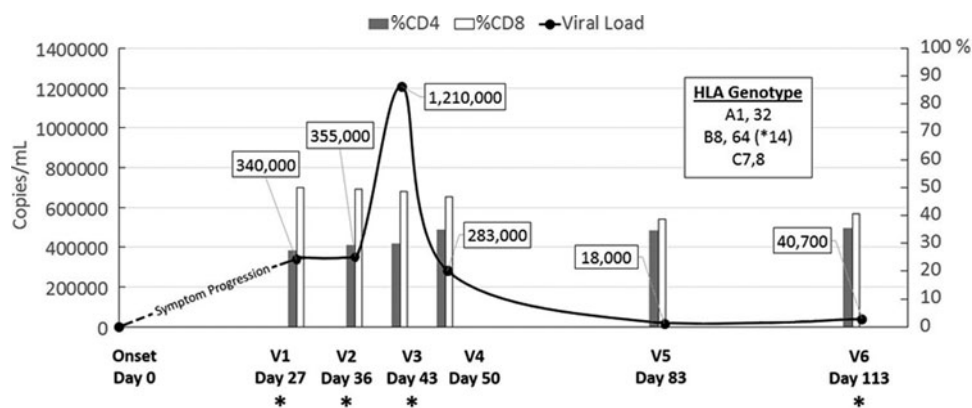
⁵Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Los Angeles, California.

⁶Division of Infectious Diseases, Department of Medicine, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, California.

⁷AIDS Healthcare Foundation, Los Angeles, California.

*These authors contributed equally to this work.

FIG. 1. The acute infection profile of a study participant. The smooth curve depicts changes in viremia (viral RNA copies/ml plasma) whereas the bar chart shows CD4% (gray) and CD8% (white) cell count per ml of peripheral blood mononuclear cells (PBMC) at each visit sample collection time point. The HLA genotype of the study subject is provided as a boxed inset. Visit samples chosen for deep sequencing are denoted by an asterisk.



at $\sim 3.5 \times 10^5$ viral RNA copies/ml plasma. At visit 3 (V3—day 43) the participant had the highest viremia detected at 1.21×10^6 viral RNA copies/ml plasma, which subsequently dropped over 4-fold to 2.83×10^5 a week later by V4 (day 50). At 83 days post-symptoms onset (V5) viremia decreased over 15-fold to its lowest point during the study duration at 1.8×10^4 viral RNA copies/ml plasma, and then rebounded slightly over 2-fold to 4.07×10^4 by V6 (day 113). The percentages of CD4⁺ and CD8⁺ T lymphocytes remained relatively constant over the duration, at a median CD4% of 32.3% (range 27.4 to 35.5%) and a median CD8% of 47.8% (range 38.8 to 50.1%). Plasma samples from V4–V5 were excluded from our analysis due to availability. V1–V3 and V6 were evaluated by deep sequencing.

Sequencing Approach

A 1 ml plasma sample from each visit was used to extract HIV-1 RNA, and the total RNA from this volume was used to generate cDNA using Superscript III Reverse Transcriptase (Invitrogen) using random hexamers. Total viral cDNA was then used as a template to amplify 12 short (~ 188 nucleotides) staggered polymerase chain reaction (PCR) fragments covering the entire *gag*. All PCR reactions used high-fidelity KOD DNA polymerase (EMD Millipore). Each primer pair contained 12 randomized bases as a unique nucleotide tag (4,194,304 possible sequences) to identify each distinct HIV-1 sequence, and a constant sequence at each terminus corresponding to a portion of the 5' and 3' adapter region specific for the Illumina sequencing platform. PCR products from each visit were electrophoresed on a 3% agarose gel to confirm product amplification; aliquots of each were then pooled by visit and column purified. The final pooled sample was measured for concentration and precisely diluted before being used as a template for a second PCR reaction at sub-saturation (~ 20 cycles) with a single primer set to add the remaining sequence of the Illumina adapter region. The PCR products were then submitted for 2×100 bp paired-end sequencing using the Illumina HiSeq 2000 machine utilizing 60% of DNA space on a single sequencing lane and totaling ~ 108 million reads. The sequencing preparation scheme is shown in Fig. 2A and all primer sequences are provided in Supplementary Table S1 (Supplementary Data are available online at www.liebertpub.com/aid). The dilution step between PCR reactions was required to effectively identify Illumina HiSeq 2000 instrument errors by generating a median of 10 copies of each uniquely tagged *gag* amplicon after

the second PCR, where after clustering the sequence output by unique tag, instrument errors (rare within a given read cluster) versus natural mutations (frequent within a given read cluster) could be distinguished, an approach conceptually similar to a method previously described.²³

All Illumina HiSeq 2000 filtered *gag* sequencing reads were mapped to the reference consensus sequence derived from clonal sequencing of V1 samples. Mapping was performed using the Burrows-Wheeler Alignment tool with 10% of mismatches allowed.²⁴ Sequencing reads were clustered by unique tag. Clusters containing three or less reads were removed and only mismatches that had an occurrence of $> 95\%$ within a cluster were considered true mutational variants. These criteria provided a high statistical confidence with a p -value $\leq 10^{-9}$ (binomial exact test, $p = 0.001$) for individual mutation calling. Custom Python scripts were used to conflate error-corrected sequences. The sensitivity of our sequencing approach to detect rare variants was ultimately limited by the error rate of reverse transcription at $\sim 0.0034\%$ ($\sim 1/29,400$) at the cDNA generation step of the process. Sensitivity above this threshold is dependent on amplicon coverage and varied by each amplicon and visit sample. Shown in Fig. 2B is a histogram of the amplicon coverage per visit sample, whereas actual amplicon counts, sensitivity, and nucleotide region covered are given in Supplementary Table S2. Amplicons with coverage of $< 5,000$ reads, corresponding to a sensitivity of $\geq 0.02\%$ ($1/5,000$), were excluded from analysis (boxed in Fig. 2B). The *gag* regions representing amplicons 3 and 7 consistently did not meet our coverage cutoff for analysis as PCR product was repeatedly low presumably due to poor annealing of our designed primers. Raw sequencing data were deposited to the NCBI Sequence Read Archive (SRA) under the accession code BioProject PRJNA244693.

Quasispecies Variation

Our high-resolution sequencing approach enabled us to track the fluctuation of extremely rare *gag* mutations not observed in previous studies within the HIV-1 quasispecies population during the course of acute infection. Shown in Fig. 3A is a heat map depicting the frequencies of individual point mutations with a maximum frequency of $> 1\%$ across all four time points. A hierarchically clustering analysis was applied to group mutations with a similar frequency trend. We observed two distinct occurrence frequency patterns that were reciprocal. In Pattern 1 mutations had a relatively low occurrence frequency in V1 and V3, but had a relatively high occurrence

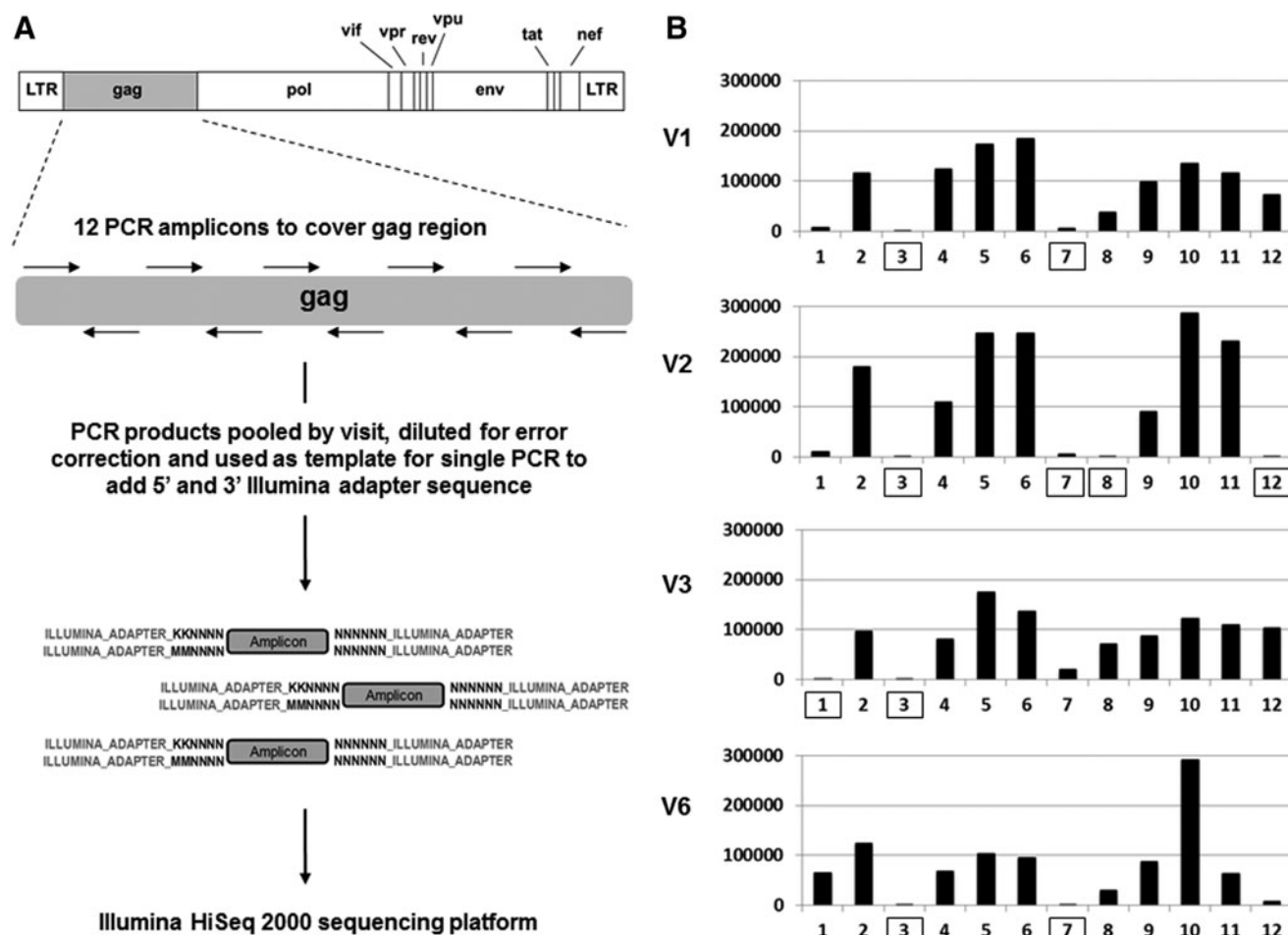


FIG. 2. Deep sequencing. (A) Schematic of sample preparation for the deep sequencing approach. HIV-1 gag cDNA was used as a template for the first step PCR using 12 primer pairs containing overhangs with unique nucleotide tag and partial overhangs with the Illumina adapter. PCR products were pooled by visit, diluted for error correction, and used in a single PCR reaction to add the remaining Illumina adapter sequence before deep sequencing. (B) Sequence coverage of each amplicon covering the HIV-1 gag region per visit. Box denotes amplicons where coverage was <5,000 error-corrected reads and not used in further analysis.

frequency in V2 and V6. Pattern 2 displayed the reciprocal trend: mutations that had a relatively low occurrence frequency in V2 and V6 showed a relatively high frequency in V1 and V3. This reciprocal fluctuation suggested there are two major subpopulations, one dominating under replication conditions present in V1 and V3, and the other dominating under the conditions present in V2 and V6. This was further supported by a pairwise correlation analysis of the genetic content across all four time points (Fig. 3B). The occurrence frequency correlation of point mutations between V1 and V3 (0.96) was higher than that of V1 and V2 (0.91) and V1 and V6 (0.60). Additionally, the occurrence frequency correlation of point mutations between V6 and V3 (0.68) was higher than that of V6 and V1 (0.60) and V6 and V2 (0.65). Together, this analysis suggests that the gag genetic content was derived from two major HIV-1 subpopulations alternating in replicative dominance during acute infection.

Sequences of Targeted CD8⁺ T Lymphocyte Epitope

By standard gamma interferon (IFN- γ) enzyme-linked immunosorbent spot assays (ELISpot) using overlapping

15-mer HIV-1 subtype B consensus sequence peptides (NIH AIDS Research and Reference Reagent Program) as previously described,^{25–27} this individual had a CD8⁺ T lymphocyte response against Gag residues 117–131 (Fig. 4A). Notably, this peptide contained two previously described overlapping epitopes matched to the HLA type of this subject, AQ9 (AADTGNSSQ) and NP10 (NSSQVSQNYP), Gag 119–127 and 124–133, respectively. In agreement with a prior study of escape mutation in these epitopes during acute HIV-1 infection in a study by Jones *et al.*,²⁸ there were numerous substitution mutations in these epitopes. At all time points, the N124S variant, described as an escape mutation by Jones *et al.*, was dominant, comprising 85–92% of the total population, suggesting it was the sequence within the founder virus or that it was an early escape variant that reached fixation before V1. The remaining 8–15% of the total population was composed of variants with other substitutions at position 124, including C, G, I, N, and R. Other variants observed by Jones *et al.* were also observed. S125N, D123E, and H126N (both positions have dominant residues that deviate from the subtype B consensus of G123 and S126) all appeared to fluctuate with viremia levels, observed at 13–14% and 18%

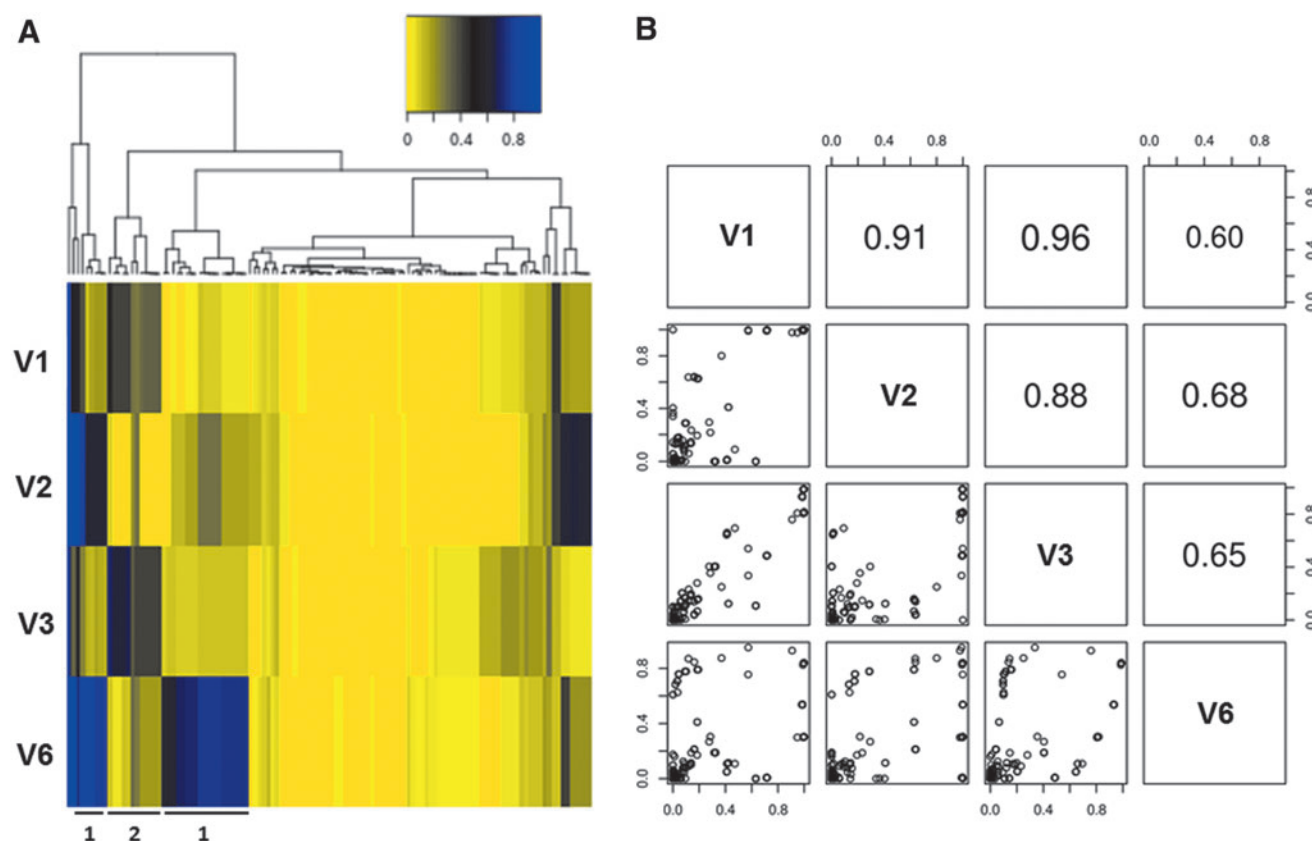


FIG. 3. Quasispecies variation analysis shows two reciprocally dominant subpopulations depending on the time point during acute infection. **(A)** Heat map depicting the occurrence frequency of individual point mutations (>1%) across all four time points hierarchically clustered by frequency at all four visits. The heat map follows the color spectrum from yellow (low frequency) to blue (high frequency). Representative mutation clusters for Patterns 1 and 2 are highlighted. **(B)** Pairwise correlation analysis of the occurrence frequencies of individual point mutations across all four time points further highlighting the two subpopulations reciprocally dominant at V1/V3 versus V2/V6.

at V1–V2, peaking at 18% and 28% in V3, and dropping to 10% and 17% in V6 for D123E and S125N, and H126N, respectively.

The frequency change of mutation group D123E, S125N, and H126N showed statistical significance using a paired *t*-test from V2–V3 and V3–V6, with $p=0.04897$ and 0.009519 , respectively, but no statistical significance for the frequency change from V1–V2 ($p=0.2258$). Rare variants included A119E/T, D121G, S125G, Q127K, and V128A, which were present as minor species (each at $\leq 2.5\%$ of the total population) throughout acute infection. The entire Gag amino acid substitution profile from 119 to 133 for each visit is shown in Fig. 4A. These data suggested ongoing CD8⁺ T lymphocyte-mediated selective pressure maintaining significant subpopulations within the total quasispecies patient pool.

As AQ9/NP10 is contained within a single amplicon (#4), this permitted interrogation of its haplotype linkage. A neighbor-joining tree was constructed using all 20 haplotypes that had a maximum occurrence frequency of >0.1% (Fig. 4B). The 20 haplotypes formed two distinct clusters on the phylogenetic tree, which could be distinguished by the two key variants, D123E and H126N. Viral subpopulations that carried D123E and H126N represented a minor fraction of the entire population, with an accumulating occurrence frequency of 13.5% in V1, 13.6% in V2, 18.1% in V3, and

10.0% in V6. We also examined the frequency distribution of all pairwise Hamming distances for the haplotypes detected in V1 (Fig. 4C), an analysis that provides insight into the origination of distinct viral subpopulations.¹⁰ This revealed a bimodal distribution, which does not conform to the Poisson distribution model predicted of a single ancestral virus. Together, our results suggest that two or more founder viruses were required to generate the distinct phylogenetic viral clusters observed, or a minor subpopulation was derived from a single founder virus by acquiring two substitutions, D123E and H126N, prior to V1 sample collection during the very early phases of acute infection.

Here we utilized a high-resolution deep sequencing approach targeting HIV-1 *gag* to uncover aspects of viral genetic dynamics and diversification during acute infection. Multiple studies have applied the 454 next-generation pyrosequencing approach to investigate viral dynamics and immune escape during HIV-1 acute infection,^{9,20,29} highlighting the importance of minor variants to early immune pressure and dramatic subpopulation shifts also observed here using the Illumina platform. The 454 pyrosequencing approach affords longer sequence reads, but has a higher error rate with lower sequence depth and sensitivity to detect rare variants versus Illumina.³⁰ The shorter sequence reads from Illumina make haplotype reconstruction a major

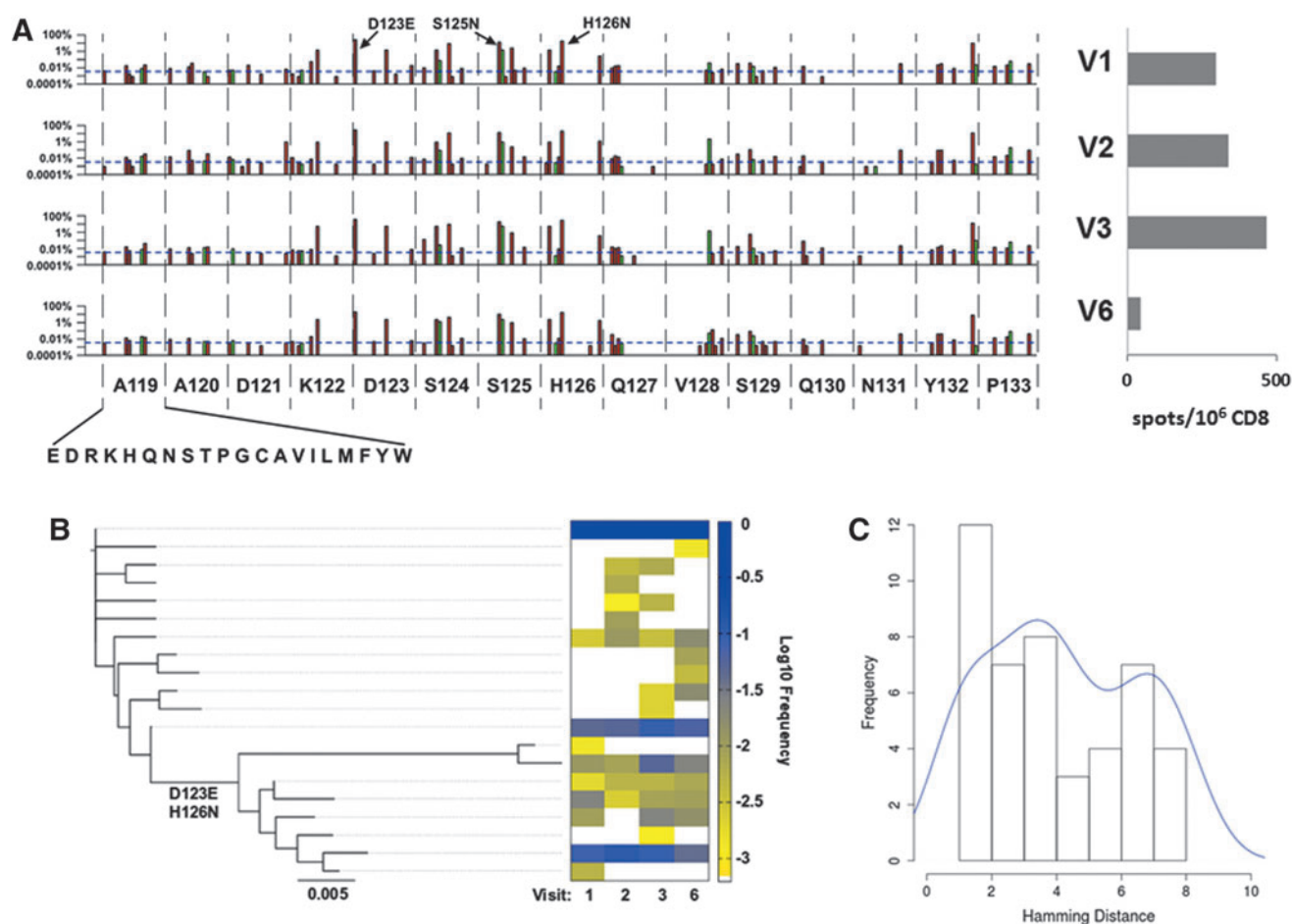


FIG. 4. AQ9/NP10 epitope focus reveals active CTL selection and insight into population-level genetic dynamics. **(A)** High-resolution focus on substitutions observed in the Gag epitope AQ9/NP10 from amino acids 119–133 by visit time point. The y-axis is in log percentage, with the blue horizontal dotted line showing the sensitivity cutoff (0.0034%), and silent versus missense mutations shown in green and red, respectively. Interferon (IFN)- γ ELISpot confirmation of patient CTL targeting AQ9/NP10 using consensus subtype B Gag peptide (117–131), with the number of positive spots/10⁶ CD8 at 298.5, 340, 468, and 43 for V1–V3 and V6, respectively. **(B)** Neighbor-joining tree constructed using AQ9/NP10 containing amplicon 4 using all 20 haplotypes with a maximum occurrence frequency of >0.1% showing two distinct clusters distinguished by the key variants D123E and H126N. The corresponding heat map shows the haplotype frequency at each visit time point at log₁₀ frequency from yellow (low frequency) to blue (high frequency). **(C)** Pairwise Hamming distances for the haplotypes detected in V1 showing a bimodal distribution indicating two or more ancestral viruses are required prior to V1 to generate the distinct phylogenetic viral clusters observed.

challenge—an issue we have recently overcome.³¹ In conclusion, our data suggested very early immune pressure with two quasispecies subpopulations reciprocally shifting in dominance. This deep sequencing approach can be applied to reveal insights about viral adaption to immune or drug pressure, including other viruses that exist as quasispecies swarms *in vivo*.

Acknowledgments

We thank our study subject for their participation. This work was supported by NIH R01 AI043203 (O.O.Y.), the UCLA Center for AIDS Research (CFAR) NIH/NIAID AI028697 (O.O.Y. and R.S.), NIH R21 AI110261 (R.S.), UCLA Jonsson Comprehensive Cancer Center (JCCC) NIH/NCA P30 CA016042 (R.S.), and the California HIV/AIDS Research Program (CHRP) Innovative, Development, Exploratory Award (IDEA) (R.S.).

Author Disclosure Statement

No competing financial interests exist.

References

- McMichael AJ, Borrow P, Tomaras GD, *et al.*: The immune response during acute HIV-1 infection: Clues for vaccine development. *Nat Rev Immunol* 2010;10(1): 11–23.
- Mellors JW, Kingsley LA, Rinaldo CR Jr, *et al.*: Quantitation of HIV-1 RNA in plasma predicts outcome after seroconversion. *Ann Intern Med* 1995;122(8):573–579.
- Mellors JW, Rinaldo CR Jr, Gupta P, *et al.*: Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science* 1996;272(5265):1167–1170.
- Carrington M and O'Brien SJ: The influence of HLA genotype on AIDS. *Annu Rev Med* 2003;54:535–551.

5. Fellay J, Shianna KV, Ge D, *et al.*: A whole-genome association study of major determinants for host control of HIV-1. *Science* 2007;317(5840):944–947.
6. Gao X, Nelson GW, Karacki P, *et al.*: Effect of a single amino acid change in MHC class I molecules on the rate of progression to AIDS. *N Engl J Med* 2001;344(22):1668–1675.
7. Kulpa DA and Collins KL: The emerging role of HLA-C in HIV-1 infection. *Immunology* 2011;134(2):116–122.
8. Abrahams MR, Anderson JA, Giorgi EE, *et al.*: Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-Poisson distribution of transmitted variants. *J Virol* 2009;83(8):3556–3567.
9. Fischer W, Ganusov VV, Giorgi EE, *et al.*: Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One* 2010;5(8):e12303.
10. Keele BF, Giorgi EE, Salazar-Gonzalez JF, *et al.*: Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA* 2008;105(21):7552–7557.
11. Salazar-Gonzalez JF, Salazar MG, Keele BF, *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(6):1273–1289.
12. Ho DD, Neumann AU, Perelson AS, *et al.*: Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature* 1995;373(6510):123–126.
13. Levy DN, Aldrovandi GM, Kutsch O, *et al.*: Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci USA* 2004;101(12):4204–4209.
14. Mansky LM: Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res Hum Retroviruses* 1996;12(4):307–314.
15. Mansky LM and Temin HM: Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol* 1995;69(8):5087–5094.
16. Perelson AS, Neumann AU, Markowitz M, *et al.*: HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996;271(5255):1582–1586.
17. Rhodes T, Wargo H, and Hu WS: High rates of human immunodeficiency virus type 1 recombination: Near-random segregation of markers one kilobase apart in one round of viral replication. *J Virol* 2003;77(20):11193–11200.
18. Borrow P, Lewicki H, Hahn BH, *et al.*: Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection. *J Virol* 1994;68(9):6103–6110.
19. Brumme ZL, Brumme CJ, Carlson J, *et al.*: Marked epitope- and allele-specific differences in rates of mutation in human immunodeficiency type 1 (HIV-1) Gag, Pol, and Nef cytotoxic T-lymphocyte epitopes in acute/early HIV-1 infection. *J Virol* 2008;82(18):9216–9227.
20. Henn MR, Boutwell CL, Charlebois P, *et al.*: Whole genome deep sequencing of HIV-1 reveals the impact of early minor variants upon immune recognition during acute infection. *PLoS Pathog* 2012;8(3):e1002529.
21. Koup RA, Safrit JT, Cao Y, *et al.*: Temporal association of cellular immune responses with the initial control of viremia in primary human immunodeficiency virus type 1 syndrome. *J Virol* 1994;68(7):4650–4655.
22. Wilson JD, Ogg GS, Allen RL, *et al.*: Direct visualization of HIV-1-specific cytotoxic T lymphocytes during primary infection. *AIDS* 2000;14(3):225–233.
23. Kinde I, Wu J, Papadopoulos N, Kinzler KW, *et al.*: Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci USA* 2011;108(23):9530–9535.
24. Li H and Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25(14):1754–1760.
25. Balamurugan A, Lewis MJ, Kitchen CM, *et al.*: Primary human immunodeficiency virus type 1 (HIV-1) infection during HIV-1 Gag vaccination. *J Virol* 2008;82(6):2784–2791.
26. Yang OO, Daar ES, Jamieson BD, *et al.*: Human immunodeficiency virus type 1 clade B superinfection: Evidence for differential immune containment of distinct clade B strains. *J Virol* 2005;79(2):860–868.
27. Yang OO, Daar ES, Ng HL, *et al.*: Increasing CTL targeting of conserved sequences during early HIV-1 infection is correlated to decreasing viremia. *AIDS Res Hum Retroviruses* 2011;27(4):391–398.
28. Jones NA, Wei X, Flower DR, *et al.*: Determinants of human immunodeficiency virus type 1 escape from the primary CD8+ cytotoxic T lymphocyte response. *J Exp Med* 2004;200(10):1243–1256.
29. Hedskog C, Mild M, Jernberg J, *et al.*: Dynamics of HIV-1 quasispecies during antiviral treatment dissected using ultra-deep pyrosequencing. *PLoS One* 2010;5(7):e11345.
30. Li JZ, Chapman B, Charlebois P, *et al.*: Comparison of illumina and 454 deep sequencing in participants failing raltegravir-based antiretroviral therapy. *PLoS One* 2014;9(3):e90485.
31. Wu NC, De La Cruz J, Al-Mawsawi LQ, *et al.*: HIV-1 quasispecies delineation by Tag linkage deep sequencing. *PLoS One* 2014;9(5):e97505.

Address correspondence to:

Otto O. Yang

AIDS Institute

University of California, Los Angeles

Los Angeles, California 90095

E-mail: oyang@mednet.ucla.edu

Ren Sun

Department of Molecular and Medical Pharmacology

David Geffen School of Medicine

University of California, Los Angeles

Los Angeles, California 90094

E-mail: rsun@mednet.ucla.edu