Elevated Viral Load is Associated with Higher Viral Recombination Rate In Vivo

Elena V. Romero, Alison F. Feder
University of Washington, Seattle, WA
evromero@uw.edu

Background

- Intra-host evolution permits HIV to escape the immune system and the antiretroviral therapies used to treat it.
- A major driver of diversification is HIV's high recombination rate – estimated to be on the order of 10^{-5}/bp/generation[3].
- HIV recombines only after two viruses infect the same cell (Fig. 1), meaning that population density could modulate the rate of recombination.

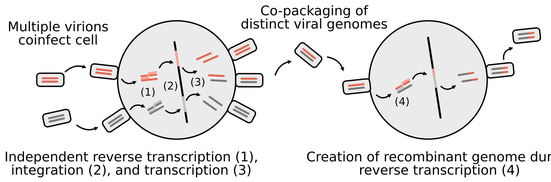


Figure 1: Diagram of HIV Recombination.

- However, HIV's intra-host demography is complex; viral loads (a proxy for viral population size) vary by orders of magnitude over time and between hosts. Viral density (and subsequent recombination) may therefore vary as well.
- We do not understand the relationship between viral load and recombination rate in vivo. As a result, we may be over- or underestimating the impact of recombination across patients with different viral loads, or within a single patient across timepoints.

Hypothesis: High viral loads lead to elevated recombination rates

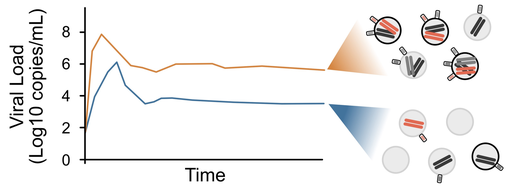


Figure 2: Hypothesis: viral load impacts recombination rate through mediating the frequency of viral coinfection.

Quantifying recombination using time series linkage decay

Because intra-host populations are frequently measured longitudinally, we can exploit the decay in genetic linkage over time to estimate the intra-host recombination rate (ρ) separately in people living with HIV (PLWH) with variable viral loads. We quantify this linkage by using the D' statistic[2].

First, we calculate D' between all pairs of loci separately at each time point. We expect D' to decay according to the distance between genetic loci and the amount of time that has passed. The decay of D' is described by

$$D'(t+\Delta t) = D'(t)e^{-\rho d\Delta t}$$

Then, we can estimate ρ by looking at this decay relationship across many pairs of loci and fitting the relationship at low values of distance x time ($d\Delta t$). This process is illustrated in Figure 3.

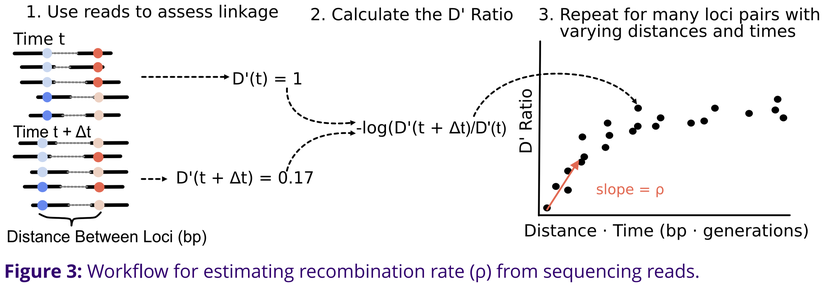


Figure 3: Workflow for estimating recombination rate (ρ) from sequencing reads.

Method accurately distinguishes simulated recombination rates

We used SLiM3[1] to simulate data from neutrally evolving populations with an effective population size of 1×10^4 HIV genomes.

We sampled 800 genomes from these simulations and then down-sampled the number of segregating loci to be 100. This was in line with the amount of in vivo data available for each individual and sequencing fragment.

We simulated populations with recombination rates (ρ) surrounding previous estimates[3], which were on the order of 10^{-5} . The saturation patterns of D' ratio curves vary alongside ρ as expected (Fig. 4A). We estimated recombination rates in this data and were able to recover the value of ρ used in the simulation (Fig. 4B).

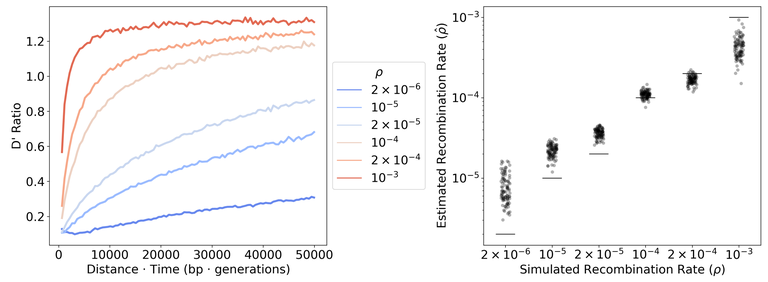


Figure 4: (A) Binned averages of D' ratios from simulations with known ρ. (B) Estimated value of ρ plotted against the known value of ρ from the simulations. Each point represents the estimate from a pair of two simulation runs.

Existing datasets enable in vivo estimation of recombination rates in low and high viral load settings

We applied our estimation method to longitudinal data from Zanini et al., 2015. The data had the following features:

- Experimental protocol minimized PCR recombination
- Paired end reads allowed for computation of D' for loci up to 700 bp apart
- Viral loads varied by orders of magnitude both within and between participants (Fig. 5)

To assess our hypothesis, we separated the pairs of timepoints into two groups based on whether the mean viral load between two timepoints was either above or below a given threshold.

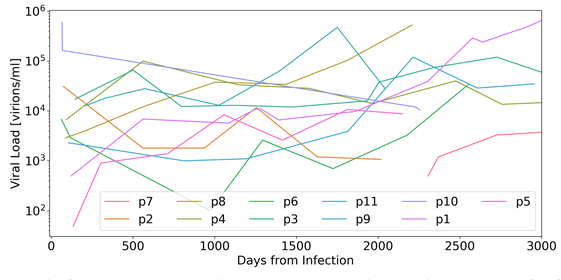


Figure 5: Viral loads from 11 untreated PLWH measured over the course of infection[3]. These viral loads vary by orders of magnitude both between and within individuals over time.

References:
1. Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Molecular Biology and Evolution*, *36*(3), 632–637.
2. D. L. Hartl, A. G. Clark, and A. G. Clark. Principles of population genetics, volume 116. Sinauer associates Sunderland, 1997.
3. Neher, R. A., & Leitner, T. (2010). Recombination Rate and Selection Strength in HIV Intra-patient Evolution. *PLOS Computational Biology*, *6*(1), e1000660.
4. Zanini, F., Brodin, J., Thebo, L., Lanz, C., Bratt, G., Albert, J., & Neher, R. A. (2015). Population genomics of intrapatient HIV-1 evolution. *ELife*, *4*, e11282.

Recombination rate varies across individuals

Using data from all participants, we estimated ρ separately within each viral load group. At a threshold of $> 5\times10^4$ copies/mL, among high viral load timepoints, ρ was estimated to be 6.2×10^{-5} (95% CI $[4.6\times10^{-5}, 8.1\times10^{-5}]$) and the low viral load ρ was significantly lower at 2×10^{-5} (95% CI $[1.3\times10^{-5}, 2.9\times10^{-5}]$) (Fig 6).

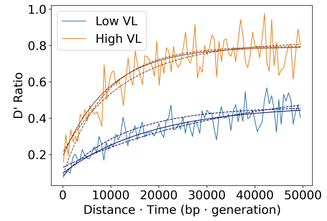


Figure 6: Example fit used for in vivo estimates at a threshold of 5×10^4 copies/mL. ρ is proportional to the derivative of the curve at 0. Fits for the estimates at the edges of the 95% confidence interval are indicated by dashed lines. Confidence intervals were created using 1000 bootstraps where loci were sampled with replacement.

We estimated ρ across a range of viral load (VL) thresholds and found that viral load has a dose-dependent effect on recombination rate (Fig 7).

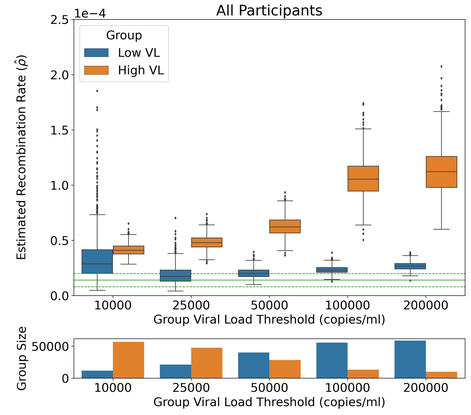


Figure 7 : In vivo estimates of ρ with units of recombinations per base pair and generation. Each box and whisker plot represents the distribution of estimates in the 1000 bootstrapped replicates. Whiskers represent 1.5xIQR. Estimates from Neher & Leitner, 2010 are indicated by the solid green line and the estimate ± one standard deviation is in dashed green. The bar plot in the lower panel indicate the number of loci pairs used for estimation.

Recombination rate varies within individuals

Participant p1 had sufficient genetic diversity and viral load variation for estimating recombination within them. Within p1, we estimate a nearly 10x elevation in ρ at times where VL exceeded 5×10^4 copies/mL compared to times below 5×10^4 (Fig 8).

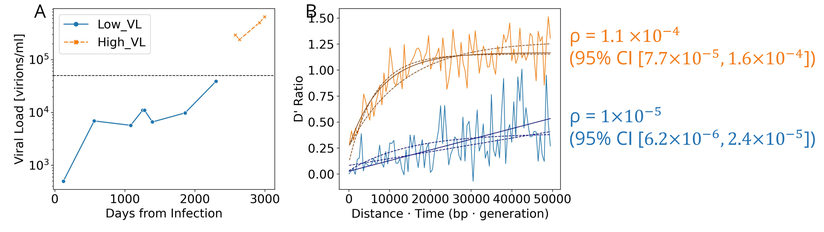


Figure 8: (A) Viral load in individual p1 for each sampling timepoint. Horizontal line represents viral load threshold of 5×10^4 copies/mL used for estimation. (B) Recombination rate estimation in p1. ρ is proportional to the derivative of the curve at 0. Fits for the estimates at the edges of the 95% confidence interval are indicated by dashed lines.

Conclusions

- High viral loads are associated with elevated in vivo recombination rates which can be up to an order of magnitude higher (10^{-4} rather than 10^{-5}) than in low viral load settings, suggesting recombination rates can vary by across PLWH and within those people over time.
- Low viral loads and corresponding low recombination rates may suppress the creation of multi-drug resistant haplotypes in combination-therapy treated HIV populations.
- On the other hand, high viral loads and elevated recombination during acute and/or uncontrolled infection may fuel ongoing diversification.
- Incorporating variation in recombination rate estimates could lead to new insights analyzing and interpreting intra-host genetic data.