HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders

Holly Janes1, Joshua T Herbeck2, Sodsai Tovanabutra3,4, Rasmi Thomas3,4, Nicole Frahm1,2, Ann Duerr1, John Hural1, Lawrence Corey1, Steve G Self1, Susan P Buchbinder1, Nelson L Michael3, James I Mullins10, Jerome H Kim3, Peter B Gilbert1 & Morgane Rolland3,4

Given the variation in the HIV-1 viral load (VL) set point across subjects, as opposed to a fairly stable VL over time within an infected individual, it is important to identify the characteristics of the host and virus that affect VL set point. Although recently infected individuals with multiple phylogenetically linked HIV-1 founder variants represent a minority of HIV-1 infections, we found—in two different cohorts—that more diverse HIV-1 populations in early infection were associated with significantly higher VL 1 year after HIV-1 diagnosis.

Approximately 20–35% of individuals become infected with multiple founder HIV-1 variants1–3. We therefore sought to evaluate whether genetic characteristics of the founder viral populations could influence markers of clinical outcomes. An association between infections with multiple HIV-1 variants and faster disease progression than the average disease course is supported by previous reports based on heteroduplex mobility assays4 and dual HIV infections5–7. More recently, large studies have sought to derive HIV-1 sequences through single-genome amplification (SGA) from samples collected in acute or early HIV-1 infection to better define the viruses that are establishing the HIV-1 infection; as such, enumerating the number of HIV-1 founder variants has been a hallmark of these studies. The availability of larger, more precise sequence data sets prompted us to test the association between HIV-1 diversity and markers of disease progression using SGA-derived HIV-1 genomic data.

We focused on HIV-1 breakthrough infections in the Step and RV144 HIV-1 vaccine efficacy trials (median of six and ten genomes, respectively)8–11, and restricted our analysis to 63 Step trial participants (infected with HIV-1 subtype B) and 100 RV144 trial participants (infected with the subtype CRF01_AE) who had VL and CD4+ T cell measurements in the presence of antiretroviral therapy (ART). In both trials, HIV-1 infections were established by a single viral variant in most individuals, with no significant difference in proportions between treatment groups (P ≥ 0.81) (refs. 10,11).

We used two measures of diversity for HIV-1 founder populations: a categorical measure that distinguishes between subjects with a single founder variant (homogeneous viral populations) or multiple founder variants (heterogeneous viral populations), and a continuous measure of diversity in the envelope gene (envelope), which corresponds to the mean pairwise diversity among sequences from a subject. Regarding the 63 Step study participants, the median diversity was 0.073% (0–0.566%) among the 47 subjects with homogeneous viral populations and 0.593% (0.026–5.98%) for those with heterogeneous populations. For the 100 RV144 participants, median diversity was 0.194% (0.027–0.847%) among the 68 participants with homogeneous founding populations and 0.825% (0.073–4.42%) for those with heterogeneous populations.

Because relevant variables and availability of baseline variables and VL and CD4+ T cell measurements differed between the trials, data were analyzed separately (as previously described8,9,12,13). Linear regression models were used to relate each diversity measure to post-infection endpoints while accounting for baseline subject characteristics. Besides the treatment assignment (vaccine or placebo), statistical models were adjusted for multiple covariates, including sex, human leukocyte antigen (HLA) genotype (Step only) and a baseline behavioral risk score. Here we present fully adjusted results (see Supplementary Tables 1–3 for unadjusted and partially adjusted results).

In the Step study, there was no association between sequence diversity and VL at the time of HIV-1 diagnosis, neither when homogeneous and heterogeneous viral populations were compared (P = 0.88), nor when env diversity measures were considered (P = 0.37) (Supplementary Fig. 1 and Supplementary Table 1). When we considered the 276 pre-ART VL measurements obtained during the first year of follow up (median = 5 (range: 1–7) per subject), subjects with heterogeneous founder viral populations showed significantly higher mean VL than did subjects with homogeneous founder populations: the estimated difference was 0.37 log10 VL (copies per ml), on the basis of analyses that were adjusted for baseline subject characteristics in addition to time since infection diagnosis (P = 0.01) (Fig. 1). Similarly to the categorical data, higher mean VL over the first year was associated with higher env diversity of the founder population (P < 0.001). The relationship was nonlinear, such that there was no...
BRIEF COMMUNICATIONS

Figure 1 The positive relationship between HIV-1 diversity and viral loads over the first year after HIV-1 diagnosis. HIV-1 diversity was measured at the time of HIV-1 diagnosis. Individual log VL values, predicted means and 95% confidence intervals are shown at each post-infection visit. (a,b) Step (a) and RV144 (b) trial participants with homogeneous or heterogeneous viral populations. (c,d) Low, medium and high levels of env diversity in Step (c) and RV144 (d) trial participants. Models were adjusted for baseline subject characteristics and assumed that the effects of heterogeneity and diversity are constant over time.

Figure 2 The negative relationship between HIV-1 diversity and CD4+ T cell counts over the first year after HIV-1 diagnosis. HIV-1 diversity was measured at the time of HIV-1 diagnosis. Individual CD4+ T cell count measurements and predicted means and 95% confidence intervals are shown at each post-infection visit. (a,b) Step (a) and RV144 (b) trial participants with homogeneous or heterogeneous viral populations. (c,d) Low, medium and high levels of env diversity in Step (c) and RV144 (d) trial participants. Models were adjusted for baseline subject characteristics and assumed that the effects of heterogeneity and diversity are constant over time.
mean square root CD4\(^+\) T cell count over the first year of infection compared to subjects with homogeneous ones \((P = 0.02; \text{Fig. 2 and Supplementary Table 1})\). When env diversity was considered, increasing env diversity was found to be significantly associated with decreasing mean square root CD4\(^+\) T cell counts \((P = 0.03)\). The relationship was nonlinear: it was not significant below the cutpoint of 0.1% env diversity \((P = 0.51)\), and there was a negative association above the cutpoint \((P = 0.02)\). These results suggest that greater heterogeneity in the HIV-1 founder population of recently infected individuals is associated with a higher VL over time, confirming an earlier heteroduplex mobility assay-based report\(^4\). When VL measurements obtained up to 2 years after diagnosis were included, the sizes of the effects were smaller but still significant or trending \((P < 0.001–0.08)\). Study subjects were vaccine or placebo recipients in the Step and RV144 vaccine trials, yet there was no evidence that the vaccine assignment modified the associations \((q > 0.3, \text{accounting for multiplicity in interaction tests across endpoints})\). To address the possibility of post-randomization selection bias—whereby breakthrough vaccine and placebo cases of infection are not comparable, owing to differences in characteristics associated with both HIV-1 infection and post-infection endpoints—we adjusted for covariates that are potentially predictive of either HIV-1 acquisition or disease progression, such as HLA (for Step) and baseline behavioral risk, and we found that including these variables had a negligible effect on the results. We also performed sensitivity analyses to show that our sequencing protocol and the lack of precision in the timing of HIV-1 infection dates in both cohorts did not noticeably affect our results (Supplementary Tables 2 and 3; Supplementary Note). The limited depth of sequencing is another potential limitation; pyrosequencing data, however, confirmed our estimates of founder variants for 48 of the 63 subjects (selected on the basis of sample availability) in the Step study (S. Iyer and J.I.M., University of Washington, personal communication).

Some results differed between the two cohorts. In the Step study, the positive association between VL and multiplicity of founder variants was seen at later time points but not at HIV-1 diagnosis (note that infections were diagnosed earlier in Step than in RV144 (ref. 14)). A similar lack of association between env diversity at a median of 6 weeks after infection and contemporary VL was reported previously\(^5\). This supports the view that VL set point is not a characteristic of the founder virus per se, and highlights the importance of the relationship between the host and the virus in the establishment of the VL set point. In addition, we failed to find an association between CD4\(^+\) T cell counts and multiplicity of founder variants in the Step study. To interpret this result, we note that the relationship between CD4\(^+\) T cell count and VL was stronger in RV144 (Spearman rank correlation between VL and CD4\(^+\) T cell counts residuals = −0.52) than in the Step study (−0.44). It has previously been noted that the link between the two predictors of virulence is not always strong\(^16\).

Given that different studies found that the viral genotype contributed to VL set point to different extents, ranging from 6% (ref. 17) to 59% (ref. 18), a better characterization of the viral factors that potentially impact VL set point is critical to our understanding of HIV-1 pathogenesis. Although the size of the HIV-1 genetic effect estimated in our analyses can be considered modest \((0.29–0.37 \log_{10})\), our findings have been replicated in two independent cohorts with different distributions of subject ethnicity, route of HIV-1 transmission and infecting HIV-1 subtype. Moreover, a difference of 0.3 \log_{10} in VL set point is clinically relevant to both disease progression\(^19\) and HIV-1 transmission; a decrease in VL of 0.74 \log_{10} \((95\% \text{ CI } 0.60–0.97)\) was estimated to reduce by 50% the risk of virus transmission in heterosexuals\(^20\). Further studies are needed to define host specificities that predict an individual's propensity to acquire a multi-variant HIV-1 infection. Although our study cannot be used to determine whether certain individuals are predisposed to acquire multiple HIV-1 variants, the fact that individuals replicating multiple HIV-1 variants presented higher VL set points illustrates the consequences of the initial steps of HIV-1 infection for clinical disease progression, and it suggests that limiting HIV-1 founder heterogeneity could be a goal for prophylactic interventions.

**METHODS**

Methods and any associated references are available in the online version of the paper.

**Note** Any Supplementary Information and Source Data files are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We would like to thank the participants in the Step study and the RV144 trial for their contributions to this research and Lisa Reilly for editorial help. Sequencing and analysis were performed under grants from the National Institute of Allergy and Infectious Diseases: US Public Health Service grant AI41955; Interagency Agreement Y1-AI-2642-12 with the US Army Medical Research and Material Command; US National Institutes of Health (NIH) grant 2R37AI05465-10 to P.B.G. This work was also supported by a cooperative agreement (W81XWH-07-2-0067) between the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., and the US Department of Defense. The opinions expressed herein are those of the authors and should not be construed as official or representing the views of the US Department of Defense or the Department of the Army. This does not alter our adherence to policies on sharing data and materials.

**AUTHOR CONTRIBUTIONS**

H.J. and M.R. designed and performed experiments, analyzed data and wrote the manuscript. S.T., A.D., J.T.H., L.C., S.G.S., S.P.B., M.J.M., R.I.O., R.M.P., S.R.-N., S.N., P.P., J.K., M.L.R., N.I.M., J.H.K. and P.B.G. oversaw the vaccine trials and clinical aspects. J.T.H., R.T., N.F., M.J.M., M.L.R., N.I.M., J.I.M., J.H.K. and P.B.G. edited the manuscript.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Keele, B.F. et al. Proc. Natl. Acad. Sci. USA 105, 7552–7557 (2008).
2. Abrahams, M.R. et al. J. Virol. 83, 3556–3567 (2009).
3. Li, H. et al. PLoS Pathog. 6, e1000890 (2010).
4. Sagar, M. et al. J. Virol. 77, 12921–12926 (2003).
5. Gottlieb, G.S. et al. Nature 363, 619–622 (2004).
6. Jost, S. et al. HIV super-infection: rapid replacement of AE subtype by B subtype. Poster presented at the 9th Conference on Retroviruses and Opportunistic Infections, Abstract no.: 757-W (Seattle, Washington, 2002).
7. Afifeld, M. et al. Nature 420, 434–439 (2002).
8. Buchbinder, S.P. et al. Lancet 372, 1881–1893 (2008).
9. Rekers-Ngarm, S. et al. N. Engl. J. Med. 361, 2209–2220 (2009).
10. Rolland, M. et al. Nat. Med. 17, 366–371 (2011).
11. Rolland, M. et al. Nature 490, 417–420 (2012).
12. Janes, H. et al. PLoS ONE 7, e43396 (2012).
13. Rekers-Ngarm, S. et al. J. Infect. Dis. 207, 1195–1205 (2013).
14. Edelstein, P.T., Gilbert, P.B. & Rolland, M. Curr. Opin. HIV AIDS 8, 432–436 (2013).
15. Rieder, P. et al. Clin. Infect. Dis. 53, 1271–1279 (2011).
16. Rieder, P. et al. J. Am. Med. Assoc. 296, 1498–1506 (2006).
17. Hodorov, E. et al. PLoS Pathog. 10, e1004112 (2014).
18. Alizon, S. et al. PLoS Pathog. 6, e1001123 (2010).
19. Melors, J.W. et al. Ann. Intern. Med. 126, 946–954 (1997).
20. Lingappa, J.R. et al. PLoS ONE 5, e12998 (2010).
ONLINE METHODS

Ethics statement. The Step study (HVTN502) and RV144 vaccine trials were registered at ClinicalTrials.gov and assigned the registration numbers NCT00095576, and NCT00223080, respectively. The RV144 protocol was approved by the ethics committees of the Ministry of Public Health, the Royal Thai Army, Mahidol University and the Human Subjects Research Review Board of the US Army Medical Research and Materiel Command. The Step protocol was approved by the ethics committees of each trial site (Sydney, Australia; Rio de Janeiro, Brazil; Sao Paulo, Brazil; Montreal, Canada; Toronto, Canada; Vancouver, Canada; Santo Domingo, Dominican Republic; Port-au-Prince, Haiti; Kingston, Jamaica; Iquitos, Peru; Lima, Peru; San Juan, Puerto Rico; United States: Atlanta, GA; Birmingham, AL; Boston, MA; Chicago, IL; Denver, CO; Houston, TX; Los Angeles, CA; Miami, FL; New York, NY; Newark, NJ; Philadelphia, PA; Rochester, NY; St. Louis, MO; San Francisco, CA; Seattle, WA). Written informed consent was obtained from all volunteers.

HIV-1 genetic data. Breakthrough infections were sequenced using endpoint dilution PCR from plasma samples collected at the time of HIV-1 diagnosis (except for one Step study participant who was sampled 28 d after diagnosis and six RV144 participants who were sampled an average of 26 d later), with 5–10 near full-length genome sequences per individual10,11. Breakthrough HIV-1 sequences were obtained from 68 Step trial participants and 121 RV144 trial participants. Exclusion criteria for our study were: lack of availability of VL and CD4+ T cell measurements; lack of sequence data; being on ART; not being infected by the prevalent HIV-1 subtype; being female (in Step only). Thus, the cohorts included 63 male participants from the Step study who were infected with HIV-1 subtype B (excluded subjects included one female, one non-subtype B infection, one subject on ART at diagnosis and two subjects with only one sequence available), and 100 RV144 participants infected with CRF01_AE (114 subjects were enrolled in the post-infection follow-up study, three subjects who lacked HIV-1 sequence data and 11 who were infected with non-CRF01_AE viruses were excluded).

For each intra-host data set, inspection of sequence alignments, phylogenetic tree topologies and sequence diversity measures were used to categorize infections as either established by a single founder variant (homogeneous viral population) or multiple founder variants (heterogeneous viral population).

The primary variables used to measure the multiplicity of HIV-1 founder variants were an indicator of homogeneous or heterogeneous infection and the mean pairwise nucleotide diversity in env (mean percent distance between all pairs of sequences for a subject, calculated using the general time-reversible model of nucleotide substitution in HyPhy21, a phylogenetic analysis software). Mean pairwise diversity was analyzed on the log10 scale, with values of 0 set to 0.0065, the midpoint between the lowest positive value of 0.013 and 0. Piecewise linear splines for log mean pairwise sequence diversity (hereafter ‘env diversity’) were considered as candidate predictors. Where the data did not support the inclusion of piecewise-linear terms, models included only the linear terms.

Clinical data. Pre-ART VL measurements from the first year of follow up were obtained at weeks 0, 1, 2, 8, 12, 26, 52, 78 and 104 in the Step trial, and at months <1, 1, 3, 6, 9, 12, 18 and 24 in RV144. Longitudinal CD4+ T cell counts were measured at the same post-infection visits (except for the diagnostic visit in Step).

Statistical methods. The VL endpoints were log10 VL at the time of HIV-1 diagnosis (which corresponded to the time of HIV sequencing), and longitudinal VL; both were censored upon ART initiation. Longitudinal pre-ART CD4+ T cell counts were analyzed on the square root scale. Linear regression models were used to relate each predictor (homogeneous or heterogeneous infection; env diversity) to each post-infection endpoint. Wald tests were used to test for statistical interactions between each predictor and the vaccine or placebo assignment. Q values were calculated to account for multiplicity in interaction tests across endpoints; we considered Q < 0.20 to be significant, implying that up to 20% of the significant interaction results are expected to be false positives (see results in Supplementary Note).

For the Step study analyses of longitudinal VL and CD4+ T cell counts, weighted generalized estimating equations (GEE) models with exchangeable working correlations were used. GEE models account for the correlation between longitudinal measurements on the same subject. Observations were weighted with respect to the inverse probability of having an observed pre-ART VL and CD4+ T cell count measurement, which was factored as the product of: 1) the probability of not dropping out; 2) the probability of not initiating ART, given not dropping out; and 3) the probability of not missing the visit, given not dropping out or initiating ART. Each probability was modeled using a separate logistic GEE regression model with independence working correlation. Weeks after infection, year of diagnosis, geographic region, age and an indicator of all three vaccinations received were used to predict dropout; weeks post-infection, geographic region, Ad5 seropositivity, an indicator of all three vaccinations received and previous VL were used to predict ART initiation; and geographic region and previous VL and CD4+ T cell count were used to predict missing visits. The following covariates were included as specified: treatment assignment, Ad5 seropositivity, self-reported circumcision status, HSV-2 serostatus, age under or over 30 years, Australian or North American residence, HLA group and baseline behavioral risk score. HLA group was defined as ‘protective’ (B27, B57, *B*58:01), unfavorable (B*35:02, *35:03, *35:04, B53 or homozygous in at least one locus), or ‘neutral’ (all others) (a protective allele assigns a subject to the protective HLA group, irrespective of the other alleles present). The baseline risk score is defined as the number of risk behaviors reported at baseline among the following: more than two male partners, any drug use, unprotected anal sex with a male partner, unprotected vaginal sex with a female partner, evidence of a sexually transmitted disease or exchanging sex for money, all over the course of the 6 months prior to enrollment. Two individuals from North America without circumcision status were considered circumcision. Where specified, models were also adjusted for linear terms of days after diagnosis.

For RV144 analyses of longitudinal VL and CD4+ T cell counts, multiple imputation was used to fill in missing values, and GEE was used for modeling with AR1 working correlation structure. The following covariates were included where specified: six centered polynomial terms of days since infection diagnosis, calendar period of infection diagnosis (2003–05; 2006; 2007; 2008–09), age, gender and baseline behavioral risk (low, medium or high). The VL model was not adjusted for time-dependent CD4+ T cell count, nor were CD4+ T cell counts adjusted for VL; this was a departure from previous analyses of the VL and CD4+ data13. Centered polynomial terms of days since infection diagnosis were included: one polynomial (linear) term was used for months 0–3, three polynomial terms for analyses of months 0–12 and four polynomial terms for analyses of months 0–24. The baseline risk score depended on the following risk factors: number of partners, unprotected sex with a regular or casual sex partner, with a sex worker, with a same sex partner or with an injecting drug user, injection drug use, sharing needles, symptoms of a sexually transmitted disease. No HLA adjustment was done in the RV144 cohort because there is limited information about which HLA alleles are protective, unfavorable or neutral in the Thai CRF01_AE–infected population. Our comparison of the genotypes from 74 placebo recipients who became infected in the RV144 trial to the genotypes of 450 placebo recipients22 who remained uninfected did not identify any HLA alleles with frequencies greater than 5% that were associated with HIV-1 acquisition (Fisher’s exact test; P ≥ 0.05; q > 0.20).

21. Pond, S.L., Frost, S.D. & Muse, S.V. Bioinformatics 21, 676–679 (2005).
22. Prentice, H.A. et al. Immunogenetics 66, 299–310 (2014).