Abstract #1877

Next Generation Sequencing of Full-Length HIV-1 *env* During Primary Infection

Melissa E. Laird[1], N. Lance Hepler[2], Colleen Ludka[2], Michael Brown[2], Stephen Espitia[3], Ben Murrell[1], Yan Guo[2], Douglas D. Richman[1,3], Sergei L. Kosakovsky Pond[1], Ellen E. Paxinos[2], Davey M. Smith[1,3]

[1] Department of Medicine, University of California, La Jolla, CA 92093; [2] Pacific Biosciences, 1380 Willow Road, Menlo Park, CA 94025; [3] Veterans Affairs San Diego Healthcare System, San Diego, CA 92161

Melissa Laird, Ph.D.
University of California, San Diego
9500 Gilman Drive, MC0679
San Diego, CA 92093
E-mail: mlaird@ucsd.edu

Introduction

The use of next-generation sequencing (NGS) to examine circulating HIV *env* variants has been limited by *env* gene length (~2.6 kb), indel polymorphism, GC deficiency, and long homopolymeric regions.

Objective

To develop and standardize protocols for plasma viral RNA isolation, RT-PCR amplification, single-molecule real-time (SMRT®) Sequencing, and bioinformatics analysis of circulating HIV-1 *env* variants to evaluate viral diversity in primary infection.

Methods

PCR

Day 1 Virus pelleted by high-speed centrifugation through a sucrose cushion

↓

Viral RNA extracted (QIAGEN kit), cDNA generated (SuperScript III pol)

↓

FL *env* amplification (Q5 Pol, NEB)

↓

Day 2 Analyze PCR by gel electrophoresis and BioAnalyzer instrument

↓

Day 3 Iterative PCR (500 ng product required)

Figure 1. HIV genome, with our amplicon highlighted in red. Primers bracket the *env* coding sequence, amplifying an expected product of 3.2 kb. Primer sequences are as follows:
EnvA (F): GCTTAGGCATCTCCTATGGCAGGAAGAA
EnvN (R): CTGCCAATCAGGGAAGTAGCCTTGTGT

Methods

SMRT Sequencing

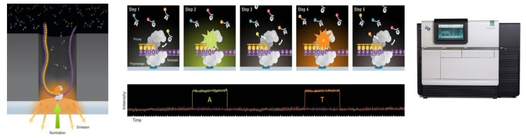

Figure 2. HIV *env* amplicons were sequenced on the PacBio RS II instrument using commercially available P4-C2 chemistry and protocols.

Results

PCR Amplification of 3.2kb HIV-1 *env*

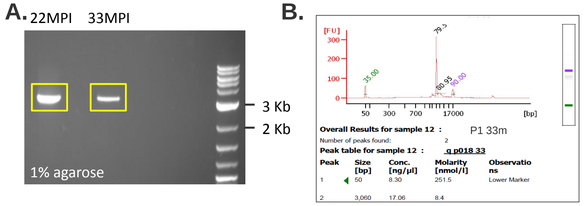

Figure 3. (A) 3.2 kb *env* amplicon from subject P1 at 22 and 33 months post-infection. (**B**) Bioanalyzer quantitation of final P1, 33 month *env* amplicon prior to SMRTbell library construction.

SMRT Sequencing of FL HIV-1 *env* Amplicons

Subject ID	Months PI	Viral Load (log IU/mL)	CCS reads	Raw reads	Diversity
P1	3	5.65	11,541	75.0K	0.74%
	22	4.55	11,316	67.6K	1.15%
	33	4.54	12,234	82.1K	2.0%
P9	3	4.36	9,246	57.4K	0.77%
H4	28	4.22	7,322	35.5K	1.3%
K4	12	4.57	5,098	32.8K	0.8%
Q8	6	5.51	8,775	63.7K	0.31%

Table 1. Summary of full-length HIV-1 *env* SMRT Sequencing. Diversity was measured as the mean nucleotide pairwise distance among circular consensus sequence (CCS) reads.

Results

Coverage and Variation of HIV-1 *env*

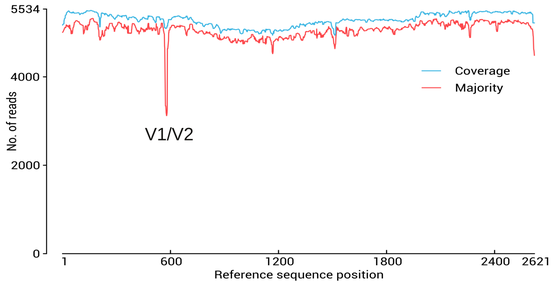

Figure 4. Examples of CCS coverage (blue) and sequence count with majority residue (red) over *env* for subject P1 at 3 months post-infection, compared to the in-sample consensus.

HIV-1 *env* Evolution

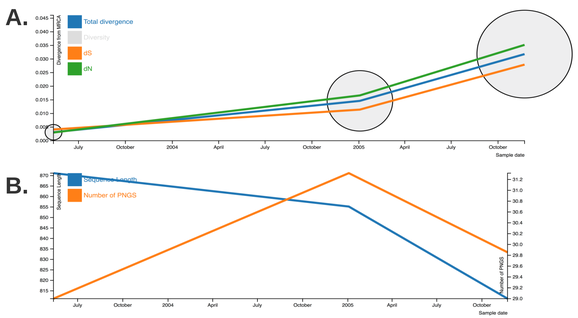

Figure 5. (**A**) Temporal evolution of *env* sequences in P1: mean synonymous, non-synonymous and total divergence (from the imputed ancestral strain), and mean within-sample nucleotide diversity. (**B**) Examination of sequence length and putative N-linked glycosylation sites (PNGS) in FL HIV-1 *env* from Subject P1 throughout primary infection (3, 22 and 33 months post-infection).

Acknowledgements

The authors would like to sincerely thank all participants of the San Diego HIV Primary Infection Research Consortium. This work was supported by the Department of Veterans Affairs, and National Institute of Health (NIH) awards AI106039, AI090970, AI100665, AI036214.

Results

HIV-1 *env* Phylogeny

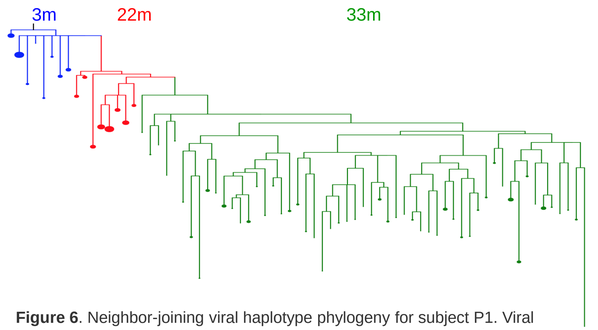

Figure 6. Neighbor-joining viral haplotype phylogeny for subject P1. Viral haplotypes at 3 (blue), 22 (red) and 33 (green) months post-infection were estimated by hierarchically clustering circular consensus sequences (CCS) and constructed using a windowed partial order alignment.

Conclusions

- This study developed a standardized procedure using PacBio SMRT technology to deep sequence full-length HIV *env* variants from the circulating viral population, achieving good coverage, and confirming the pattern of low *env* diversity during primary infection that increased over the course of disease progression.
- The number of reconstructed viral haplotypes increased from 8 to 55 throughout primary infection. Haplotype diversity increased from 0.74% (3 months) to 1.15% (22 months) and to 2.0% late in infection (33 months).
- The long, accurate reads obviate the need for short-read-based computational haplotype reconstruction, increasing our confidence in the results.
- The sequencing methodology and analysis tools developed here are immediately useful for any setting in which full-length HIV *env* analysis would be applicable.