DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY

Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus

Citation

Tully, D. C., C. B. Ogilvie, R. E. Batorsky, D. J. Bean, K. A. Power, M. Ghebremichael, H. E. Bedard, et al. 2016. "Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus." PLoS Pathogens 12 (5): e1005619. doi:10.1371/journal.ppat.1005619. http://dx.doi.org/10.1371/journal.ppat.1005619.

Published version

https://doi.org/10.1371/journal.ppat.1005619

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:27320415

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

PLOS | PATHOGENS

RESEARCH ARTICLE

Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus

Damien C. Tully[1], Colin B. Ogilvie[1], Rebecca E. Batorsky[1], David J. Bean[1], Karen A. Power[1], Musie Ghebremichael[1], Hunter E. Bedard[1], Adrianne D. Gladden[1], Aaron M. Seese[1], Molly A. Amero[1], Kimberly Lane[1], Graham McGrath[2], Suzane B. Bazner[2], Jake Tinsley[3], Niall J. Lennon[4], Matthew R. Henn[4], Zabrina L. Brumme[5], Philip J. Norris[6], Eric S. Rosenberg[2], Kenneth H. Mayer[3], Heiko Jessen[7], Sergei L. Kosakovsky Pond[8], Bruce D. Walker[1,9], Marcus Altfeld[1,10], Jonathan M. Carlson[11], Todd M. Allen[1]*

1 Ragon Institute of MGH, MIT and Harvard, Cambridge, Massachusetts, United States of America, **2** Division of Infectious Disease, Massachusetts General Hospital, Boston, Massachusetts, United States of America, **3** The Fenway Institute, Fenway Health, Boston, Massachusetts, United States of America, **4** Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **5** Faculty of Health Sciences, Simon Fraser University, Vancouver, British Columbia, Canada, **6** Blood Systems Research Institute, San Francisco, California, United States of America, **7** HIV Clinic Praxis, Jessen, Berlin, Germany, **8** Department of Medicine, University of California San Diego, La Jolla, California, United States of America, **9** Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America, **10** Heinrich-Pette-Institut, Hamburg, Germany, **11** Microsoft Research, Redmond, Washington, United States of America

* tallen2@partners.org

CrossMark
click for updates

OPEN ACCESS

Citation: Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et al. (2016) Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. PLoS Pathog 12(5): e1005619. doi:10.1371/journal.ppat.1005619

Editor: Ronald Swanstrom, University of North Carolina at Chapel Hill, UNITED STATES

Received: October 27, 2015

Accepted: April 18, 2016

Published: May 10, 2016

Copyright: © 2016 Tully et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was funded in part with federal funds from the National Institute of Allergy and Infectious Disease through P01 AI074415 (TMA and MA). DCT is supported by an American Foundation for AIDS Research (amfAR) Mathilde Krim Fellowship award (108683-55-RKMT). ZLB is supported by a Scholar Award from the Michael Smith Foundation for Health Research. TMA is supported by the Jane Brock-Wilson Fund. MG is supported by the Harvard

Abstract

Due to the stringent population bottleneck that occurs during sexual HIV-1 transmission, systemic infection is typically established by a limited number of founder viruses. Elucidation of the precise forces influencing the selection of founder viruses may reveal key vulnerabilities that could aid in the development of a vaccine or other clinical interventions. Here, we utilize deep sequencing data and apply a genetic distance-based method to investigate whether the mode of sexual transmission shapes the nascent founder viral genome. Analysis of 74 acute and early HIV-1 infected subjects revealed that 83% of men who have sex with men (MSM) exhibit a single founder virus, levels similar to those previously observed in heterosexual (HSX) transmission. In a metadata analysis of a total of 354 subjects, including HSX, MSM and injecting drug users (IDU), we also observed no significant differences in the frequency of single founder virus infections between HSX and MSM transmissions. However, comparison of HIV-1 envelope sequences revealed that HSX founder viruses exhibited a greater number of codon sites under positive selection, as well as stronger transmission indices possibly reflective of higher fitness variants. Moreover, specific genetic "signatures" within MSM and HSX founder viruses were identified, with single polymorphisms within gp41 enriched among HSX viruses while more complex patterns, including clustered polymorphisms surrounding the CD4 binding site, were enriched in MSM viruses.

University Center for AIDS Research 2P30 AI060354-11 and HIVRAD P01-AI104715. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: During the life of this grant, TMA's spouse was an employee of Bristol Myers Squibb, which has a focus in Virology, specifically treatments for hepatitis B and C and HIV/AIDS. TMA's spouse no longer works for BMS and only retained a small stock interest in the public company. TMA's interests were reviewed and managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. This does not alter our adherence to all PLOS policies on sharing data and materials.

While our findings do not support an influence of the mode of sexual transmission on the number of founder viruses, they do demonstrate that there are marked differences in the selection bottleneck that can significantly shape their genetic composition. This study illustrates the complex dynamics of the transmission bottleneck and reveals that distinct genetic bottleneck processes exist dependent upon the mode of HIV-1 transmission.

Author Summary

While the global spread of HIV-1 has been fueled by sexual transmission the genetic determinants underlying the transmission bottleneck remains poorly understood. Here we characterized founder virus population diversity from next generation sequencing data in a cohort of 74 acute and early HIV-1 infected individuals. We observe that the risk of multi-variant infection in men-who-have-sex-with-men (MSM) is not greater than that observed for heterosexuals (HSX), contrary to reports of higher rates of multiple founder virus infections in higher-risk MSM transmissions. These findings were further supported through a metadata analysis of 354 acute and early HIV-1 subjects. We did, however, observe differences between HSM and MSM founder viruses, including a higher selection barrier in HSX transmission with founder viruses being more cohort consensus-like that may be reflective of increased replicative fitness. We also identified a number of residues within Envelope that behave in a risk-dependent manner and could be key for HIV-1 transmission. These novel insights improve our understanding of the HIV-1 transmission bottleneck and underscore the differential selective pressures that founder viruses within the two major transmission risk groups are subjected to.

Introduction

The global spread of HIV-1 has been fueled predominantly by HSX transmission, with MSM representing a second major risk group [1]. As was first established over two decades ago, HIV-1 undergoes a severe population bottleneck upon transmission with only a limited number of variants from the diverse pool of strains in the source establishing productive infection in the recipient individual [2–6]. While the biological mechanisms underlying this genetic bottleneck remain poorly understood, recent studies support a combination of host factors, including the effective physical barrier of the mucosa [7], availability of target cells [8] and levels of immune activation and genital inflammation that may enhance HIV-1 transmission [9–11]. Recently, in a cohort of transmission pairs it was also demonstrated that factors associated with an increased risk of HSX transmission can mitigate this process and reduce the strength of selection of transmission [12]. The application of single-genome amplification and sequencing (SGA/S) to subjects sampled during acute and early infection has allowed for the inference of the founder virus [13, 14]. In 80% of HSX transmissions a single founder virus is responsible for productive clinical infection [7, 14–18], whereas among MSM the incidence of multi-variant transmission is reported to be higher with up to 40% of infections established by multiple viral variants [14, 19]. These data, coupled with epidemiological data illustrating differential risks of infection based on the route of exposure [20], suggest that the mode of transmission additionally influences the selection of the viral variant(s) establishing systemic infection.

A number of genetic, immunologic and phenotypic signatures of founder viruses, predominantly located within the envelope glycoprotein (Env), have been identified that affect HIV-1

entry [16, 21–28]. In particular HIV-1 clade C founder viruses appear to favor shorter, less glycosylated Envs [7, 21] that are closer to ancestral sequences than their contemporary chronic Env counterparts [29, 30], though this does not appear to be the case in subtype B [31–33]. Korber and colleagues extended these earlier findings and identified a number of Env signature sites enriched upon subtype B transmission [34], many of which may affect Env expression resulting in higher Env incorporation within the budding HIV-1 virion [35]. Aside from these genetic features no dominant phenotypic correlate has been associated with viral transmission other than the preference for CCR5 and CD4+ T cell tropism [4, 14, 25, 36, 37]. Recent studies have shown that transmitted viruses appear to be more resistant to inhibition by interferon-α (IFN-α) than viruses derived from chronic infection [27, 38]. However, it remains unclear whether IFN-α makes a major contribution to the HIV-1 transmission bottleneck as not all studies have observed this property in founder viruses [39, 40].

To date, attributes particular to HIV-1 founder viruses have been identified by comparing these sequences to larger datasets of contemporaneous chronic viruses [23, 25–27], with smaller studies examining epidemiologically linked transmission pairs [7, 21, 41–46]. It is unclear, however, what genetic constraints the mode of sexual transmission imposes on the nascent founder virus. Importantly no studies have directly compared HIV-1 founder viruses specific to MSM versus HSX infections. In this study we used whole-genome amplification and 454 deep sequencing to characterize HIV-1 intra-host genetic diversity in acutely infected subjects and demonstrate, using a genetic distance-based approach, that contrary to previous findings [14, 19], only a single founder virus is detectable in the majority of MSM infections. Moreover comparative analyses of MSM and HSX founder viruses suggests that the mode of transmission imposes differential selection pressures, with HSX viruses experiencing broader, modest selection while MSM viruses exhibit stronger selection but at fewer sites. Furthermore, we identify a number of sites within Env that are differentially enriched upon MSM and/or HSX transmission, suggesting that HIV-1 founder viruses specifically evolve or are selected to overcome the different mucosal barriers imposed by the route of transmission. This study provides important insights into how the mode of transmission shapes the HIV-1 founder virus, as well as into the differing selective pressures between MSM and HSX infection, and may facilitate the design of a more effective HIV-1 vaccine or other therapeutic and prevention strategies.

Results

Application of a Hamming distance approach to discriminate between homogeneous and heterogeneous HIV-1 founder virus infections

To adapt the application of more sensitive and higher-throughput next-generation deep sequencing data, where the shorter sequencing reads makes the inference of haplotype reconstruction difficult, we implemented a Hamming distance-based genetic approach, termed here the average (over read pairs) pairwise hamming distance (APHD). This method calculates the number of mismatches between reads in a sliding window of a defined size to analyze the level of genetic diversity within the viral quasispecies. To validate this approach we applied it to previously published SGA/S-derived *env* founder virus sequences from 127 subjects (Dataset 1) for which the multiplicity of infection had previously been determined [14, 19] (Fig 1A and Supplementary Materials and Methods in S1 Text). A classifier based on a logistic regression clearly segregated the subjects into those previously identified to have exhibited infection by a single virus (n = 98) versus those infected by two or more viruses (n = 29; Fig 1B), with the model correctly classifying 97% of the subjects. Cross-validated prediction errors and area under the receiver operating characteristic curve (AUC) were used to assess model

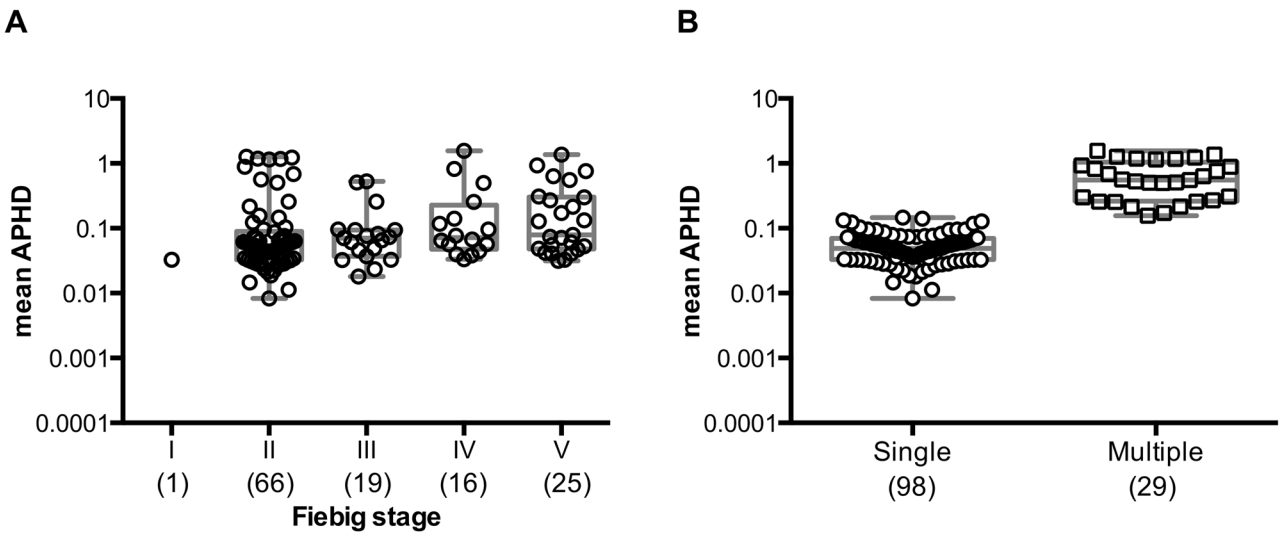

Fig 1. Mean average pairwise Hamming distance (APHD) of HIV-1 Env SGA/S sequences distinguishes between single and multiple founder viruses. (A) A training set of SGA/S Env sequences derived from 127 previously published acute HIV-1 infected subjects illustrating a wide range of *env* diversity. The APHD is calculated using a sliding window of 120bp with a step size of 21bp. The mean APHD is plotted according to Fiebig stages as defined by HIV-1 clinical laboratory test results. **(B)** A classifier based on a logistic regression segregated 127 subjects into single or multiple infections and correctly assigned 97% of subjects into the respective groups. Each point corresponds to an individual subject with the number of subjects denoted on the x-axis in parenthesis under each Fiebig stage.

doi:10.1371/journal.ppat.1005619.g001

performance on data not used to build the model. The prediction error based on 10-fold cross-validated AUC was estimated to be 3.95% and the 10-fold cross-validated AUC was shown to be approximately 0.993 with the corresponding 95% confidence intervals [0.981, 1.00]. For each subject, ART_454 software was used to simulate *in silico* sequencing reads with multiple replicates generated resulting in a dataset of 7.9 million *env* reads. Calculation of the APHD score for each subject from each replicate demonstrated a lack of significant difference between the simulated data and actual data suggesting that modeling sample variation and incorporating the error profiles associated with 454 sequencing has not had an undue effect on the classification of subjects. We conclude, therefore, that the APHD approach is suitable for discriminating between homogeneous and heterogeneous infections, and represents an alternative approach to discriminate between single and multiple founder viruses.

Validation of the APHD approach to deep sequencing data

To validate application of the APHD approach for deep sequencing data we performed both SGA/S and Roche 454 pyrosequencing of *env* (see Methods and S1 Text) in 6 individuals from a range of Fiebig stages (II/III to V). In subject 571373 (Fiebig stage II/III) application of the APHD approach to the 454 sequencing data suggested infection by a single founder virus. A codon-diversity heat map illustrates only minor variation (<10%) present in the viral population (Fig 2A), with a low calculated mean APHD of 0.063 (Fig 2B) placing it inside the 75% percentile of APHD scores obtained for the single founder virus group as calculated above. Comparison of over a dozen SGA/S sequences supported infection by a single founder virus with each lineage containing a unique set of near identical sequences (Fig 2C) in agreement with a star-like phylogeny and Poisson distributed hamming distance that conform to a mathematical model of random evolution (Fig 2D).

In contrast, subject 654207 demonstrated a high level of diversity inconsistent with a single virus transmission (Fig 3A), resulting in a mean APHD score of 0.752 (Fig 3B). SGA/S

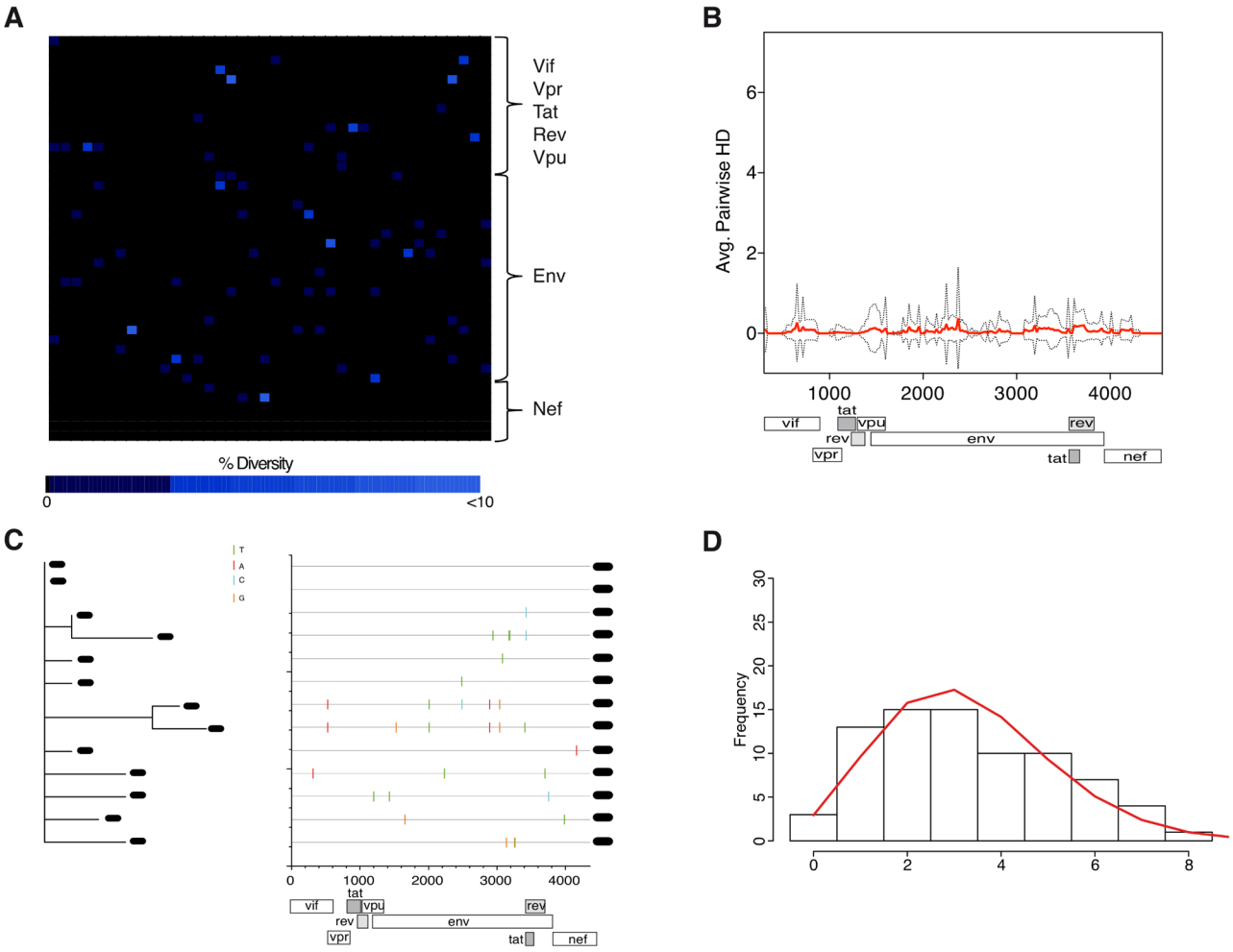

Fig 2. Subject 571373 exhibits low *env* diversity by both 454 and SGA/S reflective of a single founder virus. 454 and SGA/S analysis of 3′ half sequences from subject 571373 (Fiebig stage II/III). **(A)** Heat map illustrating a small number of sites exhibiting low-level amino acid sequence diversity across the 3′ half of the HIV-1 genome as detected by 454 deep sequencing. Plotted is the percentage of amino acid diversity at each position with the first amino acid of Vif located in the top left corner of the grid and last amino acid of Nef located in the bottom right corner. Completely conserved residues are black and low-level variant residues (<10%) are dark blue. **(B)** The average pairwise hamming distance calculated from 454 sequencing reads for the 3′ half of the genome is plotted with the APHD of 0.063 (red line) and standard deviation (dotted black line) shown. The plot shows a relatively uniform population with random sites throughout the genome exhibiting low-level diversity. **(C)** SGA/Ss from subject 571373 covering the 3′ half of the HIV-1 genome display limited structure on a neighbor-joining (NJ) phylogenetic tree (left) and few nucleotide changes from the intrasubject consensus. The highlighter plots (right) compare sequences for each subject's sequence set to an intrasubject consensus (uppermost sequence) and illustrates the pattern of nucleotide base mutations within sequences using short color-coded bars. **(D)** Hamming distance analysis of SGA/Ss showing infection by a single virus with Hamming distance frequencies conforming precisely to model predictions of a single virus infection (red line). As further support for a single founder virus the estimated time to a single most recent common ancestor (MRCA) of 36 days (23–49 days) overlapped with the estimated clinical duration of infection based on Fiebig stages (18–37 days).

doi:10.1371/journal.ppat.1005619.g002

supported infection by at least 4 founder lineages along with extensive interlineage recombination (Fig 3C) while a mathematical model of evolution demonstrated that the SGA/S did not conform to a Poisson distribution (mean Hamming distance per base of 0.005). Notably, however, splitting of variants into their respective sub-lineages did demonstrate conformity to the Poisson distribution and star-like phylogeny resulting in most common recent ancestors (MRCAs) in agreement with clinical estimates (Fig 3D). In next three subjects 882283 (S1 Fig), 702865 (S2 Fig), and 574194 (S3 Fig) the APHD approach again suggested infection by three

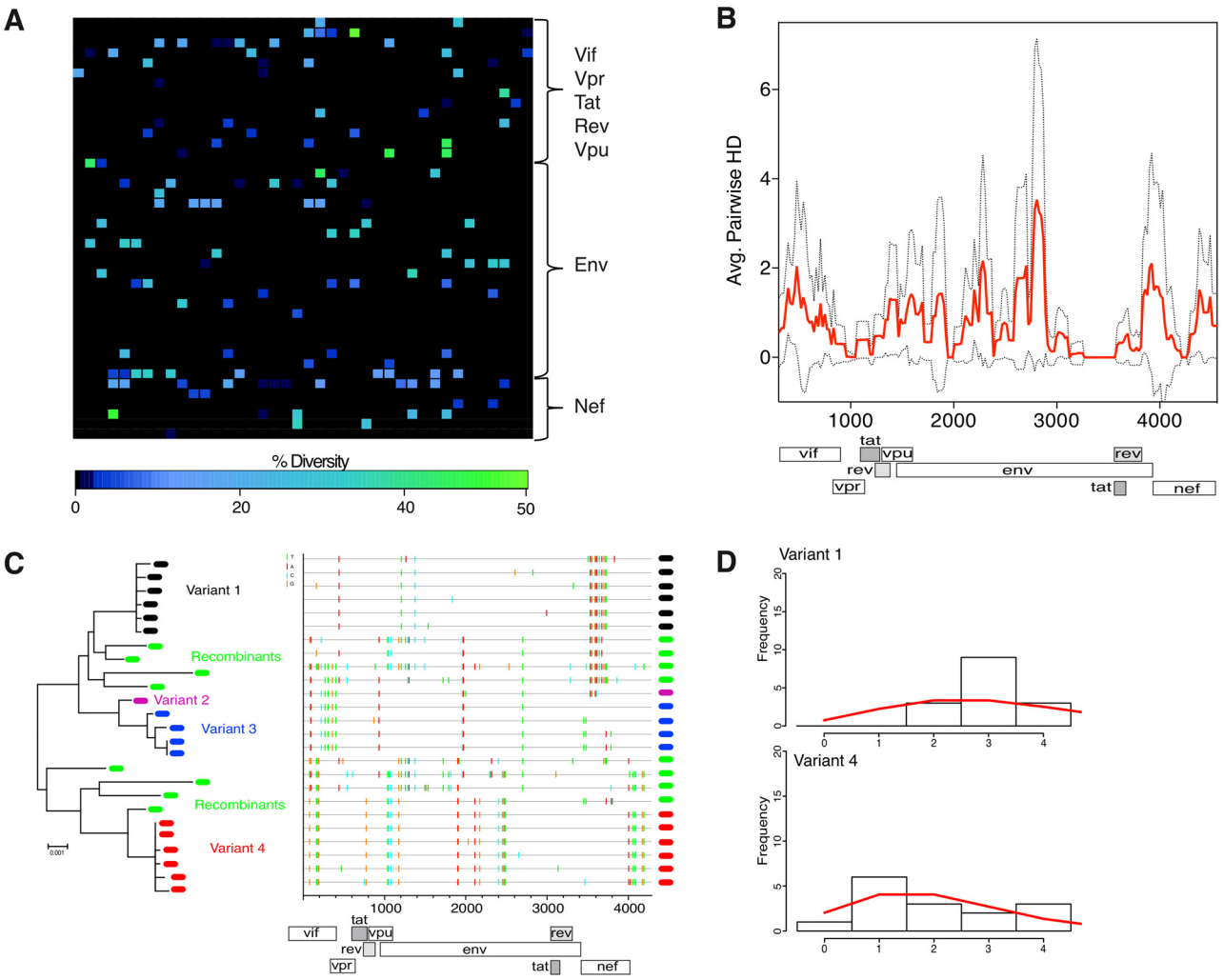

Fig 3. Subject 654207 exhibits high *env* diversity by both 454 and SGA/S reflective of multiple founder viruses. (A) Heat maps illustrating a number of sites exhibiting amino acid sequence diversity across the 3′ half of the genome as detected by 454 deep sequencing. Plotted is the percentage of amino acid diversity at each position with the first amino acid of Vif located in the top left corner of the grid and last amino acid of Nef located in the bottom right corner. Completely conserved residues are black and low-level variant residues (<10%) are dark blue, moderately variable residues (20%) are *sky blue* and highly variant residues (>40%) are green. **(B)** The average pairwise hamming distance calculated from 454 sequencing reads for the 3′ half of the genome is plotted with the APHD of 0.752 (red line) and standard deviation (dotted black line) shown. The plot shows a variable population with a high number of sites exhibiting throughout the genome exhibiting high-level diversity. **(C)** SGA/Ss from subject 654207 covering the 3′ half of the HIV-1 genome display a phylogeny (left) revealing productive infection by at least four viruses with inter-lineage recombination. Founder virus lineages are color-coded and labeled variant 1–4. Recombinant sequences are shown by green symbols. The highlighter plots (right) compare sequences for each subject's sequence set to an intrasubject consensus (uppermost sequence) and illustrates the pattern of nucleotide base mutations within sequences using short color-coded bars. **(D)** Hamming distance analysis of SGA/Ss showing infection by multiple viruses with Hamming distance frequencies (mean Hamming distance of 35.38) not conforming to model predictions of a single virus infection. Splitting of variants into their respective sub-lineages such as variants 1 and 4 demonstrate Hamming distance frequencies that do conform to model predictions of a single virus infection (red line). Subject 654207 is viral RNA positive but Western blot negative (stage II/III of infection).

doi:10.1371/journal.ppat.1005619.g003

or more distantly related viruses with high APHD scores of 0.267, 0.680 and 1.580 that was confirmed by SGA/S (which also indicated interlineage recombination). Finally, in subject 1051, deep sequencing demonstrated a high APHD score (0.490) reflective of infection by multiple viruses while our SGA/S data supported at least three founder viruses (S4 Fig). Coincidentally, this subject was also previously examined in detail by SGA/S [14] where at least 4

founder viruses were found (S4 Fig). Therefore, application of our APHD approach to 454 deep sequencing data successfully distinguished between single and multiple founder viruses as validated by parallel SGA/S.

Majority of MSM HIV-1 infections are established by a single founder virus

To further explore the use of deep sequencing data to examine the multiplicity of infection during the acute phase we next assembled a broader cohort of 74 subjects who had recently acquired HIV-1 either through sexual contact via MSM (n = 64) or HSX (n = 2) exposure including source plasma donors (SPD) (n = 6), or through percutaneous exposure (n = 1) or injecting drug use (IDU, n = 1), (see S1 Text and S1 Table). To reduce the potential of cohort-induced bias, subjects were predominantly selected from two distinct HIV-1 acute cohorts in Massachusetts and Germany. The majority of subjects (88%) were captured very early after infection, with 12 subjects in Fiebig stage I, 53 in Fiebig stage II/III, 5 in Fiebig stage IV and 4 in Fiebig V. All but two subjects were infected with HIV-1 subtype B, and all subjects exhibited high viral loads typical of acute infection with a median of 923,000 copies/ml (IQR: 270,378–4,100,000) and a median CD4+ T cell count of 413 cells/ml (IQR: 310–552).

To provide a more comprehensive and informative genome-wide analysis of multiple founder viruses, we PCR amplified the entire protein-coding region of HIV-1 in conjunction with 454 deep sequencing. While sequence coverage varied between samples and amplicons there was no significant difference in sequence coverage between amplicons with a median sequencing depth of 501 for *gag* (IQR: 285–794), 402 for *pol* (IQR: 203–679) and 600 for the 3′ genomic half (IQR: 359–862). Only 4 subjects had a 3' genomic half sequencing depth $<$ 200-fold with the remainder of subjects harboring sufficient sequence coverage to detect variants at 1%. Utilizing *env* sequences the APHD scores derived from the 74 acute HIV-1 infected subjects ranged from 0.0002 to 1.580 across different Fiebig stages (Fig 4A). The logistic classifier assigned 63 of these individuals to the single founder virus class, with the

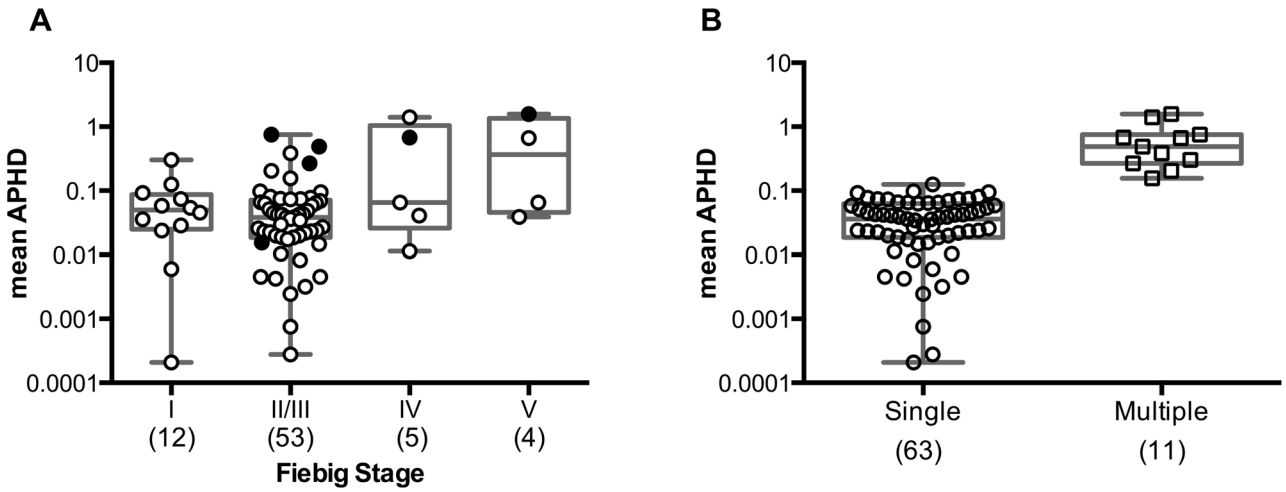

Fig 4. Complexity of acute HIV-1 infection revealed using deep sequencing data and the APHD approach. (A) Mean APHD of 74 newly deep sequenced acute HIV-1 infected subjects, illustrating a wide range of *env* diversity plotted according to Fiebig stages. Black circles depict the 6 samples in which SGA/S was also performed. **(B)** Classification of the 74 subjects into single vs. multiple founder viruses resulted in 63 subjects exhibiting a more homogeneous infection suggestive of productive clinical infection originating from a single virus, and 11 subjects exhibiting distinctly higher diversity indicative of heterogeneous infection and infection by multiple founder viruses. Each point corresponds to an individual subject with the number of subjects denoted on the x-axis in parenthesis under each Fiebig stage.

doi:10.1371/journal.ppat.1005619.g004

remaining 11 subjects categorized as infected by multiple variants (Fig 4B). Notably, 6 subjects were previously examined by Keele and colleagues [14] and our approach correctly assigned 5 of these as founded by a single genetic lineage and 1 by multiple lineages. Thus, among the 74 subjects studied the majority (85%) displayed low *env* diversity consistent with single variant transmission. Moreover, given that our cohort was predominantly MSM, this data indicates that the majority of these MSM infections (83%) were established by a single founder virus.

Impact the mode of transmission has upon the multiplicity of HIV-1 infection

Given that our findings were contrary to the prevailing understanding that higher risk infections such as MSM transmission frequently exhibit multiple founder viruses and that the results of such studies exploring the multiplicity of infection in HSX, MSM and IDU vary widely (Table 1) [7, 14, 15, 19, 44, 47–49], we undertook a meta-analysis of the multiplicity of HIV-1 transmission from a number of studies. We limited our analysis to the 354 subjects (Dataset 2) for whom full-length envelope sequences had been generated, encompassing MSM, HSX and IDU transmissions (the set included the 74 subjects sequenced during this study) [7, 14, 15, 19, 34, 44, 47, 49, 50]. Using the previously determined infection outcome (single or multiple) from these studies, we found the frequencies of multi-variant transmissions were comparable between HSX, MSM and IDU transmission ($P = 0.167$, Chi-Square test; Table 1). Recognizing that 6 of these subjects were counted twice between Keele and our study a re-analysis counting these subjects once still demonstrated no significant difference ($P = 0.177$, Chi-Square test). Taken together, these findings compiled from a number of different studies indicate that at the time of sampling there are no detectable differences in terms of multiplicity of infection between different modes of transmission.

Selection pressures are different between HSX and MSM transmissions

Given that we observed no differences in the number of founder viruses between MSM and HSX infection we sought to investigate whether other differences in the founder viruses may exist between the two modes of transmission. To do so we assembled a collection of 131 founder viruses (Dataset 3) derived from acute clade B HIV-1 infected subjects with a defined HSX (55 subjects) or MSM (76 subjects) exposure [14, 16, 19, 34, 44], of which 52 were newly generated by this study (S2 Table). We restricted our analysis to subjects sampled early (Fiebig stages I-III) and those classified as having a single variant infection. To distinguish the pattern of selection between MSM and HSX founder viruses we used the RELAX test [51] on a multiple sequence alignment of inferred founder strains, without variable loops (which are difficult to align, and could introduce false signal for selection). RELAX is a comparative codon-based phylogenetic framework implemented in HyPhy that formally tests whether selective pressures are intensified or relaxed in one subset of branches ("test") relative to a "reference" subset of branches, while allowing the strength of selection to vary from site to site in the alignment [51]. Traditionally, the intensity of selection is measured by estimating the ratio (ω) of rates at which nonsynonymous (dN) and synonymous substitutions (dS) are fixed [52], with an excess of dN ($\omega > 1$) an indicator of diversifying positive selection [53]. Here, the point estimates of ω distributions for MSM and HSX branch sets under the Partitioned Exploratory model assigned more codon sites in the HSX lineages to the positively selected category (5.4% [5.0–6.4%] in HSX vs 2.6% [2.3–2.9%] in MSM), although inferred that selection on these sites was stronger in MSM ($\omega = 15.8$ [14.4–17.5] in MSM vs $\omega = 9.2$ [8.2–9.6] in HSX. Therefore HSX founder sequences are subject to broader, albeit weaker, diversifying selective pressure than their MSM counterparts.

Table 1. Multiplicity of HIV-1 infection in HSX, MSM and IDU subjects.

Route of transmission	Study	Virus subtype	Total no. of subjects	Single variant transmission[a]		Multiple variant transmission[a]	
				n	%	n	%
HSX	Tully	B	8	7	87.5%	1	12.5%
	Abrahams [15]	C	69	54	78.3%	15	21.7%
	Haaland [7]	A or C	27	22	81.5%	5	18.5%
	Keele [14]	B	79	65	82.3%	14	17.7%
	Gnanakaran [34]	B	7	4	57.1%	3	42.9%
	Total		**190**	**152**	**80.0%**	**38**	**20.0%**
MSM	Tully	B	64	53	82.8%	11	17.2%
	Keele [14]	B	22	13	59.1%	9	40.9%
	Li [19]	B	28	18	64.3%	10	35.7%
	Herbeck [44]	B	9	8	88.9%	1	11.1%
	Gnanakaran [34]	B	9	7	77.8%	2	22.2%
	Total		**132**	**99**	**75.0%**	**33**	**25.0%**
IDU	Tully	B	1	1	100.0%	0	0.0%
	Keele [14]	B	1	0	0.0%	1	100.0%
	Bar [49]	B	10	4	40.0%	6	60.0%
	Masharsky [47]	A	13	9	69.2%	4	30.8%
	Dukhovlinova [50]	A	7	7	100.0%	0	0.0%
	Total		**32**	**21**	**65.6%**	**11**	**34.4%**

[a] Chi-Square test was used to examine the association between mode of transmission and infection outcome (p value = 0.167)

doi:10.1371/journal.ppat.1005619.t001

To determine whether other differences exist between HSX and MSM founder viruses we next compared their 'transmission index', which was shown to be predictive of which sequence in the donor will establish infection in the recipient [12]. In this recent study examining 137 heterosexual transmission pairs Carlson et al. revealed the preferential selection of viruses exhibiting a more wild-type or consensus-like sequence, perhaps reflective of an optimal HIV-1 genome or one exhibiting higher replicative fitness [12]. We hypothesized that the elevated risk of infection among MSM compared to HSX would result in reduced transmission selection bias, or as such a lower transmission index, upon MSM transmission. Using model weights taken from Carlson et al. [12], we indeed observed significantly lower transmission indices among MSM founder viruses compared to HSX founder viruses ($P = 3 \times 10^{-5}$, Fig 5). These data indicate that founder viruses from HSX are more closely related to a clade B consensus sequence, are likely to exhibit higher transmission fitness, and are more likely to have undergone proteome-wide selection at the transmission bottleneck as compared to their MSM counterparts, results consistent with the RELAX analysis. Taken together, these data suggest that viral populations from HSX and MSM infections are exposed to distinct selective pressures upon transmission.

Previously identified founder signature sites are enriched in HSX infection

Recently, Gnanakaran and colleagues reported on a number of sequence motifs in HIV-1 Env associated with founder viruses [34]. In particular, the presence of a histidine at position H12 and the absence of a potential N-linked glycosylation site (PNGS) at position N415 were found

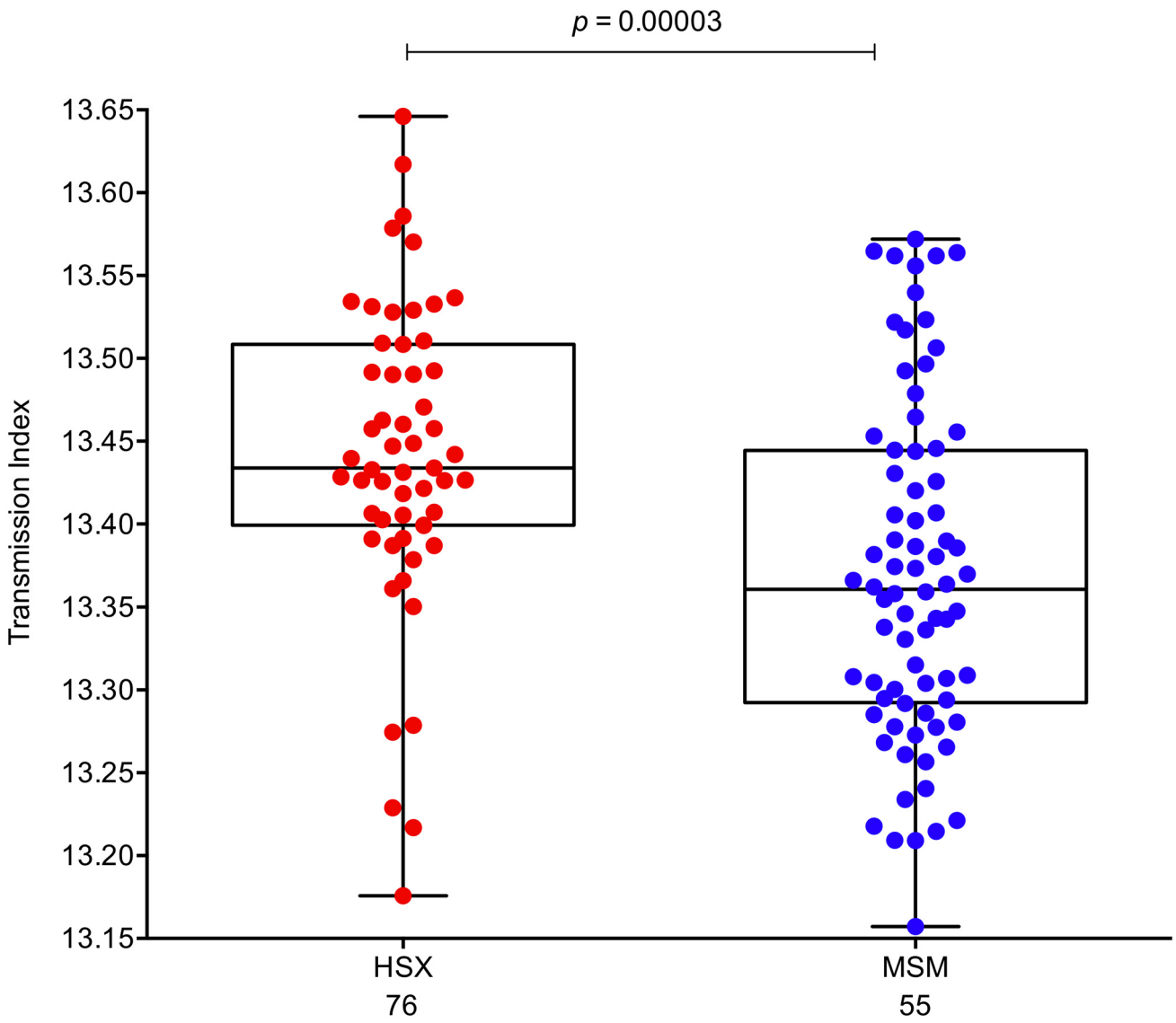

Fig 5. Selection bias intensified for HSX founder viruses compared to MSM founder viruses. The transmission index of a sequence was calculated using logistic regression with model weights taken from [12]. Black lines represent the median transmission index for the two risk groups. The overall transmission index of HSX (red circles) viruses is significantly higher than from MSM (blue circles) founder viruses ($P = 0.00003$, Mann-Whitney two-tailed test). The number of subjects in each category is denoted under each group.

doi:10.1371/journal.ppat.1005619.g005

to be selected in acute versus chronic viruses. Throughout this study we used the convention of "!" to express the loss of an amino acid at that position. For instance, mutating away from Asn at position 415 would be expressed as! N415. We investigated whether any of these 30 previously identified signature sites were enriched in HSX or MSM founder viruses in our dataset of 131 subtype B sequences. We applied a phylogenetically corrected logistical-regression model [54, 55], and employed a false-discovery rate approach (FDR) to account for multiple comparisons [56]. Adopting a q-value cutoff is critical in this study as thousands of tests were conducted. We generally chose a relatively high q-value cut off in our initial analysis; thus we expect approximately 20% of our sites from our first round of analysis to be by chance.

From this analysis, 3 of the 30 previously identified sites in Env were found significantly enriched in HSX founder viruses: R192 and N362 with a q value <0.2 and R633 with a

Table 2. Previously described signature sites enriched in HSX Founder viruses.

Predictor Variable[a]	Target Position (HXB2 #)[b]	Target Amino Acid[c]	Target Consensus	Direction [d]	*p* value	*q* value	Mutation Direction	Domain
HSX	R192	R	R	Adapted	0.0039	0.142	!R—> R	V2 Base
HSX	N362	N	N	Adapted	0.0017	0.1199	!N—> N	C3; PNGS
HSX	N362	K	N	Non-Adapted	0.00106	0.2312	K ->! K	C3; PNGS
HSX	R633	K	R	Non-Adapted	0.0128	0.2312	K ->! K	gp41

Using a phylogenetically corrected logistical-regression model we identified whether any previously published signature sites [34] were associated with HSX mode of transmission. All associations identified were corrected for multiple compairsons with sites with *q*-values <0.3 displayed.

[a] Predictor Variable: The risk behavior tested. In this case HSX mode of transmission.

[b] Target Position: The position and identity of the target variable in HXB2 numbering with cohort consensus amino acid listed.

[c] Target Amino Acid: The position and identity of the target amino acid.

[d] Direction: Adapted: amino acid is positively correlated with respect to HSX mode of transmission, Non-Adapted: amino acid is present only when risk group is absent

doi:10.1371/journal.ppat.1005619.t002

q value <0.3 (Table 2; nomenclature denotes cohort consensus residue and HXB2 numbering). More specifically, for residues R192 and N362 we observed selection for maintenance of a consensus residue in HSX founder viruses and at residues N362 and R633 we observed selection away from the non-consensus residue lysine (K) in HSX founder viruses. R192 is located at the base of the V2 loop and Gnanakaran et al. previously reported a loss of arginine (R→!R) in chronic viruses [34]. Similarly, they observed N362 mutating away from an asparagine (N→!N) while at residue R633 the pattern of R→!R was found enriched during chronic infection [34]. This analysis demonstrates that some of the previously found signature patterns of founder viruses may be primarily driven by HSX transmission events.

Identification of novel signature sites in Env associated with mode of transmission

Given the RELAX results, which identified a large number of sites under weak positive selection in the HSX group but a small number of sites under strong positive selection in the MSM group, we next conducted an unbiased search under a phylogenetic corrected framework to identify additional signature sites positively or negatively associated with MSM or HSX transmission. Although all associations identified below using one variable (i.e. HSX) were significant at P<0.05 when using the opposite variable (i.e. MSM), the models are nonetheless distinct and did identify different associations at the q-value cutoffs. From this analysis 7 sites (8 residues) were statistically associated with HSX founder viruses at a *q* of <0.2 (Q389, P724, A823, V832, R845 and A854) or <0.4 (K617, Table 3). All of these sites were found within the gp41 domain with the exception of Q389. Using MSM as the predictor variable revealed 16 sites (17 residues) that were statistically associated with MSM founder viruses at a *q* of <0.2 (T283, K343, Q389, E429, P724, E735, F752, R770, A823, H842, R845) or <0.4 (I165, N362, T465, G471, M518; Table 3). From these MSM-associated sites, 8 are within gp120 while the remaining 8 are within the gp41 domain.

Four of these sites (Q389, P724, A823 and R845) were found to overlap in both analyses and thus were found to be under significant, but opposing, selection in both HSX and MSM founder viruses. For example, at position Q389, the presence of proline is positively associated with HSX founder viruses while conversely the absence of proline is associated with MSM

Table 3. Signature sites identified between MSM and HSX Founder viruses in Env using a phylogenetic corrected method.

Predictor Variable [a]	Target Position (HXB2 #) [b]	Target Amino Acid [c]	Target Consensus	Direction [d]	p value	q value	Association Conditions [e]	Mutation Direction	Domain
MSM	I165	I	I	Non-Adapted	0.00135	0.36973		I ->! I	V2 Loop
MSM	T283	I	T	Non-Adapted	0.00082	0.13190	E275@K	T ->! I	C2 region; CD4 contact
MSM	K343	E	K	Adapted	0.00085	0.13190	N135@E, L515@L	!E -> E	C3 region
MSM	N362	N	N	Non-Adapted	0.00099	0.36973		N ->! N	C3 region
MSM	Q389	P	Q	Non-Adapted	0.00003	0.04125	P417@P	Q ->! P	V4 loop (Beside Lectin DC-SIGN BS)
HSX				Adapted	0.00006	0.04150		!P -> P	
MSM	E429	Q	E	Non-Adapted	0.00026	0.07752	T189@N, K304@R, E347@G	E ->! Q	C4 Region; CD4 contact
MSM	T465	N	T	Adapted	0.00365	0.33233	E293@Q, N404@S, T467@T, H721@H, A792@I	!N -> N	V5 Loop
MSM	G471	A	G	Adapted	0.00048	0.36973		!A -> A	C5 region; CD4
MSM	M518	M	M	Non-Adapted	0.00156	0.36973		M ->! M	gp41; fusion peptide
HSX	K617	K	K	Non-Adapted	0.00215	0.31619	F317@L, M426@R, R444@R	K ->! K	gp41 fusion domain
MSM	P724	P	P	Non-Adapted	0.00013	0.07103	N136@N, T139@T, N463@E	P ->! P	cytoplasmic tail; Kennedy epitope
HSX				Adapted	0.00007	0.04150		!P -> P	
MSM	E735	E	E	Adapted	0.00199	0.19669		!E -> E	cytoplasmic tail; Kennedy epitope
MSM	F752	F	F	Adapted	0.00100	0.13190	S209@S	!F -> F	cytoplasmic tail
MSM	R770	R	R	Adapted	0.00018	0.07103	T139@T, L856@Q	!R -> R	cytoplasmic tail; LLP-2 amphipathic helix
MSM	A823	G	A	Adapted	0.00128	0.13803	N396@E, Q834@Q	!G -> G	cytoplasmic tail
HSX				Non-Adapted	0.00111	0.18648		G ->! G	
MSM	A823	A	A	Non-Adapted	0.00128	0.13803	N396@E, Q834@Q	A ->! A	cytoplasmic tail
HSX				Adapted	0.00111	0.18648		!A -> A	
HSX	V832	I	V	Adapted	0.00082	0.18648	D412@N, T676@S	!I -> I	cytoplasmic tail; LLP-1 amphipathic helix
MSM	H842	N	H	Adapted	0.00040	0.09374	P724@Q, E275@K, N461@D	!N -> N	cytoplasmic tail; LLP-1 amphipathic helix
MSM	R845	T	R	Non-Adapted	0.00098	0.13190	I6@I, E87@A, N141@A, T818@T	R ->! T	cytoplasmic tail; LLP-1 amphipathic helix
HSX				Adapted	0.001017	0.186479	I6@I, N141@A	!T -> R	
HSX	A854	A	A	Adapted	0.0005	0.186479	T413@N, D621@D	!A -> A	cytoplasmic tail; LLP-1 amphipathic helix

[a] Predictor Variable: The risk behavior MSM or HSX.

[b] Target Position: The position and identity of the target variable in HXB2 numbering with cohort consensus amino acid listed.

[c] Target Amino Acid: The position and identity of the target variable.

[d] Direction: Adapted: amino acid is positively correlated with the risk group, Non-Adapted: amino acid is present only when risk group is absent.

[e] Association Conditions: The conditions under which the p-value was calculated. Models are built using forward selection, so these predictor variables were added to the model prior to the addition of the site. Sites are labeled with HXB2 numbering and residues correspond to cohort consensus

doi:10.1371/journal.ppat.1005619.t003

founder viruses. Specifically, only 1% (1 subject) of MSM sequences contain a proline at position Q389 while 20% (11 subjects) of HSX sequences contain this proline substitution. This Q389 signature site was only seen as significant when taken into account its association with a strongly covarying and highly conserved proline at residue P417 ($q = 0.086$; Table 3). Residue Q389 is located in the α4 helix of the V4 loop in relatively close proximity to the CD4-binding loop [57], and also neighbors the lectin DC-SIGN binding site (N386-T388). Selection here for the bulky hydrophobic residue proline might result in conformational changes in the pocket of the V4 loop and affect virus entry. Notably, this site has previously been found to be under positive selection, with Q389P variants specifically found to reduce HIV-1 replication capacity as well as increase neutralization resistance 15-fold to both b12 and sCD4 [58].

Within the sites identified with MSM as the predictor variable we observed a cluster of sites spatially concentrated (<15 Å) around the CD4 binding loop with at least 6 of the 8 MSM signature sites identified in gp120. Namely I165, T283, N362, Q389, E429, and G471, have previously been shown to alter HIV-1 replication, CD4 dependence and/or lie proximal to the CD4 biding site [57–69]. The identification of signature sites specific to either mode of sexual transmission, further supports that HSX and MSM transmissions are undergoing distinct selection pressures.

Impact of chronic virus residue frequencies on transmission signature sites

Finally, we sought to determine whether the aforementioned HSX and MSM signature sites described in Table 3 might simply reflect residue frequency differences in the respective HSX and MSM chronic 'donor' populations. To assess this we compared our acute sequence data to a panel of chronic viruses comprising over 1300 SGA/S sequences derived from 59 subjects with known HSX or MSM modes of transmission [34] (described as Dataset 4 in Methods). From the newly described signatures we found 12 sites (16 residues) (K343, N362, Q389, E429, T465, K617, E735, A823, V832, H842, R845, A854) that could be influenced by similar residue frequency differences between HSX and MSM chronically infected individuals (S5A Fig). In each of these cases the observed trend in founder viruses is mirrored at chronic infection. For instance, K343E we previously found to be associated with MSM where the presence of glutamic acid (E) is significantly higher in MSM founder viruses compared to HSX founder viruses. However, examination of chronic viruses also revealed the same significant trend where K343E is higher in MSM viruses compared to HSX viruses ($P = 0.026$, Chi-Square test). Thus, the higher frequency of K343E could be driven in part by its higher frequency in chronic circulating strains. Conversely, for 8 of the signature sites (9 residues) (I165, T283, G471, M518, P724, F752, R770 and R845) there was no evidence supporting an influence of differences in chronic residue frequencies impacting the transmission of these variants (S5B Fig). For example, at position P724 in gp41, there is clear selection for a proline (P) in HSX founder viruses while in circulating chronic HSX viruses the frequency of proline was significantly lower than in MSM viruses ($P < .0001$, Chi-Square test). Therefore, the presence of proline appears to be strongly selected for at transmission during HSX infection. Although it is difficult to decipher the origin of these selection events, since higher chronic frequencies could be similarly driven by differences selected at the time of transmission, at least 9 residues are strongly supportive of selection occurring upon transmission followed by regression during chronic infection. Thus, taken together we hypothesize that some of these variants may modestly improve overall fitness and hence be selected for during transmission and conversely selected against (or are neutral) over the course of ensuing chronic infection.

Discussion

In the present study we build upon a large body of work characterizing acute HIV-1 infection by shedding new light on the genetic properties of founder viruses distinguished by the two primary risk groups responsible for driving the global HIV-1 pandemic. Development of an average pairwise Hamming distance (APHD) approach, benchmarked to SGA/S data as the gold standard method, enabled us to distinguish between single and multiple founder virus infections using deep sequencing data. Application of this approach can also be extended to other genomic regions other than *env* (such as *gag* and *pol*) to assess the complexity of founder virus populations with the caveat that the sensitivity may vary. More importantly, these data demonstrated that the majority (83%) of MSM infections in our cohort exhibited a single founder virus—levels similar to those typically characteristic of HSX infections. Further characterization of these data demonstrated that HSX founder viruses do, however, appear to be under different selective pressures than MSM founder viruses, with HSX founder viruses subject to broader, albeit weaker, diversifying selective pressure than their MSM counterparts. Distinct genetic footprints were also found to be specific to HSX and MSM founder virus populations, supporting discrete selection pressures exhibited by each mode of sexual transmission.

The increasing use of next-generation sequencing has led to the development of specialized computational tools to reconstruct the viral haplotypes that constitute the quasispecies within a single host [70–74], and other methodological approaches have been used to screen for dual infection from deep sequencing data [75]. Although such tools improve our ability to probe the viral diversity they still suffer from a high rate of false positives [76]. As such, our method while based on a sliding window approach exhibits a low-error rate and is easily integrated into our existing data analysis pipeline [77]. While our estimate of the rate of multiple-variant infections in MSM (17%) is lower than the upper range of literature estimates (36–41%) [14, 19] it is not dissimilar from other published reports. Comparative full-length Env analyses have reported frequencies of only 11–14% of multiple founder viruses in MSM [44, 48], with 25% of MSMs in the STEP trial exhibiting multiple founder virus infections [78]. Additional reports using partial regions of Env demonstrated rates of only between 7–9% [79, 80]. One possible source of rate estimate discrepancy could be the inclusion of subjects sampled following peak viremia since immune-induced adaptations, including CD8+ T cell viral escape mutations, APOBEC-induced mutations or recombinants, selected during this window may distort the Poisson distribution model [81]. In our study of 74 subjects we were able to limit the majority of subjects (88%) to the earliest stages of HIV-1 infection (Fiebig I-III) as compared to rates as low as 60% of subjects in previous studies [14, 19]. Consistent with this hypothesis we found the odds of observing a multivariant infection in MSM was almost four times higher at later Fiebig stages IV-VI than in subjects sampled during earlier Fiebig I-III stages of infection (odds ratio, 3.86; 95% CI, 1.64 to 9.08; Fisher's exact test, $P = 0.003$). Importantly, although differences in the number of founder viruses could be attributable to compartmentalization in the source, examination of viruses in a cohort of Zambian transmission pairs found no evidence for preferential selection in the donor genital tract [41]. Thus, our findings support that although the risk of HIV-1 acquisition is significantly greater in MSM, this increased risk is not reflected by the transmission of an increased number of founder viruses.

While it is clear that various bottlenecks can limit the number of HIV-1 founder viruses successfully transmitted from a diverse, chronically infected donor (as recently reviewed in [82]), factors determining what viruses survive the bottleneck are not well understood. A recent study by Carlson et al. examining 137 heterosexual transmission pairs revealed the preferential selection of viruses exhibiting a more wild-type or consensus-like sequence, perhaps reflective of an optimal HIV-1 genome or one exhibiting higher replicative fitness [12]. Termed here a

'transmission index', this effect was more pronounced in female-to-male transmission compared to male-to-female transmission, and the effect could be attenuated by donor viral load and presentation of genital ulcers or inflammation (GUI). In our current study we also observed that HSX founder viruses exhibit significantly higher transmission indices than MSM founder viruses. In the absence of any donor sequence information, these data support a model in which there exists stronger selection forces, or increased opportunity for selection, upon the incoming viral quasispecies during HSX versus MSM transmission to optimize for wild-type or high-fitness variants for successful dissemination. Thus, this data fits with the prediction of modeling transmission as a binomial mixture process in which infection risk is inversely correlated with the strength of selection [12]. Given the elevated risk of infection in MSM compared to HSX we expected that the selection bias experienced by MSM founder viruses to be less stringent than that observed for HSX transmission. Such an overall reduced selective bias may make infection more conducive to even subtly weaker viruses. However, these viruses may need to optimize for enhanced CD4 binding in order to gain an advantage and successfully disseminate. Hence, the MSM-selected cluster of sites around the CD4 binding site may be evidence for such a scenario. Moreover, the differences in the selection bias at the transmission bottleneck may transcend to differences in clinical outcomes with reduced bias resulting in increased virulence with faster rates of reversion leading to higher fitness viruses emerging. On the other hand increasing the transmission selection bias may incite founder viruses that are optimized for increased fitness resulting in higher viral loads and poorer clinical outcomes in the newly infected individual.

Given that the majority of our HSX subjects were men (79% of our 131 founder virus dataset), and that within MSM the risk from unprotected receptive anal intercourse is >10 fold higher than for insertive anal intercourse [20], our study is effectively comparing penile (HSX) versus rectal (MSM) receptive routes of HIV-1 transmission. As such, our data would suggest that the rectal route may exert less selection pressure upon the incoming viral quasispecies than transmission through penile exposure, consistent with model predictions [12]. In the rhesus macaque model, even after penile exposures to a high dose SIV inoculum only a single variant founder population establishes infection [83, 84], while high dose intra rectal exposures are associated with greater numbers of founder viruses [85]. The rectal compartment is highly vulnerable to HIV-1 transmission with a single more fragile layer of columnar epithelium separating the lumen from the lamina propria as compared to the stratified squamous epithelium found in the ectocervix and vagina or the inner foreskin and the glans epithelia of the penis in uncircumcised men [86]. This, coupled with the density of HIV-1 target cells populating the rectum such as activated CD4+ T cells, macrophages and dendritic cells, may contribute to the greater risk of HIV-1 transmission associated with men who have receptive anal sex with men compared with HSX transmission in men or women [87, 88], but may also result in relaxation of the selective pressures upon the incoming quasispecies. Indeed, a briefer and narrower eclipse phase has been observed for HIV-1 infections acquired rectally compared to those acquired through the vaginal or penile tissues [83, 89, 90] where local viral expansion is necessary before the dissemination of infection to the bloodstream. Thus, MSM viruses may not need to undergo the same level of selection that HSX viruses must endure for successful replication and systemic dissemination.

Given previous studies have espoused differences in variable loop length and potential N-linked glycosylation site count in the transmitted virus for HIV-1 subtypes A and C [21, 24, 32], although less clear in subtype B infections [31–33], we searched for any association between variable loop diversity and mode of transmission. Detailed analyses of the variable loops revealed only one such putative association with MSM founder viruses encoding a more compact V2 loop compared to HSX founder viruses (mean of 40.9 residues for MSM and 42.5

for HSX) although such an effect did not reach statistical significance after correcting for multiple comparisons. While one study has identified through the comparison of acute versus chronic HIV-1 sequences signature mutations associated with founder viruses [34], our study extends these findings by identifying an array of genetic signatures that may be distinct between MSM and HSX risk groups. Interestingly, the majority of residues found to be associated with HSX risk were located within the gp41 domain, and in particular within the cytoplasmic tail. At residue K617 we observed maintenance of a consensus lysine residue where mutations at this position of the gp41 fusion domain have been shown to significantly reduce viral entry [91]. This unusually long and highly conserved domain of approximately 150 amino acids modulates a diverse array of functions, including viral replication, Env incorporation into virions, and intracellular trafficking and endocytosis to regulate levels of Env surface expression (reviewed in [92]). The C-terminal half of the cytoplasmic domain is characterized by the presence of three structurally conserved α-helices designated lentivirus lytic peptide 1 (LLP-1), LLP-2, and LLP-3 [93–95]. Notably, three of the HSX signature residues (V832, R845 and A854), where we saw selection for the consensus residue in HSX, are located within the LLP-1 region, which is associated with Env incorporation into virions [96, 97]. Thus, these specific signature sites within gp41 may influence Env virion incorporation levels and viral entry, thus increasing transmissibility. It is also conceivable that these specific mutations may alter the conformation of the envelope trimer in such a manner that is favored for initial infection. Notably, many of the strongest transmission signature sites observed by Gnanakaran et al were also in the cytoplasmic domain in addition to enrichment for histidine at residue H12 in the signal peptide [34], the later of which has been demonstrated to increase Env incorporation and infectious titers [35]. Regardless of the precise mechanism, these data support a role for selection upon the cytoplasmic domain of HIV-1 gp41 during transmission.

In contrast, nearly half of the residues associated with MSM risk were located in gp120 with six residues (T283, N362, Q389, E429, T465, G471) clustered around the CD4-binding pocket with the potential to influence CD4 binding (Fig 6). In addition to residue Q389 described earlier, which is located in close proximity to the CD4-binding loop [57], position T283 has been shown to affect CD4 binding site exposure and CD4 binding of gp120s derived from brain and other tissues [60]. Similarly, presence of the N362 PNLG site in the C3 region has been shown to enhance CD4 binding to gp120 as well as cell-cell fusion [68, 98], potentially reducing CD4 dependence by stabilizing the CD4-bound confirmation of gp120 [68]. Meanwhile, at residue E429 located in the C4 domain of gp120 we observed selection for glutamine (E429Q) where prior work has identified this residue as being critically important for the binding of CD4-blocking MAbs [65] and implicated in altering resistance to the entry inhibitors BMS-806 and #155 [63], as well as enhancing HIV-1 replication *in vitro* [59]. Within the V5 loop, residue T465 has also been associated with a neutralization-resistant phenotype [99], while finally at residue G471 where we observed selection for an alanine (G471A) the variants G471R/E have been shown to impart resistance towards CD4 mimetic compounds [61]. Thus, many of the signature sites identified in MSM in gp120 may influence gp120-CD4 interactions for enhanced interactions with CD4.

A limitation of this study is the inclusion of subjects designated as source plasma donors (SPD). These subjects had limited behavioral information available but as part of routine blood-banking practice underwent extensive questioning for HIV-1 risk behaviors and denied having sex for money, homosexual activity or i.v. drug use. Nevertheless, self-reporting of risk behaviors among paid plasma donors is imperfect and it is plausible that some subjects whom were designated as belonging to the HSX risk group as previously categorized [19, 34] may have additional risk behaviors. However, the exclusion of all SPD subjects from our comparative analysis indicates that the frequency of multi-variant transmission between HSX and

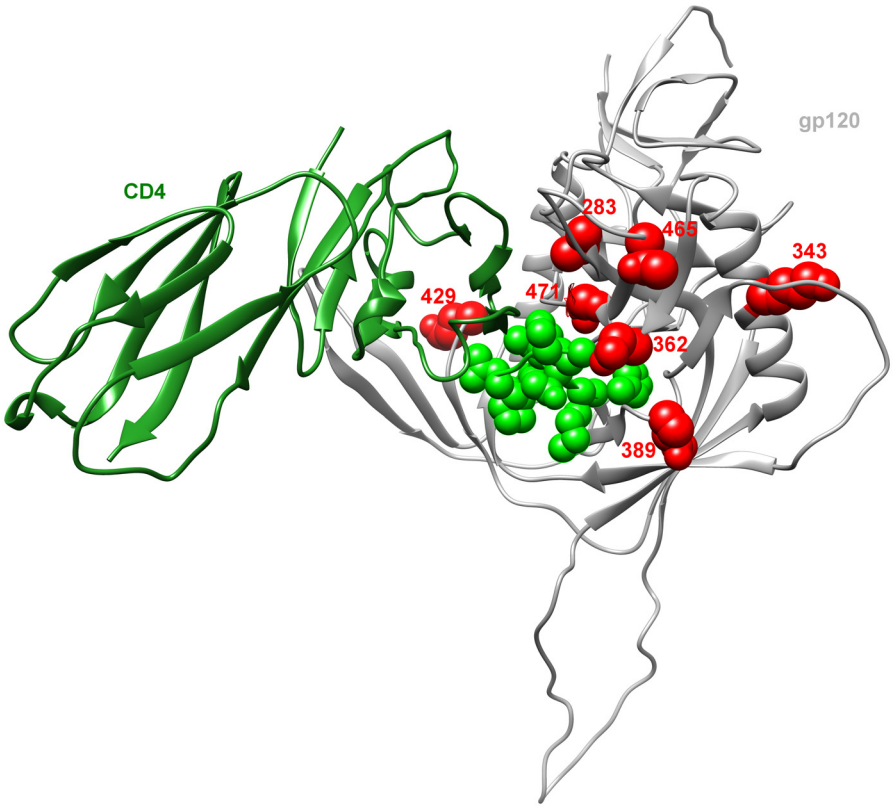

Fig 6. Mapping of signature sites on the three-dimensional structure of gp120 shows clustering around the CD4-binding site. A ribbon representation of the crystal structure from the JRFL gp120 molecule (grey) bound to CD4 molecule (green) (PDBID: 2B4C). The CD4 binding site is highlighted in transparent green while signature sites 283, 343, 362, 389, 429, 465 and 471 are all depicted as *red* space-filling residues.

doi:10.1371/journal.ppat.1005619.g006

MSM transmission remained the same (22% vs. 25%, odds ratio, 1.16; 95% CI, 0.66 to 2.06; Fisher's exact test; $P = 0.66$). The reported higher transmission indices for HSX founder viruses also continued to be significant when compared with MSM founder viruses ($P = 0.0007$, Fisher's exact test). Thus, the overall study findings are unlikely to have been influenced by the inclusion of this subject group alone.

While HIV-1 acquisition in MSM may be immunologically and virologically distinct from that of heterosexual exposure, we observe no differences in the number of HIV-1 founder viruses. Although the severe HIV-1 transmission bottleneck has a stochastic component in which any reasonably fit CCR5-tropic virus may be capable of establishing productive infection, our data do argue that any selection bias may be comparatively relaxed with ano-rectal MSM transmission, potentially due to the greater frequency of target cells at the site of transmission and the distinct kinetics of virus dissemination [87–90]. Conversely, upon heterosexual exposure we observe an increased bias for consensus-like viruses with a potentially higher replicative fitness that must undergo local viral replication prior to systemic dissemination. In the era of new therapeutic approaches such as AAV delivered HIV-1 inhibitors [100] and effective pre-exposure prophylaxis [101], it remains to be seen whether breakthrough infections will lead to higher fitness viruses being selected for resulting in more severe clinical outcomes, akin to what was observed during the CAPRISA 004 vaginal microbicide gel trial [102]. More accurate estimation of the frequency of multi-variant infection will also aid in evaluating the clinical

impact that infection with multiple variants has on disease progression as a number of studies have reported associations with increased viral load [103–105] and faster CD4+ T-cell decline [106, 107] resulting in a shorter time to AIDS. Finally, given the recent application of discerning the number of founder viruses as a measurement of relative protection from infection [108, 109], more critical delineation of the genetic make-up and complexity of the founder virus population may be important towards the development of an effective HIV-1 vaccine.

Materials and Methods

Study subjects

Plasma samples were obtained from subjects with acute or early HIV-1 infection enrolled in HIV-1 cohorts in Berlin, Germany, and Massachusetts, California, North Carolina, and South Carolina, USA. The clinical and sociodemographic characteristics of the study participants can be found in S1 Table and Fiebig staging criteria are described in Supplementary Materials and Methods in S1 Text. All study subjects gave written informed consent and plasma collections were performed with local institutional review board and other regulatory approvals. This study was approved by the Institutional Review Board of Massachusetts General Hospital.

PCR amplification and 454 sequencing

HIV-1 was PCR amplified and 454 sequenced using a nested RT-PCR with 3 amplicons overlapping the genome. Briefly, viral RNA was isolated from 1ml of plasma using the QiAmp Viral RNA Mini Kit (Qiagen, Valencia, CA) and RT-PCR of near full-length HIV-1 genomes performed using nested-PCR primers specific for gag, pol and 3′ half of the viral genome (see S1 Text and S3 Table). Pooled PCR products were prepared for sequencing on the 454 Genome Sequencer Junior (Roche) using the Nextera DNA Sample Prep Kit and data processed performed using our previously published sequence analysis pipeline [77, 110] (see S1 Text).

Average pairwise Hamming distance measurement of 454 sequence diversity

Following alignment of cleaned reads to the consensus assembly sequence a sliding window approach was used to collect all reads that covered a window of 120bp with a step size of 21bp. We then calculated the pairwise Hamming distance (HD) (defined as the number of base positions at which the genomes differ, excluding gaps) for all reads and averaged this value over all windows to obtain the average pairwise Hamming distance (APHD). To formalize the criteria and evaluate the APHD approach for its discriminative ability we obtained SGA/S sequence data from studies where each sample had been previously designated as infection by a single virus or infection by multiple viruses [14, 19]. To introduce sequencing errors, a synthetic read dataset was simulated from the SGA/S data using the 454 sequencing error profile from ART a next-generation read simulator [111]. Reads were simulated with varying degrees of coverage in order to achieve the same representative coverage obtained from our deep sequencing data with multiple replicates performed to rigorously assess the uncertainty due to 454-like sampling.

Discrimination of 454 sequence data for single and multiple founder viruses

To differentiate between single and multiple founder viruses using 454 sequencing reads intra-patient HIV-1 genetic diversity was first characterized as previously described, with nucleotide-phasing information used to distinguish true variants from sequencing errors (see S1 Text

and S6 Fig). The mean APHD is calculated using the sliding window approach as previously discussed. Reads were then carefully inspected to rule out factors that might compromise the amount of diversity such as the emergence of CTL escape mutations or early reversions whereby diversity would be restricted to narrow windows specific to known CD8+ T cell epitopes specific to the subject's HLA and would appear as distinct peaks on the APHD landscape. More specifically, for each individual the optimal "A-list" of CTL epitopes restricted by a subject's HLA alleles was generated and local haplotype windows were reconstructed across each of the epitopes to assess for any evidence of putative CTL escape mutations. Any windows in which CTL escape mutations were found were further examined and the contribution of this window to the overall APHD score was evaluated to unsure that it did not unduly influence any subject designation as infected by multiple viruses.

Single genome amplification and sequencing of 3′ half genome

cDNA was serially diluted and amplified as previously described [14, 17, 112], and the amplified products sequenced using a 454 GS Junior (see S1 Text).

Datasets

A number of previous published datasets were used throughout this study and are numbered accordingly. Briefly, Dataset 1 comprising the SGA/S data collected from 127 acute individuals sampled at varying times post-infection and clinically staged as described by Fiebig et al [113] were used [14, 19] to test the performance of the APHD approach. To explore the relationship between multiplicity of infection and mode of transmission, Dataset 2 encompassing 354 subjects, for whom full-length SGA/S envelope sequences had been generated, encompassing MSM, HSX and IDU transmissions (including the 74 subjects newly deep sequenced during this study) [7, 14, 15, 19, 34, 44, 47, 49, 50] were obtained. From the total of 354 subjects a subset of 131 HIV-1 founder viruses from subjects reporting a sexual exposure were selected [14, 16, 19, 34, 44] (Dataset 3). The criteria for subject selection was restricted to clade B infections sampled early (Fiebig stages I-III) and included only subjects previously classified as being infected with a single virus (subjects listed in S2 Table). Dataset 4 included chronic samples obtained from Gnanakaran et al [34]. This dataset contained over 1300 SGA/S sequences derived from 59 subjects with known exposure status defined as HSX or MSM. These sequences were from individuals who were not on anti-retroviral therapy, and infected for a minimum of two years and represented clade B infections predominantly collected in the United States.

Mathematical model of random evolution

PoissonFitter was used to test the hypothesis that a single virus establishes infection [81]. PoissonFitter performs two tests: one test is based on the fit of the Poisson model to the frequency distribution of the Hamming distance observed in each sample; the other is a topological test to verify that observed frequencies are distributed according to a star-like phylogeny (for this test, no formal statistic is available and consequently no p-value is obtained). In this model the main assumption is that a single founder virus evolves under neutral evolution, generating a star-like phylogeny, with a distribution of mutations conforming to a Poisson distribution [13, 14].

Identification of signatures sites in Env using a phylogenetically corrected approach

We used the phylogenetically corrected logistic regression to identify sites positively or negatively associated with MSM or HSX state [54, 55]. Briefly, this approach uses standard logistic regression, with the modification that information from the phylogeny is used to inform the bias parameters. Rather than assuming the sequences are independent and from the same distribution, the phylogenetically corrected logistic regression model assumes the sequences are drawn from a known phylogenetic structure. Using this structure separate phylogenetically corrected logistic regression models were learned from each amino acid at each site. Phyml v3.0 [114] was used to infer the phylogenetic structure, using all sequences available in this study. Using this structure, separate phylogenetically corrected logistic regression models were learned for each amino acid at each site. For "indirect" models, only a single feature representing MSM or HSX status was used. For "direct" models, forward selection was used to learn a model that possibly included covariation from other sites in addition to MSM or HSX status. Note that although MSM = 1-HSX, the models are distinct and may identify different associations. In our case, while the association's differed when using q-value cutoff, all associations identified using one variable were significant at $P<0.05$ when using the opposite variable. The q-value is the minimal false discovery rate that adjusts for multiple tests [56]. The appropriate choice of q-value threshold is context specific and depends on how the results will be interpreted. In the present study, we typically report all tests where q is <0.2 (implying that we expect 20% of reported tests to be false positives) but sometimes report higher q values to include sites in a hypothesis-raising framework. The associations identified in this study are referred to as *adapted* and *nonadapted* forms. Adapted forms (commonly referred in many studies as escape variants) are amino acids significantly enriched for in the presence of the risk behavior in question. Nonadapted forms (also commonly called wild-type or susceptible forms) are amino acids significantly depleted in the presence of the risk behavior.

Detection of selection under a phylogenetic framework

To distinguish the pattern of selection between MSM and HSX founder viruses we used a comparative codon-based phylogenetic framework test implemented in HyPhy that formally tests whether selective pressures are intensified or relaxed relative to a subset of branches [51]. In this case we assessed whether selective strength on the test subset of branches is compressed toward or repelled away from neutrality, relative to the reference subset of branches. Under this analysis we labeled HSX viruses as reference branches while MSM viruses were labeled as test branches. Internal branches were labeled as MSM or HSX if all of their descendants were also labeled MSM or HSX, respectively. The Null model, which forces HSX and MSM to share the same selective regime can be rejected in favor of the Partitioned Exploratory model ($p = 0.009$, Likelihood Ratio Test). The Partitioned Exploratory model is merely the model where the "test" and "reference" branches in the tree are endowed with completely independent discrete distributions of omega parameters. The Alternative model, which is a restriction of the Partitioned Exploratory model forces the proportions of sites under different types of selection to be the same can be similarly rejected ($p = 0.008$). All confidence intervals listed are 95% profile likelihood approximations.

Transmission index

In the context of heterosexual linked transmission pairs, we previously trained a model that estimates the probability that any particular amino acid will be transmitted from a donor to a

recipient [12]. Although this model was trained grouping together all residues at all sites using a generalized linear mixed model, we showed that a simple extension to full sequences, which we called the "Transmission Index" was predictive of which sequence would establish infection. Here we computed the transmission index of each founder virus in MSM vs HSX founder viruses using logistic regression, with model weights taken from Table 2 of [12]. Amino acid conservation and covariation was taken from the clade B envelope sequence data of [115].

Statistical analysis

All statistical analysis was performed using JMP Pro, version 12 (SAS Institute). Descriptive measures were used to summarize the data. Continuous variables were summarized using median and inter quartile range (IQR); categorical variables were summarized using frequency and percent (%). Chi-square and Mann-Whitney tests were used to compare categorical and continuous variables between the study groups, respectively.

Supporting Information

S1 Text. Supplementary materials & methods.
(DOC)

S1 Fig. High *env* diversity in subject 882283 reflecting a heterogeneous infection suggestive of infection by at least 4 founder viruses. (A) Heatmap from the 454 sequencing data illustrating a diverse number of sites throughout the 3′ half of the genome showing up to 40% codon diversity. **(B)** The APHD plots showing a mean APHD of 0.306 (red line) and standard deviation (dotted black line) demonstrating a high level of diversity across the 3′ half of the HIV-1 genome. **(C)** SGA sequences displaying a phylogeny (left) revealing infection by at least four viruses with inter-lineage recombinants. Founder virus lineages are color-coded while recombinant sequences are shown by green symbols. Highlighter plots (right) compare sequences for each subject's sequence set to an intrasubject consensus (uppermost sequence) and depict the pattern of nucleotide base mutations. Subject 882283 was viral RNA positive but Western blot negative (Fiebig stage II/III of infection).
(TIF)

S2 Fig. High *env* diversity in subject 702865 reflecting a heterogeneous infection suggestive of infection by at least 4 founder viruses with inter-lineage recombination. (A) Heatmap from the 454 sequencing data illustrating a diverse number of sites throughout the 3′ half of the genome showing up to 40% codon diversity. **(B)** The APHD plots showing a mean APHD of 0.593 (red line) and standard deviation (dotted black line) demonstrating a high level of diversity across the 3′ half of the HIV-1 genome. **(C)** SGA sequences displaying a phylogeny (left) revealing infection by at least four viruses with inter-lineage recombinants. Founder virus lineages are color-coded while recombinant sequences are shown by green symbols. Highlighter plots (right) compare sequences for each subject's sequence set to an intrasubject consensus (uppermost sequence) and depict the pattern of nucleotide base mutations. Subject 702865 was viral RNA positive but Western blot indeterminate (Fiebig stage IV of infection).
(TIF)

S3 Fig. High *env* diversity in subject 574194 reflecting a heterogeneous infection suggestive of infection by at least 3 founder viruses. (A) Heatmap from the 454 sequencing data illustrating a diverse number of sites throughout the 3′ half of the genome showing up to 40% codon diversity. **(B)** The APHD plots showing a mean APHD of 1.718 (red line) and standard deviation (dotted black line) demonstrating a high level of diversity across the 3′ half of the HIV-1 genome. **(C)** SGA sequences displaying a phylogeny (left) revealing infection by at least

four viruses with inter-lineage recombinants. Founder virus lineages are color-coded while recombinant sequences are shown by green symbols. Highlighter plots (right) compare sequences for each subject's sequence set to an intrasubject consensus (uppermost sequence) and depict the pattern of nucleotide base mutations. Subject 574194 was viral RNA positive but Western blot positive with 3 bands (Fiebig stage V of infection).
(TIF)

S4 Fig. Comparison of SGA sequences for subject 1051 who displays evidence of infection with multiple founder viruses. SGA sequences derived from this study (red) were compared to sequences derived from Keele et al. [14] in which 3 additional timepoints were sequenced. Previous analyses by Keele revealed infection by at least 4 founder viruses while in this study we found infection by at least 3 viruses [14].
(TIF)

S5 Fig. Frequency comparison of signature sites between acute and chronic viruses during HSX and MSM infection. Frequency comparison of previously found signatures sites in our panel of HSX and MSM founder viruses compared to a dataset of chronic HSX and MSM viruses. Chronic viruses comprised 462 SGA/S sequences derived from 24 chronically infected subjects who reported heterosexual as their risk factor for HIV-1 infection and 867 SGA/S sequences from 35 chronically infected MSM subjects. Sites and amino acids examined are listed on the x-axis with the frequency of that amino acid at that position shown on the y-axis. Acute HSX (red bars), acute MSM (blue bars), chronic HSX (red striped bars) and chronic MSM (blue striped bars). **(A)** Sites that show the same frequency trend in chronic and acute infection are depicted. **(B)** Sites that show the opposite trend in chronic and acute infection are shown. Sites that showed a statistically significant difference between the chronic stage of infection at a *P* value of less than 0.05 (Chi-Square test) are indicated by a single *asterisk* (*).
(TIF)

S6 Fig. Distribution of point mutation errors found within a plasmid control. (A) Distribution of point mutation errors across the HIV-1 genome from 3 independent sequencing runs of an HIV-1 NL4-3 plasmid control. **(B)** Effect of PCR amplification on point mutation variant mismatches across three independent PCR and sequencing runs.
(TIF)

S1 Table. Clinical description of cohort study subjects with acute or early HIV-1 infection.
(XLSX)

S2 Table. A description of the 131 founder viruses used in this study.
(XLSX)

S3 Table. A list of all primers used to amplify the HIV-1 genome.
(XLSX)

Acknowledgments

We thank the Enterprise Research Infrastructure & Services at Partners Healthcare for their in-depth support and for the provision of computational resources.

Author Contributions

Conceived and designed the experiments: DCT TMA. Performed the experiments: CBO DJB KAP HEB ADG AMS. Analyzed the data: DCT CBO DJB SLKP MG JMC. Contributed

reagents/materials/analysis tools: REB SLKP JMC MAA KL GM SBB JT NJL MRH ZLB PJN ESR KHM HJ BDW MA. Wrote the paper: DCT TMA SLKP ZLB JMC.

References

1. UNAIDS. 2013.
2. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J Virol. 1993; 67(6):3345–56. Epub 1993/06/01. PMID: 8497055
3. Wolinsky SM, Wike CM, Korber BT, Hutto C, Parks WP, Rosenblum LL, et al. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science. 1992; 255 (5048):1134–7. Epub 1992/02/28. PMID: 1546316
4. Zhu T, Mo H, Wang N, Nam DS, Cao Y, Koup RA, et al. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science. 1993; 261(5125):1179–81. Epub 1993/08/27. PMID: 8356453
5. Wolfs TF, Zwart G, Bakker M, Goudsmit J. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology. 1992; 189(1):103–10. Epub 1992/07/01. PMID: 1376536
6. McNearney T, Hornickova Z, Markham R, Birdwell A, Arens M, Saah A, et al. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc Natl Acad Sci U S A. 1992; 89(21):10247–51. Epub 1992/11/01. PMID: 1438212
7. Haaland RE, Hawkins PA, Salazar-Gonzalez J, Johnson A, Tichacek A, Karita E, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. PLoS Pathog. 2009; 5(1):e1000274. Epub 2009/01/24. doi: 10.1371/journal.ppat.1000274 PMID: 19165325
8. Zhang ZQ, Wietgrefe SW, Li Q, Shore MD, Duan L, Reilly C, et al. Roles of substrate availability and infection of resting and activated CD4+ T cells in transmission and acute simian immunodeficiency virus infection. Proc Natl Acad Sci U S A. 2004; 101(15):5640–5. Epub 2004/04/06. PMID: 15064398
9. Naranbhai V, Abdool Karim SS, Altfeld M, Samsunder N, Durgiah R, Sibeko S, et al. Innate immune activation enhances hiv acquisition in women, diminishing the effectiveness of tenofovir microbicide gel. J Infect Dis. 2012; 206(7):993–1001. Epub 2012/07/26. PMID: 22829639
10. Taha TE, Hoover DR, Dallabetta GA, Kumwenda NI, Mtimavalye LA, Yang LP, et al. Bacterial vaginosis and disturbances of vaginal flora: association with increased acquisition of HIV. AIDS. 1998; 12 (13):1699–706. Epub 1998/10/09. PMID: 9764791
11. Galvin SR, Cohen MS. The role of sexually transmitted diseases in HIV transmission. Nat Rev Microbiol. 2004; 2(1):33–42. Epub 2004/03/24. PMID: 15035007
12. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, et al. HIV transmission. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science. 2014; 345 (6193):1254031. Epub 2014/07/12. doi: 10.1126/science.1254031 PMID: 25013080
13. Lee HY, Giorgi EE, Keele BF, Gaschen B, Athreya GS, Salazar-Gonzalez JF, et al. Modeling sequence evolution in acute HIV-1 infection. J Theor Biol. 2009; 261(2):341–60. Epub 2009/08/08. doi: 10.1016/j.jtbi.2009.07.038 PMID: 19660475
14. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. 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. Epub 2008/05/21. doi: 10.1073/pnas.0802203105 PMID: 18490657
15. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping LH, 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):3556–67. Epub 2009/02/06. doi: 10.1128/JVI.02132-08 PMID: 19193811
16. 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(6):1273–89. Epub 2009/06/03. doi: 10.1084/jem.20090378 PMID: 19487424
17. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol. 2008; 82(8):3952–70. Epub 2008/02/08. doi: 10.1128/JVI.02660-07 PMID: 18256145

18. Fischer W, Ganusov VV, Giorgi EE, Hraber PT, Keele BF, Leitner T, 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. Epub 2010/09/03. doi: 10.1371/journal.pone.0012303 PMID: 20808830

19. Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, et al. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. PLoS Pathog. 2010; 6(5):e1000890. Epub 2010/05/21. doi: 10.1371/journal.ppat.1000890 PMID: 20485520

20. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. Estimating per-act HIV transmission risk: a systematic review. AIDS. 2014; 28(10):1509–19. Epub 2014/05/09. doi: 10.1097/QAD.0000000000000298 PMID: 24809629

21. Derdeyn CA, Decker JM, Bibollet-Ruche F, Mokili JL, Muldoon M, Denham SA, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science. 2004; 303 (5666):2019–22. Epub 2004/03/27. PMID: 15044802

22. Liao HX, Tsao CY, Alam SM, Muldoon M, Vandergrift N, Ma BJ, et al. Antigenicity and immunogenicity of transmitted/founder, consensus, and chronic envelope glycoproteins of human immunodeficiency virus type 1. J Virol. 2013; 87(8):4185–201. Epub 2013/02/01. doi: 10.1128/JVI.02297-12 PMID: 23365441

23. Parker ZF, Iyer SS, Wilen CB, Parrish NF, Chikere KC, Lee FH, et al. Transmitted/founder and chronic HIV-1 envelope proteins are distinguished by differential utilization of CCR5. J Virol. 2013; 87 (5):2401–11. Epub 2012/12/28. doi: 10.1128/JVI.02964-12 PMID: 23269796

24. Ping LH, Joseph SB, Anderson JA, Abrahams MR, Salazar-Gonzalez JF, Kincer LP, et al. Comparison of viral Env proteins from acute and chronic infections with subtype C human immunodeficiency virus type 1 identifies differences in glycosylation and CCR5 utilization and suggests a new strategy for immunogen design. J Virol. 2013; 87(13):7218–33. Epub 2013/04/26. doi: 10.1128/JVI.03577-12 PMID: 23616655

25. Wilen CB, Parrish NF, Pfaff JM, Decker JM, Henning EA, Haim H, et al. Phenotypic and immunologic comparison of clade B transmitted/founder and chronic HIV-1 envelope glycoproteins. J Virol. 2011; 85(17):8514–27. Epub 2011/07/01. doi: 10.1128/JVI.00736-11 PMID: 21715507

26. 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 alpha4beta7. PLoS Pathog. 2012; 8(5):e1002686. Epub 2012/06/14. doi: 10.1371/journal.ppat.1002686 PMID: 22693444

27. Parrish NF, Gao F, Li H, Giorgi EE, Barbian HJ, Parrish EH, et al. Phenotypic properties of transmitted founder HIV-1. Proc Natl Acad Sci U S A. 2013; 110(17):6626–33. Epub 2013/04/02. doi: 10.1073/pnas.1304288110 PMID: 23542380

28. Nawaz F, Cicala C, Van Ryk D, Block KE, Jelicic K, McNally JP, et al. The genotype of early-transmitting HIV gp120s promotes alpha (4) beta(7)-reactivity, revealing alpha (4) beta(7) +/CD4+ T cells as key targets in mucosal transmission. PLoS Pathog. 2011; 7(2):e1001301. Epub 2011/03/09. doi: 10.1371/journal.ppat.1001301 PMID: 21383973

29. Herbeck JT, Nickle DC, Learn GH, Gottlieb GS, Curlin ME, Heath L, et al. Human immunodeficiency virus type 1 env evolves toward ancestral states upon transmission to a new host. J Virol. 2006; 80 (4):1637–44. Epub 2006/01/28. PMID: 16439520

30. Redd AD, Collinson-Streng AN, Chatziandreou N, Mullis CE, Laeyendecker O, Martens C, et al. Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. J Infect Dis. 2012; 206(9):1433–42. Epub 2012/09/22. doi: 10.1093/infdis/jis503 PMID: 22997233

31. Frost SD, Liu Y, Pond SL, Chappey C, Wrin T, Petropoulos CJ, et al. Characterization of human immunodeficiency virus type 1 (HIV-1) envelope variation and neutralizing antibody responses during transmission of HIV-1 subtype B. J Virol. 2005; 79(10):6523–7. Epub 2005/04/29. PMID: 15858036

32. Chohan B, Lang D, Sagar M, Korber B, Lavreys L, Richardson B, et al. Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1–V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. J Virol. 2005; 79 (10):6528–31. Epub 2005/04/29. PMID: 15858037

33. Liu Y, Curlin ME, Diem K, Zhao H, Ghosh AK, Zhu H, et al. Env length and N-linked glycosylation following transmission of human immunodeficiency virus Type 1 subtype B viruses. Virology. 2008; 374 (2):229–33. Epub 2008/03/04. doi: 10.1016/j.virol.2008.01.029 PMID: 18314154

34. Gnanakaran S, Bhattacharya T, Daniels M, Keele BF, Hraber PT, Lapedes AS, et al. Recurrent signature patterns in HIV-1 B clade envelope glycoproteins associated with either early or chronic infections. PLoS Pathog. 2011; 7(9):e1002209. Epub 2011/10/08. doi: 10.1371/journal.ppat.1002209 PMID: 21980282

35. Asmal M, Hellmann I, Liu W, Keele BF, Perelson AS, Bhattacharya T, et al. A signature in HIV-1 envelope leader peptide associated with transition from acute to chronic infection impacts envelope

processing and infectivity. PLoS One. 2011; 6(8):e23673. Epub 2011/08/31. doi: 10.1371/journal.pone.0023673 PMID: 21876761

36. Ochsenbauer C, Edmonds TG, Ding H, Keele BF, Decker J, Salazar MG, et al. Generation of transmitted/founder HIV-1 infectious molecular clones and characterization of their replication capacity in CD4 T lymphocytes and monocyte-derived macrophages. J Virol. 2011; 86(5):2715–28. Epub 2011/12/23. doi: 10.1128/JVI.06157-11 PMID: 22190722

37. Margolis L, Shattock R. Selective transmission of CCR5-utilizing HIV-1: the 'gatekeeper' problem resolved? Nat Rev Microbiol. 2006; 4(4):312–7. Epub 2006/03/17. PMID: 16541138

38. Fenton-May AE, Dibben O, Emmerich T, Ding H, Pfafferott K, Aasa-Chapman MM, et al. Relative resistance of HIV-1 founder viruses to control by interferon-alpha. Retrovirology. 2013; 10:146. Epub 2013/12/05. doi: 10.1186/1742-4690-10-146 PMID: 24299076

39. Etemad B, Gonzalez OA, White L, Laeyendecker O, Kirk GD, Mehta S, et al. Characterization of HIV-1 envelopes in acutely and chronically infected injection drug users. Retrovirology. 2014; 11(1):106. Epub 2014/11/29.

40. Deymier MJ, Ende Z, Fenton-May AE, Dilernia DA, Kilembe W, Allen SA, et al. Heterosexual Transmission of Subtype C HIV-1 Selects Consensus-Like Variants without Increased Replicative Capacity or Interferon-alpha Resistance. PLoS Pathog. 2015; 11(9):e1005154. Epub 2015/09/18. doi: 10.1371/journal.ppat.1005154 PMID: 26378795

41. Boeras DI, Hraber PT, Hurlston M, Evans-Strickfaden T, Bhattacharya T, Giorgi EE, et al. Role of donor genital tract HIV-1 diversity in the transmission bottleneck. Proc Natl Acad Sci U S A. 2011; 108 (46):E1156–63. Epub 2011/11/09. doi: 10.1073/pnas.1103764108 PMID: 22065783

42. Pena-Cruz V, Etemad B, Chatziandreou N, Nyein PH, Stock S, Reynolds SJ, et al. HIV-1 envelope replication and alpha4beta7 utilization among newly infected subjects and their corresponding heterosexual partners. Retrovirology. 2013; 10:162. Epub 2013/12/29. doi: 10.1186/1742-4690-10-162 PMID: 24369910

43. Sagar M, Laeyendecker O, Lee S, Gamiel J, Wawer MJ, Gray RH, et al. Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. J Infect Dis. 2009; 199(4):580–9. Epub 2009/01/16. doi: 10.1086/596557 PMID: 19143562

44. Herbeck JT, Rolland M, Liu Y, McLaughlin S, McNevin J, Zhao H, et al. Demographic processes affect HIV-1 evolution in primary infection before the onset of selective processes. J Virol. 2011; 85 (15):7523–34. Epub 2011/05/20. doi: 10.1128/JVI.02697-10 PMID: 21593162

45. Isaacman-Beck J, Hermann EA, Yi Y, Ratcliffe SJ, Mulenga J, Allen S, et al. Heterosexual transmission of human immunodeficiency virus type 1 subtype C: Macrophage tropism, alternative coreceptor use, and the molecular anatomy of CCR5 utilization. J Virol. 2009; 83(16):8208–20. Epub 2009/06/12. doi: 10.1128/JVI.00296-09 PMID: 19515785

46. Alexander M, Lynch R, Mulenga J, Allen S, Derdeyn CA, Hunter E. Donor and recipient envs from heterosexual human immunodeficiency virus subtype C transmission pairs require high receptor levels for entry. J Virol. 2010; 84(8):4100–4. Epub 2010/02/12. doi: 10.1128/JVI.02068-09 PMID: 20147398

47. Masharsky AE, Dukhovlinova EN, Verevochkin SV, Toussova OV, Skochilov RV, Anderson JA, et al. A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia. J Infect Dis. 2010; 201(11):1697–702. Epub 2010/04/29. doi: 10.1086/652702 PMID: 20423223

48. Gottlieb GS, Heath L, Nickle DC, Wong KG, Leach SE, Jacobs B, et al. HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. J Infect Dis. 2008; 197(7):1011–5. Epub 2008/04/19. doi: 10.1086/529206 PMID: 18419538

49. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, et al. Wide variation in the multiplicity of HIV-1 infection among injection drug users. J Virol. 2010; 84(12):6241–7. Epub 2010/04/09. doi: 10.1128/JVI.00077-10 PMID: 20375173

50. Dukhovlinova E, Masharsky A, Verevochkin S, Toussova O, Shevchenko A, Montefiori D, et al. AHI Detection Among People Who Inject Drugs in Russia Reveals the HIV-1 Transmission Bottleneck. Conference On Retroviruses And Opportunistic Infections (CROI 2014): International Antiviral Society—USA; 2014. p. 299.

51. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015; 32(3):820–32. Epub 2014/12/30. doi: 10.1093/molbev/msu400 PMID: 25540451

52. Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature. 1977; 267(5608):275–6. Epub 1977/05/19. PubMed PMID: 865622

53. Yang Z, Bielawski JP. Statistical methods for detecting molecular adaptation. Trends Ecol Evol. 2000; 15(12):496–503. Epub 2000/12/15. PMID: 11114436

54. Carlson JM, Brumme ZL, Rousseau CM, Brumme CJ, Matthews P, Kadie C, et al. Phylogenetic dependency networks: inferring patterns of CTL escape and codon covariation in HIV-1 Gag. PLoS Comput Biol. 2008; 4(11):e1000225. Epub 2008/11/22. doi: 10.1371/journal.pcbi.1000225 PMID: 19023406

55. Bhattacharya T, Daniels M, Heckerman D, Foley B, Frahm N, Kadie C, et al. Founder effects in the assessment of HIV polymorphisms and HLA allele associations. Science. 2007; 315(5818):1583–6. Epub 2007/03/17. PMID: 17363674

56. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003; 100(16):9440–5. Epub 2003/07/29. PMID: 12883005

57. Chen B, Vogan EM, Gong H, Skehel JJ, Wiley DC, Harrison SC. Structure of an unliganded simian immunodeficiency virus gp120 core. Nature. 2005; 433(7028):834–41. Epub 2005/02/25. PMID: 15729334

58. Bunnik EM, van Gils MJ, Lobbrecht MS, Pisas L, Nanlohy NM, van Baarle D, et al. Emergence of monoclonal antibody b12-resistant human immunodeficiency virus type 1 variants during natural infection in the absence of humoral or cellular immune pressure. J Gen Virol. 2010; 91(Pt 5):1354–64. Epub 2010/01/08. doi: 10.1099/vir.0.017319-0 PMID: 20053822

59. Bontjer I, Melchers M, Eggink D, David K, Moore JP, Berkhout B, et al. Stabilized HIV-1 envelope glycoprotein trimers lacking the V1V2 domain, obtained by virus evolution. J Biol Chem. 2010; 285 (47):36456–70. Epub 2010/09/10. doi: 10.1074/jbc.M110.156588 PMID: 20826824

60. Dunfee RL, Thomas ER, Gorry PR, Wang J, Taylor J, Kunstman K, et al. The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. Proc Natl Acad Sci U S A. 2006; 103(41):15160–5. Epub 2006/10/04. PMID: 17015824

61. Grupping K, Selhorst P, Michiels J, Vereecken K, Heyndrickx L, Kessler P, et al. MiniCD4 protein resistance mutations affect binding to the HIV-1 gp120 CD4 binding site and decrease entry efficiency. Retrovirology. 2012; 9:36. Epub 2012/05/04. doi: 10.1186/1742-4690-9-36 PMID: 22551420

62. Ince WL, Zhang L, Jiang Q, Arrildt K, Su L, Swanstrom R. Evolution of the HIV-1 env gene in the Rag2-/- gammaC-/- humanized mouse model. J Virol. 2010; 84(6):2740–52. Epub 2010/01/01. doi: 10.1128/JVI.02180-09 PMID: 20042504

63. Madani N, Perdigoto AL, Srinivasan K, Cox JM, Chruma JJ, LaLonde J, et al. Localized changes in the gp120 envelope glycoprotein confer resistance to human immunodeficiency virus entry inhibitors BMS-806 and #155. J Virol. 2004; 78(7):3742–52. Epub 2004/03/16. PMID: 15016894

64. Mo H, Stamatatos L, Ip JE, Barbas CF, Parren PW, Burton DR, et al. Human immunodeficiency virus type 1 mutants that escape neutralization by human monoclonal antibody IgG1b12. off. J Virol. 1997; 71(9):6869–74. Epub 1997/09/01. PMID: 9261412

65. Nakamura GR, Byrn R, Wilkes DM, Fox JA, Hobbs MR, Hastings R, et al. Strain specificity and binding affinity requirements of neutralizing monoclonal antibodies to the C4 domain of gp120 from human immunodeficiency virus type 1. J Virol. 1993; 67(10):6179–91. Epub 1993/10/01. PMID: 7690420

66. Pugach P, Kuhmann SE, Taylor J, Marozsan AJ, Snyder A, Ketas T, et al. The prolonged culture of human immunodeficiency virus type 1 in primary lymphocytes increases its sensitivity to neutralization by soluble CD4. Virology. 2004; 321(1):8–22. Epub 2004/03/23. PMID: 15033560

67. Shibata J, Yoshimura K, Honda A, Koito A, Murakami T, Matsushita S. Impact of V2 mutations on escape from a potent neutralizing anti-V3 monoclonal antibody during in vitro selection of a primary human immunodeficiency virus type 1 isolate. J Virol. 2007; 81(8):3757–68. Epub 2007/01/26. PMID: 17251298

68. Sterjovski J, Churchill MJ, Roche M, Ellett A, Farrugia W, Wesselingh SL, et al. CD4-binding site alterations in CCR5-using HIV-1 envelopes influencing gp120-CD4 interactions and fusogenicity. Virology. 2011; 410(2):418–28. Epub 2011/01/11. doi: 10.1016/j.virol.2010.12.010 PMID: 21216423

69. Wrin T, Loh TP, Vennari JC, Schuitemaker H, Nunberg JH. Adaptation to persistent growth in the H9 cell line renders a primary isolate of human immunodeficiency virus type 1 sensitive to neutralization by vaccine sera. J Virol. 1995; 69(1):39–48. Epub 1995/01/01. PMID: 7983734

70. Skums P, Mancuso N, Artyomenko A, Tork B, Mandoiu I, Khudyakov Y, et al. Reconstruction of viral population structure from next-generation sequencing data using multicommodity flows. BMC Bioinformatics. 2013; 14 Suppl 9:S2. Epub 2013/08/09. doi: 10.1186/1471-2105-14-S9-S2 PMID: 23902469

71. Giallonardo FD, Topfer A, Rey M, Prabhakaran S, Duport Y, Leemann C, et al. Full-length haplotype reconstruction to infer the structure of heterogeneous virus populations. Nucleic Acids Res. 2014; 42 (14):e115. Epub 2014/06/29. doi: 10.1093/nar/gku537 PMID: 24972832

72. Zagordi O, Geyrhofer L, Roth V, Beerenwinkel N. Deep sequencing of a genetically heterogeneous sample: local haplotype reconstruction and read error correction. J Comput Biol. 2010; 17(3):417–28. Epub 2010/04/10. doi: 10.1089/cmb.2009.0164 PMID: 20377454

73. Zagordi O, Klein R, Daumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Res. 2010; 38(21):7400–9. Epub 2010/07/31. doi: 10.1093/nar/gkq655 PMID: 20671025

74. Prosperi MC, Salemi M. QuRe: software for viral quasispecies reconstruction from next-generation sequencing data. Bioinformatics. 2011; 28(1):132–3. Epub 2011/11/18. doi: 10.1093/bioinformatics/btr627 PMID: 22088846

75. Pacold M, Smith D, Little S, Cheng PM, Jordan P, Ignacio C, et al. Comparison of methods to detect HIV dual infection. AIDS Res Hum Retroviruses. 2010; 26(12):1291–8. Epub 2010/10/20. doi: 10.1089/aid.2010.0042 PMID: 20954840

76. Schirmer M, Sloan WT, Quince C. Benchmarking of viral haplotype reconstruction programmes: an overview of the capacities and limitations of currently available programmes. Brief Bioinform. 2012; 15(3):431–42. Epub 2012/12/22. doi: 10.1093/bib/bbs081 PMID: 23257116

77. Henn MR, Boutwell CL, Charlebois P, Lennon NJ, Power KA, Macalalad AR, 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. Epub 2012/03/14. doi: 10.1371/journal.ppat.1002529 PMID: 22412369

78. Rolland M, Tovanabutra S, deCamp AC, Frahm N, Gilbert PB, Sanders-Buell E, et al. Genetic impact of vaccination on breakthrough HIV-1 sequences from the STEP trial. Nat Med. 2011; 17(3):366–71. Epub 2011/03/02. doi: 10.1038/nm.2316 PMID: 21358627

79. Wagner GA, Pacold ME, Kosakovsky Pond SL, Caballero G, Chaillon A, Rudolph AE, et al. Incidence and prevalence of intrasubtype HIV-1 dual infection in at-risk men in the United States. J Infect Dis. 2013; 209(7):1032–8. Epub 2013/11/26. doi: 10.1093/infdis/jit633 PMID: 24273040

80. Rieder P, Joos B, Scherrer AU, Kuster H, Braun D, Grube C, et al. Characterization of human immunodeficiency virus type 1 (HIV-1) diversity and tropism in 145 patients with primary HIV-1 infection. Clin Infect Dis. 2011; 53(12):1271–9. Epub 2011/10/15. doi: 10.1093/cid/cir725 PMID: 21998286

81. Giorgi EE, Funkhouser B, Athreya G, Perelson AS, Korber BT, Bhattacharya T. Estimating time since infection in early homogeneous HIV-1 samples using a poisson model. BMC Bioinformatics. 2010; 11:532. Epub 2010/10/27. doi: 10.1186/1471-2105-11-532 PMID: 20973976

82. Joseph SB, Swanstrom R, Kashuba AD, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. Nat Rev Microbiol. 2015; 13(7):414–25. Epub 2015/06/09. doi: 10.1038/nrmicro3471 PMID: 26052661

83. Ma ZM, Keele BF, Qureshi H, Stone M, Desilva V, Fritts L, et al. SIVmac251 is inefficiently transmitted to rhesus macaques by penile inoculation with a single SIVenv variant found in ramp-up phase plasma. AIDS Res Hum Retroviruses. 2011; 27(12):1259–69. Epub 2011/07/08. doi: 10.1089/aid.2011.0090 PMID: 21732792

84. Fieni F, Stone M, Ma ZM, Dutra J, Fritts L, Miller CJ. Viral RNA levels and env variants in semen and tissues of mature male rhesus macaques infected with SIV by penile inoculation. PLoS One. 2013; 8(10):e76367. Epub 2013/10/23. doi: 10.1371/journal.pone.0076367 PMID: 24146859

85. Keele BF, Li H, Learn GH, Hraber P, Giorgi EE, Grayson T, et al. Low-dose rectal inoculation of rhesus macaques by SIVsmE660 or SIVmac251 recapitulates human mucosal infection by HIV-1. J Exp Med. 2009; 206(5):1117–34. Epub 2009/05/06. doi: 10.1084/jem.20082831 PMID: 19414559

86. Dinh MH, Anderson MR, McRaven MD, Cianci GC, McCoombe SG, Kelley ZL, et al. Visualization of HIV-1 interactions with penile and foreskin epithelia: clues for female-to-male HIV transmission. PLoS Pathog. 2015; 11(3):e1004729. Epub 2015/03/10. doi: 10.1371/journal.ppat.1004729 PMID: 25748093

87. Poles MA, Elliott J, Taing P, Anton PA, Chen IS. A preponderance of CCR5(+) CXCR4(+) mononuclear cells enhances gastrointestinal mucosal susceptibility to human immunodeficiency virus type 1 infection. J Virol. 2001; 75(18):8390–9. Epub 2001/08/17. PMID: 11507184

88. McElrath MJ, Smythe K, Randolph-Habecker J, Melton KR, Goodpaster TA, Hughes SM, et al. Comprehensive assessment of HIV target cells in the distal human gut suggests increasing HIV susceptibility toward the anus. J Acquir Immune Defic Syndr. 2013; 63(3):263–71. Epub 2013/02/09. doi: 10.1097/QAI.0b013e3182898392 PMID: 23392465

89. Ribeiro Dos Santos P, Rancez M, Pretet JL, Michel-Salzat A, Messent V, Bogdanova A, et al. Rapid dissemination of SIV follows multisite entry after rectal inoculation. PLoS One. 2011; 6(5):e19493. Epub 2011/05/17. doi: 10.1371/journal.pone.0019493 PMID: 21573012

90. Miyake A, Ibuki K, Enose Y, Suzuki H, Horiuchi R, Motohara M, et al. Rapid dissemination of a pathogenic simian/human immunodeficiency virus to systemic organs and active replication in lymphoid tissues following intrarectal infection. J Gen Virol. 2006; 87(Pt 5):1311–20. Epub 2006/04/11. PMID: 16603534

91. Jacobs A, Sen J, Rong L, Caffrey M. Alanine scanning mutants of the HIV gp41 loop. J Biol Chem. 2005; 280(29):27284–8. Epub 2005/05/27. PMID: 15917239

92. Postler TS, Desrosiers RC. The tale of the long tail: the cytoplasmic domain of HIV-1 gp41. J Virol. 2013; 87(1):2–15. Epub 2012/10/19. doi: 10.1128/JVI.02053-12 PMID: 23077317

93. Eisenberg D, Wesson M. The most highly amphiphilic alpha-helices include two amino acid segments in human immunodeficiency virus glycoprotein 41. Biopolymers. 1990; 29(1):171–7. Epub 1990/01/01. PMID: 2328285

94. Venable RM, Pastor RW, Brooks BR, Carson FW. Theoretically determined three-dimensional structures for amphipathic segments of the HIV-1 gp41 envelope protein. AIDS Res Hum Retroviruses. 1989; 5(1):7–22. Epub 1989/02/01. PMID: 2541749

95. Kliger Y, Shai Y. A leucine zipper-like sequence from the cytoplasmic tail of the HIV-1 envelope glycoprotein binds and perturbs lipid bilayers. Biochemistry. 1997; 36(17):5157–69. Epub 1997/04/29. PMID: 9136877

96. Kalia V, Sarkar S, Gupta P, Montelaro RC. Rational site-directed mutations of the LLP-1 and LLP-2 lentivirus lytic peptide domains in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41 indicate common functions in cell-cell fusion but distinct roles in virion envelope incorporation. J Virol. 2003; 77(6):3634–46. Epub 2003/03/01. PMID: 12610139

97. Piller SC, Dubay JW, Derdeyn CA, Hunter E. Mutational analysis of conserved domains within the cytoplasmic tail of gp41 from human immunodeficiency virus type 1: effects on glycoprotein incorporation and infectivity. J Virol. 2000; 74(24):11717–23. Epub 2000/11/23. PMID: 11090171

98. Sterjovski J, Churchill MJ, Ellett A, Gray LR, Roche MJ, Dunfee RL, et al. Asn 362 in gp120 contributes to enhanced fusogenicity by CCR5-restricted HIV-1 envelope glycoprotein variants from patients with AIDS. Retrovirology. 2007; 4:89. Epub 2007/12/14. PMID: 18076768

99. Watkins BA, Reitz MS Jr., Wilson CA, Aldrich K, Davis AE, Robert-Guroff M. Immune escape by human immunodeficiency virus type 1 from neutralizing antibodies: evidence for multiple pathways. J Virol. 1993; 67(12):7493–500. Epub 1993/12/01. PMID: 7693973

100. Gardner MR, Kattenhorn LM, Kondur HR, von Schaewen M, Dorfman T, Chiang JJ, et al. AAV-expressed eCD4-Ig provides durable protection from multiple SHIV challenges. Nature. 2015; 519 (7541):87–91. Epub 2015/02/25. doi: 10.1038/nature14264 PMID: 25707797

101. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med. 2010; 363(27):2587–99. doi: 10.1056/NEJMoa1011205 PMID: 21091279

102. Garrett NJ, Werner L, Naicker N, Naranbhai V, Sibeko S, Samsunder N, et al. HIV disease progression in seroconvertors from the CAPRISA 004 tenofovir gel pre-exposure prophylaxis trial. J Acquir Immune Defic Syndr. 2015; 68(1):55–61. Epub 2014/09/24. doi: 10.1097/QAI.0000000000000367 PMID: 25247433

103. Pacold ME, Pond SL, Wagner GA, Delport W, Bourque DL, Richman DD, et al. Clinical, virologic, and immunologic correlates of HIV-1 intraclade B dual infection among men who have sex with men. AIDS. 2011; 26(2):157–65. Epub 2011/11/03.

104. Sagar M, Lavreys L, Baeten JM, Richardson BA, Mandaliya K, Chohan BH, et al. Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. J Virol. 2003; 77(23):12921–6. Epub 2003/11/12. PMID: 14610215

105. Tsai L, Tasovski I, Leda AR, Chin MP, Cheng-Mayer C. The number and genetic relatedness of transmitted/founder virus impact clinical outcome in vaginal R5 SHIVSF162P3N infection. Retrovirology. 2014; 11:22. Epub 2014/03/13. doi: 10.1186/1742-4690-11-22 PMID: 24612462

106. Gottlieb GS, Nickle DC, Jensen MA, Wong KG, Grobler J, Li F, et al. Dual HIV-1 infection associated with rapid disease progression. Lancet. 2004; 363(9409):619–22. Epub 2004/02/28. PMID: 14987889

107. Cornelissen M, Pasternak AO, Grijsen ML, Zorgdrager F, Bakker M, Blom P, et al. HIV-1 dual infection is associated with faster CD4+ T-cell decline in a cohort of men with primary HIV infection. Clin Infect Dis. 2012; 54(4):539–47. Epub 2011/12/14. doi: 10.1093/cid/cir849 PMID: 22157174

108. Burton DR, Hessell AJ, Keele BF, Klasse PJ, Ketas TA, Moldt B, et al. Limited or no protection by weakly or nonneutralizing antibodies against vaginal SHIV challenge of macaques compared with a strongly neutralizing antibody. Proc Natl Acad Sci U S A. 2011; 108(27):11181–6. Epub 2011/06/22. doi: 10.1073/pnas.1103012108 PMID: 21690411

109. Santra S, Tomaras GD, Warrier R, Nicely NI, Liao HX, Pollara J, et al. Human Non-neutralizing HIV-1 Envelope Monoclonal Antibodies Limit the Number of Founder Viruses during SHIV Mucosal Infection in Rhesus Macaques. PLoS Pathog. 2015; 11(8):e1005042. Epub 2015/08/04. doi: 10.1371/journal.ppat.1005042 PMID: 26237403

110. Macalalad AR, Zody MC, Charlebois P, Lennon NJ, Newman RM, Malboeuf CM, et al. Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. PLoS Comput Biol. 2012; 8(3):e1002417. Epub 2012/03/23. doi: 10.1371/journal.pcbi.1002417 PMID: 22438797

111. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. Bioinformatics. 2012; 28(4):593–4. Epub 2011/12/27. doi: 10.1093/bioinformatics/btr708 PMID: 22199392

112. Palmer S, Kearney M, Maldarelli F, Halvas EK, Bixby CJ, Bazmi H, et al. Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J Clin Microbiol. 2005; 43(1):406–13. Epub 2005/01/07. PMID: 15635002

113. 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(13):1871–9. Epub 2003/09/10. PMID: 12960819

114. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59(3):307–21. Epub 2010/06/09. doi: 10.1093/sysbio/syq010 PMID: 20525638

115. Carlson JM, Listgarten J, Pfeifer N, Tan V, Kadie C, Walker BD, et al. Widespread impact of HLA restriction on immune control and escape pathways of HIV-1. J Virol. 2012; 86(9):5230–43. Epub 2012/03/02. doi: 10.1128/JVI.06728-11 PMID: 22379086