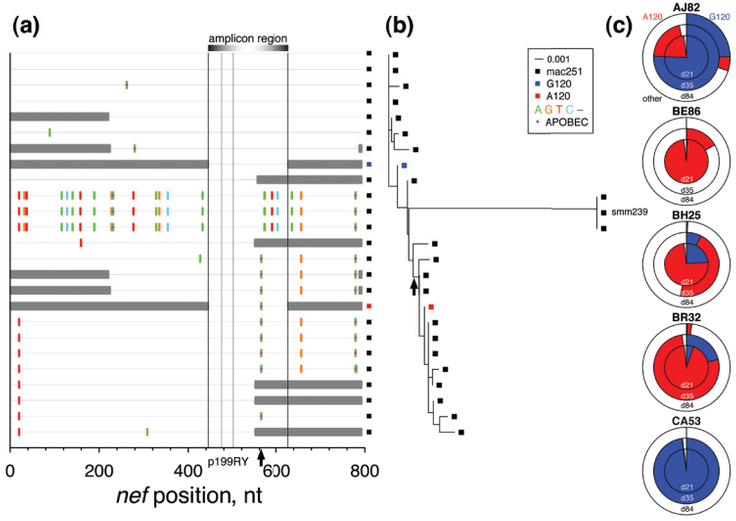
Early Detection and Dynamics of Rare Viral Variants via Ultra-deep Sequencing

Peter Hraber, Will Fischer, Elena Giorgi, Thomas Leitner, Bette Korber, T-6; Cheryl Gleasner, Lance Green, Cliff Han, B-6; Evan Cale, Norm Letvin, Harvard University

Despite epitope-specific immune control of viral replication, some viral mutations confer immune escape. Vaccines that stimulate broad immune response limit the potential for escape. Previous work reported a CTL response to escape mutations 35 dpi, earlier than detected with conventional sequencing. We used ultra-deep sequencing to quantify viral variants and characterize escape dynamics in an immunodominant epitope. Sequence diversity analysis indicates multiple founders. In intravenously infected rhesus macaques, the two predominant sequences (>98% of sample) at 21 dpi differ by a silent substitution. However, the earliest escape mutations occur within the host, rather than being transmitted from inoculum. Non-synonymous-to-synonymous substitution ratios show positive selection in the epitope by 21 dpi. Escape variants occur in each infection at 21 dpi, which subsequently outcompete founders. Ultra-deep sequencing provides sensitivity to detect rare variants and shows immunity influencing viral evolution before peak CTL response, earlier than previously detected with conventional sequencing.

Fig. 1. Sequence diversity is limited in nef. (a) Highlighter plot shows locations of sequence polymorphisms (vertical lines indicate the amplicon sequenced and the p199RY epitope; grey bars indicate sequence gaps). (b) Corresponding phylogenetic tree depicts sequence relations. (c) At 21 dpi, over 98% of samples from each of 5 macaques (AJ82, BE86, BH25, BR32, CA53) consist of founder sequences identical except for a clade-defining silent G or A polymorphism in one position outside the epitope (arrow). Though the founders diminish later (35 and 84 dpi), they do not disappear completely.



An understanding of the timing of immune selection and how viruses evade the host immune response is needed to inform vaccine-design strategies. The ability to detect rare viral variants as they accumulate under cytotoxic-T-lymphocyte (CTL) selection illuminates the evolution of viral escape from the immune system. While Sanger sequencing can yield highly accurate sequences, detecting rare variants requires intensive sampling. In contrast, next-generation sequencing technologies yield orders of magnitude more sequences and provide sensitivity to detect rare variants. Previous sequencing results informed design of an ultra-deep pyrosequencing strategy to study evolution in the SIV p199RY nef epitope in experimentally infected macaques.

Five rhesus macaques (AJ82, BE86, BH25, BR32, CA53) were intravenously infected with 60,000 copies per mL of SIVmac251, and their plasma was sampled at 21, 35, and 84 days post-infection (dpi) [1]. (Note that multiple founders are expected from intravenous infections.) We deployed cDNA generation from viral RNA and subsequent PCR steps that used high template volume and low cycle numbers, then pooled replicates and performed 454 amplicon pyrosequencing. Because insertions/deletions in homopolymer regions are common sequencing errors, we corrected them to match the consensus sequence and reduce data loss [2]. We also excluded sequences containing stop codons or short inversions near homopolymers.

Sequence variation was limited early after infection, and each animal was infected by multiple founders, represented by a predominant sequence polymorphism that differed by a silent substitution (Fig. 1). A simple model of sequence diversity in acute infection [3] rejected the null hypothesis of Poisson-distributed pairwise distances in all five animals ($P<0.001$), consistent with the presence of selection and multiple founder variants. Comparing synonymous and non-synonymous substitution rates detected positive selection at 21 dpi in four of five animals ($P<0.05$) [4].

Epitope frequencies change over time, from the transmitted form to rapidly escaping variants (Fig. 2). In all five animals, emerging variants S3L and I7T are present, though rare, in multiple copies at 21 dpi, and S3L predominates at 84 dpi. Linkage of these rapidly escaping variants to only the more abundant inoculum polymorphism suggests de novo

For further information contact Peter Hraber at phraber@lanl.gov.

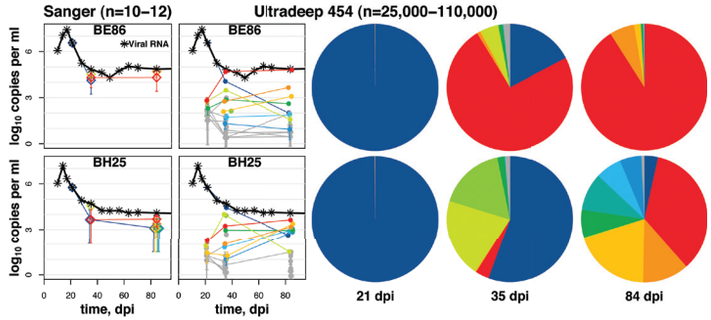


Fig. 2. SIV epitope frequency dynamics in two representative macaques, BE86 and BH25. While both Sanger and 454 sequencing platforms find common epitopes, ultradeep 454 sequencing detects rare variants earlier than limited Sanger sequencing. Error bars show 95% confidence intervals from binomial distributions. Colors indicate epitopes as listed in Fig. 2, from the transmitted form (blue) to escape variants (red to light blue).

mutations, rather than transmission from the inoculum. Double mutants for both S3L and I7T and new variants (T2S and Y9F) emerge in a later wave of escape mutations by 35 dpi. Though the transmitted epitope diminishes over time, it persists at low levels, probably in reservoirs of latently infected cells.

Maximum-likelihood phylogenies [5] from each sample relate amplicon nucleotide sequences to epitopes (Fig. 3). That diversity increases with time and replacement of the transmitted epitope after selective sweeps is clearly evident. In ongoing work, kinetic models of escape will quantify relative fitness of variants and the CTL killing rate [6].

Clearly, ultra-deep sequencing provides sensitivity to detect rare variants earlier than with conventional sequencing. We found an early wave of escape mutants at 21 dpi that had not been detected previously, revealing that immune selection is active within 3 weeks of infection.

Special Thanks

We thank our collaborators Gayathri Athreya, Heidi Bazick, Tanmoy Bhattacharya, Mira Bussod, Vitaly Ganusov, Beatrice Hahn, Barton Haynes, Sandrine Hulot, Chien-Chi Lo, Andrew McMichael, Alan Perelson, Ruy Ribeiro, George Shaw, James Whitney, and Wendy Yeh.

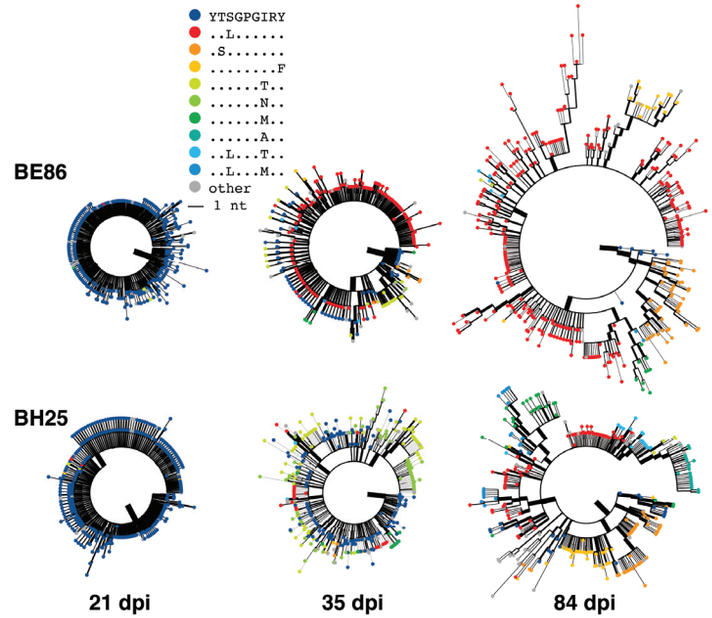


Fig. 3. SIV phylogenies from two representative macaques show quasispecies diversity increasing with time, and population turnover as dominant epitope variants (colored circles) undergo CTL-induced selective sweeps. Each tree is rooted hubward on the transmitted form, with branch widths cumulative and proportional to the log of variant frequency in the sample. The scale bar depicts distances of one nucleotide.

[1] Newberg, M.H., et al., *J Immunol* **176,** 319 (2005).
[2] Fischer, W., et al., *PLoS ONE* **5,** e12303 (2010).
[3] Giorgi, E.E., et al., *BMC Bioinformatics* **11,** 532 (2010).
[4] Korber, B., in *Computational Analysis of HIV Molecular Sequences,* Kluwer Academic, A.G. Rodrigo and G.H. Learn, Eds., 55-72 (2000).
[5] Guindon, S., and O. Gascuel, *Syst Biol* **52,** 696 (2003).
[6] Ganusov, V.V., and R.J. De Boer, *PLoS Comput Biol* **2,** e24 (2006).

Funding Acknowledgments

LANL Laboratory Directed Research and Development; National Institutes of Health, Center for HIV/AIDS Vaccine Immunology