Molecular Characterization of Highly Diverse HIV-1 Viruses Using Metagenomic Next-Generation Sequencing (mNGS) as Reference Reagents for Panel Development

Jiangqin Zhao[1], Hanxia Huang[1], Venice Servellita[2], Alicia Sotomayor Gonzalez[2], Helene Highbarger[3], Robin L. Dewar[3], Charles Chiu[2], and Indira Hewlett[1]

[1] Lab of Molecular Virology, DETTD, CBER, FDA, Silver Spring, MD; [2] Dept of Laboratory Medicine, UCSF Clinical Microbiology Laboratory, UCSF, CA; [3] Virus Isolation and Serology Laboratory (VISL), SAIC-Frederick, Inc. Frederick, MD

Abstract

HIV continues to be a highly diverse virus and therefore poses significant challenges to HIV diagnostics and therapeutics[1,2]. We characterized several hundred HIV positive plasma and viruses obtained from Cameroon to identify highly divergent HIV strains for reference panel development[3]. The MSSPE panel that was used potentially enriches for 14 different RNA viral pathogens including HIV, facilitating both identification of low-titer infections and viral whole-genome sequencing[4]. NGS data were analyzed using the SURPI+ software. De novo assembly using the SPADES algorithm followed by reference-based assembly using the SHIVER algorithm was performed for assembly of HIV-genomes[5]. A total of 101 cultured HIV-1 viruses and 257 patient plasma samples were analyzed using mNGS to identify all HIV-1 strains and various viral pathogens present based on alignment to comprehensive viral genomic sequence databases from GenBank and Los Alamos National Laboratory (LANL). Ninety nine percent of samples from HIV-positive individuals with 102-106 copies/mL had detectable HIV genomic sequence by NGS analysis. Several viruses were detected with very low reads in some samples such as pegiviruses, human herpesvirus 5, and hepatitis GB virus B. The 37 cDNA libraries will be re-sequenced for further genomic characterization. Among them, 18 HIV-1 variants were used in the pilot study for determining the viral load (VL) titer of highly diverse HIV variant reference panels. The current study will help to develop the next generation of HIV reference panels for use in evaluating performance of nucleic acid tests and antigen/antibody serologic tests, as well as new assays for HIV latency.

Introduction

HIV-1 exhibits high genetic diversity due to an error prone reverse transcriptase that results in mutations and high rate of recombination. The continued diversification of HIV poses significant challenges to diagnostics and therapeutics. The dynamic evolution of emerging variants in Cameroon is demonstrated by the prevalence of all known subtypes and circulating recombinant forms (CRFs) in this region. We obtained several hundred HIV positive plasma and viruses from this region for characterization and identification of highly divergent HIV strains for HIV reference panel development.

Materials and Methods

To characterize the genomic background more accurately for new HIV variants or any co-existing pathogens present in those samples, we collaborated with the Chiu laboratory at University of California, San Francisco to perform metagenomic next-generation sequencing (mNGS) using the Metagenomic Sequencing with Spiked Primer Enrichment technique (MSSPE) to detect and discover emerging viral pathogens (Fig. 1). The MSSPE panel that was used potentially enriches for 14 different RNA viral pathogens including HIV, measles, Marburg, Lassa, Ebola, Zika, yellow fever, West Nile, Rift Valley Fever, hepatitis E, human immunodeficiency, dengue, chikungunya, CEV, Crimean-Congo hemorrhagic fever, and hepatitis C viruses, facilitating both identification of low-titer infections and viral whole-genome sequencing (Fig. 3).

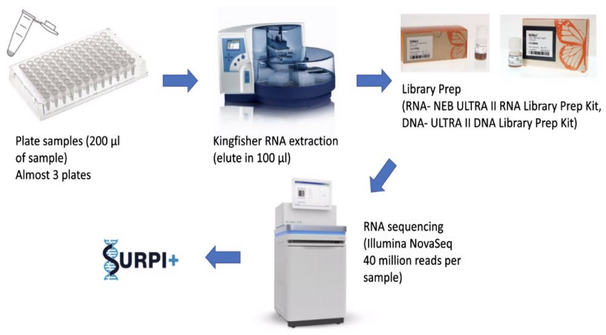

Figure 1. SURPI Library Preparation Pipeline

Results and Discussion

A total of 101 cultured HIV-1 viruses and 257 clinical plasma samples were analyzed using mNGS to identify all HIV-1 strains and various viral pathogens based on alignment to comprehensive viral genomic sequence databases from GenBank and Los Alamos National Laboratory (LANL). After library preparation, samples were run on Illumina NovaSeq and NGS data were analyzed using in-house SURPI+ software (Fig. 1, 3). De novo assembly using the SPADES algorithm followed by reference-based assembly (from 3,571 HIV reference sequences) using the shiver algorithm were performed for assembly of HIV-genomes[5]. A total of 100 of 101 (99%) viral samples from HIV-positive individuals from Cameroon with 102-106 copies/mL had detectable HIV genomic sequence by NGS analysis. Several viruses were detected with very low reads in some samples such as pegiviruses, human herpesvirus 5 (cytomegalovirus), and hepatitis GB virus B. The 37 cDNA libraries containing candidate viral pathogens will be re-sequenced for further genomic characterization (Fig. 2). Among them, 18 HIV-1 variants were used in the pilot study for determining the viral load (VL) titer of highly diverse HIV variant reference panels (Tab. 1).

Table 1. Pilot Study for Highly Diverse HIV Variants Panel Preparation

Patient ID	LMV Subtype	LMV Identify	LANL Subtype	Pilot Study tested VL(copies/mL) 100	500	1000	5000
11CMMDC224	F2	Subtype	F2	55	188	381	2252
10CMLB030	G	Subtype	G	93	290	664	4721
11CMLB115	02_AG	CRF	02_AG	54	286	526	2860
06CMARC13	06_cpx	CRF	06_cpx	77	551	915	4592
10CMLB018	18_cpx	CRF	18_cpx	114	603	1157	6170
10CMLB092	22_01A1	CRF		84	322	770	3644
11CMMDC211	22_01A1	CRF		123	616	1225	6768
08CMNYU997	02,A1	URF	02A1	121	551	1291	6339
06CMBDSH13	02,U	URF	02G	73	315	689	3799
10CMLB008	02,18	URF	U	57	265	591	3091
10CMLB040	G,BG,	URF		85	495	914	4691

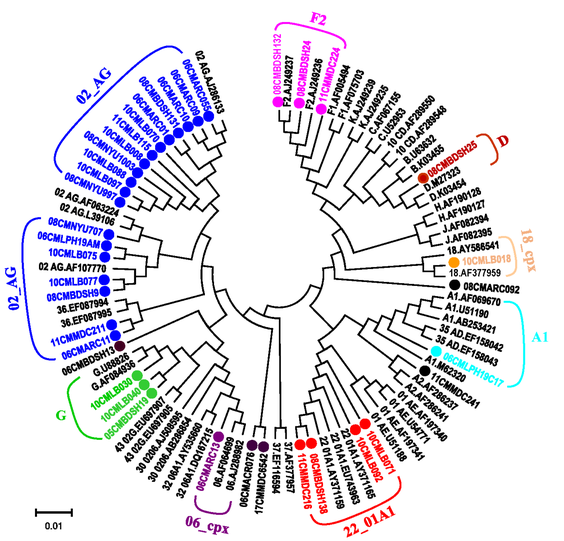

Figure 2. Phylogenetic tree of near full-length HIV-1 genome sequences

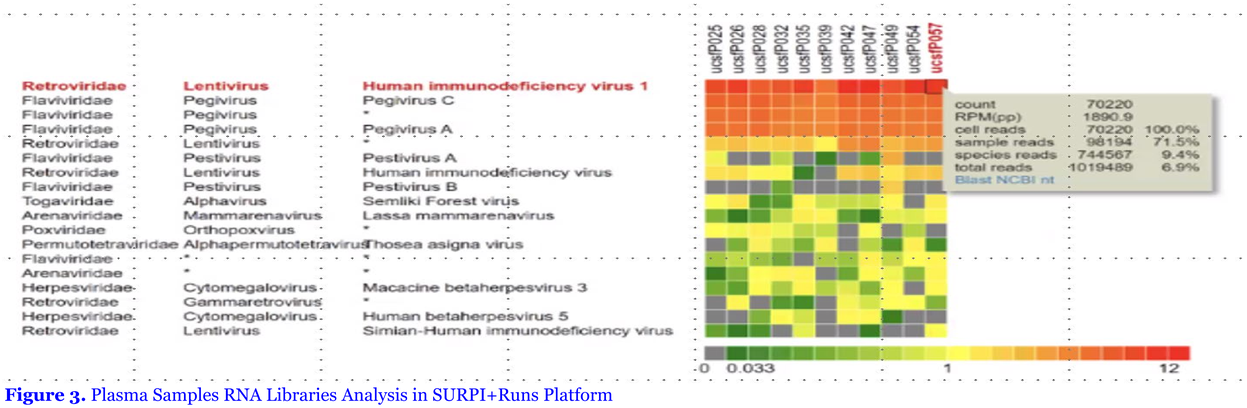

Figure 3. Plasma Samples RNA Libraries Analysis in SURPI+Runs Platform

Conclusion

Current studies will determine if there are unknown strains related to HIV and/or other emerging co-infecting pathogens that may be present in these samples. It will provide more information for full characterization of the panel materials in terms of HIV strain assignment and that of co-infecting pathogens. The current study will help develop the next-generation of HIV reference panels for use in evaluating performance of serologic antigen/antibody combo tests and nucleic acid tests as well as new assays for HIV latency.

References

1. Manak M, et al., AIDS Res Hum Retr 2012;28: 594-606.
2. Price MA, et al., J Infect Dis 2019;220: 432-41.
3. Zhao J, et al., Viruses 2021;13: 1-12.
4. Deng X, et al., Nat Microbiol 2020;5: 443-54.
5. Wymant C, et al., Virus Evol 2018;4: vey007.

* Project supported by the NHLBI and CERSI