SIV Envelope Diversity During Untreated Infection

Haoyue Li[1], Emily J Fray[1], Vivek Hariharan[1], Dan H. Barouch[2], Janet D. Siliciano[1], Robert F. Siliciano[1,3]

[1]Johns Hopkins University School of Medicine, Baltimore, MD, USA, [2]Beth Israel Deaconess Medical Center, Boston, MA, USA, [3]Howard Hughes Medical Institute, Baltimore, MD 21205, USA

BACKGROUND

Viral diversification complicates immune-based HIV-1 cure strategies. Simian Immunodeficiency Virus (SIV)-infected Rhesus Macaques (RMs) serve as an important model for studying viral evolution under consistent immune pressure and HIV-1 reservoir persistence. While antiretroviral therapy (ART) suppresses viral replication and halts disease progression, it does not eliminate the latent reservoir, which seeds viral rebound upon treatment interruption. The HIV-1 envelope (*env*) protein is the only viral protein on the virion surface and mediates viral entry into the host cell, making it a key target for therapeutic and preventative vaccine design. However, the significant diversification occurring within the *env* gene under immune pressure poses a major challenge[1-4]. Here we examine *env* sequences from pre-ART plasma virus and on-ART reservoir virus to determine the full extent of viral diversification and identify sequence features that may be targeted in vaccine development.

METHODS

Because sequencing errors give a false indication of viral evolution, we used single-genome sequencing (SGS) to collect ~10,000 high quality *env* sequences from 8 Rhesus Macaques (RM) inoculated with SIVmac251 by repetitive intra-rectal challenge. Animals were placed on ART 48 weeks after initial challenge (Fig. 1). Data from 4 representative RM are shown. We analyzed the extent of variation at individual amino acid positions and performed ecological methods[5] to estimate the total diversity of viral variants within a single host.

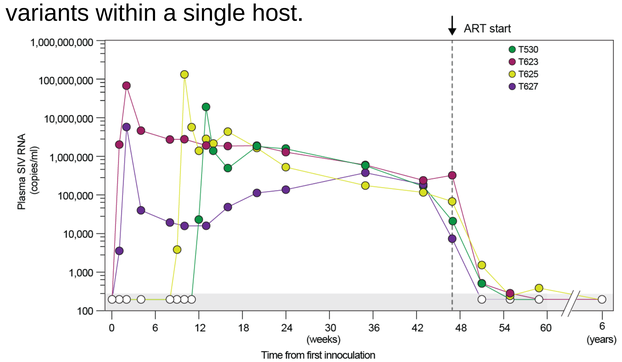


Fig. 1. Plasma SIV RNA levels. High level viremia was present before ART with stable suppression on ART, making these RM a good model for people treated during chronic infection and maintaining suppression on long term ART. Open circles indicate measurements below the assay detection limit (shaded area).

RESULTS

Despite inoculation with the same stock, distinct evolutionary patterns emerged in each animal (Fig. 2A). Dominant mutations in each animal were largely different (Fig. 2B) and often found in V regions of the *env* gene, con

Extraordinary **intra-host diversification** complicates therapeutic vaccine design

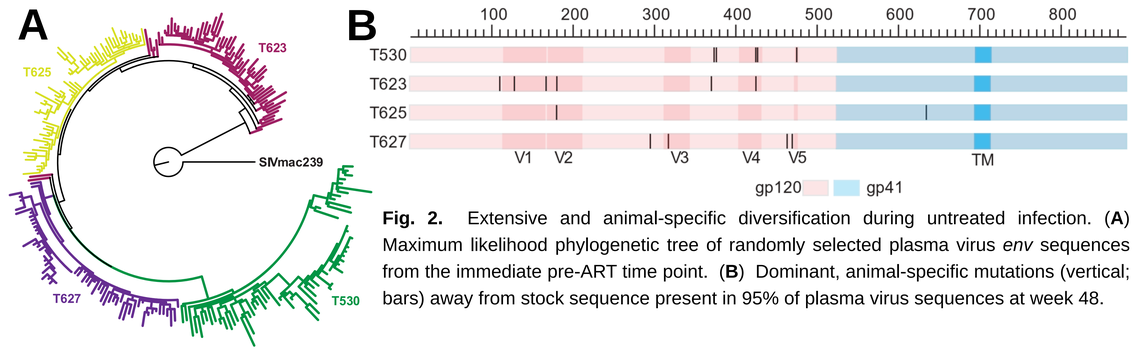


Fig. 2. Extensive and animal-specific diversification during untreated infection. (**A**) Maximum likelihood phylogenetic tree of randomly selected plasma virus *env* sequences from the immediate pre-ART time point. (**B**) Dominant, animal-specific mutations (vertical; bars) away from stock sequence present in 95% of plasma virus sequences at week 48.

RESULTS (cont)

-sistent with host-specific evolutionary pathways driven by autologous neutralizing antibodies (aNAbs). Standard analysis of diversification involves calculation of average pairwise distance, but this measure does not provide an estimate of total within-host sequence diversity. Therefore, we used the Chao1 estimator which provides a minimal estimate of the total number of distinct sequences based on the number of sequences found only once or twice[6] (Fig 3A). This analysis predicts that thousands of distinct *env* variants are present in each animal. In longitudinal sampling of pre-ART plasma virus and on-ART proviral sequences, the same sequence was rarely encountered (Fig. 3B). This is likely a reflection of the extensive diversification resulting from high viral loads throughout the period of untreated infection (Fig. 1). Over longer periods on ART, clonal expansion is likely to make sampling of the same

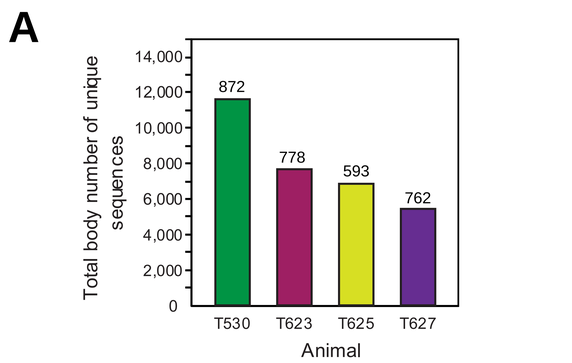
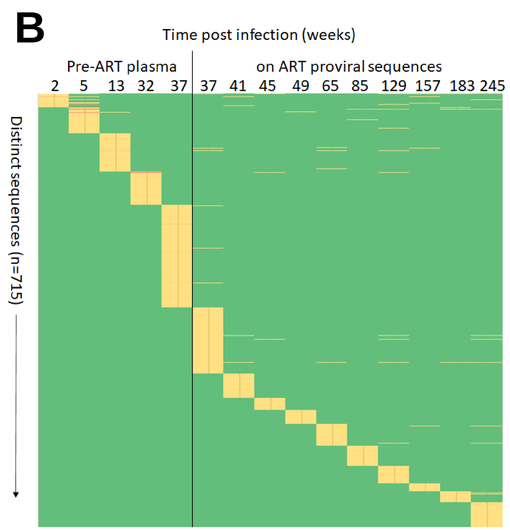


Fig. 3. Extraordinary intra-host diversification. (**A**) Chao1 estimates of the total body number of distinct *env* sequences. Numbers above bars indicate number of sequences analyzed. (**B**) Sampling of pre-ART plasma virus and on-ART CD4⁺ T cell proviral sequences mostly reveals new sequences at each time point due to the extensive pre-ART diversification. Each horizonal line represents a single sequence from T530.

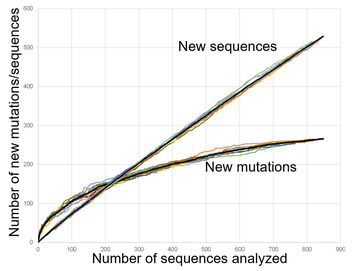


RESULTS (cont)

sequence more likely[7-9].

Despite the extensive diversification, constraints on viral evolution, even within the V regions, were readily apparent. Rarefaction curves showed that the number of unique sequences within a given host increased steadily with the number of sequences analyzed while the number of unique mutations plateaued (Fig. 4). Thus, much of the diversity represents different combinations of the same mutations.

Fig. 4. Number of new (unique) sequences or mutations found in a single host as a function of the number of single genome *env* sequences analyzed. Data are shown for a representative animal (T530) based on 872 sequences.



Even within the *env* V regions, only a small fraction of possible mutations are tolerated (Fig. 5).

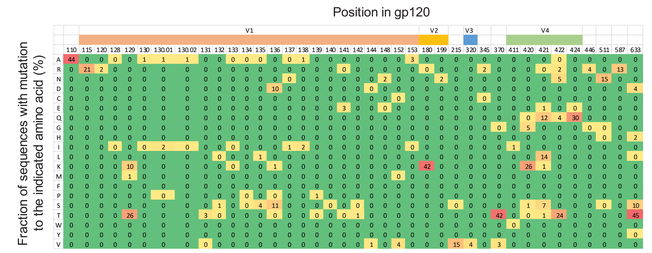


Fig. 5. Mutations from stock in T623. Table show % of 619 plasma and proviral sequences with mutations to the indicated amino acid at variable positions in gp120.

CONCLUSIONS

1) During chronic infection with high viremia, thousands of distinct *env* variants arise and persist in a single host.
2) Failure to find matching sequences in spatial or temporal reservoir studies is likely due to undersampling.
3) Despite extensive diversification, there are striking constraints on the number of alternative amino acids tolerated at positions in the V regions. Understanding these constraints could aid vaccine design.

ACKNOWLEDGEMENTS

Supported by the NIH MDC grant UM1AI164556 to D.H.B and by the Howard Hughes Medical Institute.

REFERENCES

1) Shankarappa et al, JV, 1999. 2) Wei et al Nature, 2003. 3) Richman et al, PNAS, 2003. 4) Zanini et al, eLIFE, 2015. 5) Reeves et al, Nat Comm, 2018. 6) Chao, A. 1984. 7) Bailey et al, JV,2006. 8) Maldarelli et al, Science, 2014. 9) Wagner et al., Science, 2014.