

# NextGeneration Sequencing to Characterize Near Full Length Human Immunodeficiency Virus (HIV-1) Subtypes Evolving Within Infected Host

Supraja Kanipakam, Viswanath Ragupathy Ph.D, Indira Hewlett Ph.D, and Andrew Huang

**Abstract** — Human Immunodeficiency Virus (HIV) is a retrovirus that globally affects all ages, sexes, and races. As of December 2015, approximately 36.7 million people worldwide are living with HIV. The goal of this study was to (1) sequence the near full length of HIV-1 virus and (2) discover heterogeneity of HIV-1 in the infected individuals, whether the subtypes evolved independently or as evolved as recombinants of subtypes (CRFs). By evaluating samples from Cameroon as a model to study HIV-1 diversity, we used infected plasma samples and characterized near full-length HIV strains. In order to accomplish this goal, several steps were taken, including nucleic acid extraction, reverse transcription of RNA, polymerase chain reaction (PCR), and gel electrophoresis. Next-Generation Sequencing (NGS) technology was utilized to sequence the DNA. Our results yield that out of the 3 HIV-infected linked samples evaluated: one male sample was infected with CRF02\_AG, which includes a CRF and subtypes A and G, and two of his partners were infected with a recombinant known as F2/CRF02. CRF02 virus from the male partner recombined with F2 infected female partners and evolved as F2/CRF02 recombinants. This study provides a useful tool in order to thoroughly understand the HIV epidemiology and pathogenesis, HIV diagnosis, and produce effective vaccine.

## I. INTRODUCTION

HIV type 1 (HIV-1) is a global pandemic that affects over 6,850 people every day worldwide. The HIV virus is composed of four groups: M, N, O, and P. The M group comprises of sub-subtypes (A-K) and circulating recombinant forms (CRFs), which are highly diversified strains that contain more than one subtype. Previously, it has been reported that West central Africa is an epicenter for circulation of diverse HIV-1 subtypes [1]. Blood specimens from Cameroon were collected in order to study the viral genetic diversity of the HIV-1 virus and how the HIV-1 virus' subtypes evolve in the infected host. Currently, the CRF, CRF02\_AG, which is composed of subtypes A and G, is the most dominant strain in Cameroon. Recombinants of CRF02\_AG may form as the virus evolves. The evolution of HIV-1 recombinants poses a threat to diagnostics, vaccines, and treatment. Therefore, there is need for continuous surveillance for these emerging viruses' and its recombinants.

## II. METHODS

Our laboratory is evaluating samples from Cameroon as a model to study HIV-1 diversity. The rationale for selecting Cameroon is that these blood specimens may contain complex HIV strains and other retroviruses. Cameroon in West Central Africa is known to harbor multiple genetically diverse and recombinant strains and CRF02\_AG continues to be the pre-dominant strain for over a decade. However, emergence of recombinants of CRF02\_AG may be likely as the epidemic evolves in this region which may affect diagnostics, vaccine and therapies.

In order to accomplish this goal, the nucleic acids were extracted from the plasma and reverse transcription of the RNA was performed to obtain full length HIV-1 cDNA. First-strand cDNA was synthesized into total RNA and PCR was

performed to amplify these DNA fragments; in order to confirm the Amplicon size, gel electrophoresis was conducted with 1% Agarose gel. MiSeq NGS technology was used in order to complete whole-viral genome sequencing and provide a comprehensive view of the entire genome. Ultimately, data analysis was completed to process/annotate the data, report genomic variants and biological context.

## III. RESULTS

Through this investigation, we were able to discover the heterogeneity of HIV-1 in the infected individuals. The genotypes CRF02 and F2 both co-evolved within two partners.

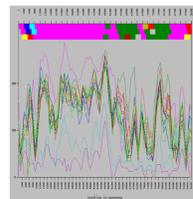


Figure 3A. Sample 1

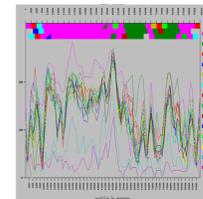


Figure 3B. Sample 2

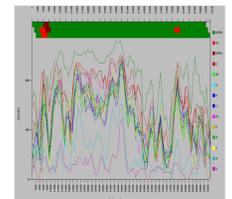


Figure 3C. Sample 3

In Figure 3, each of the patient samples was blasted against HIV-1 reference sequences in distinctive colors. Fig. 3A displays Sample 1; the high concentration of pink shown displays CRF02/F2 as the predominant subtype. Fig. 3B displays Sample 2, which also displays the recombinant CRF02/F2 as the dominant subtype, while Fig. 3C displays Sample 3, with subtype F2 as the lead subtype. Samples 1 and 2 are CRFs, while Sample 3 is a pure genomic form of the F2.

The recombinant in the female partner (F2/CRF02) may be contributed by the male partner. It is important to note that in the male partner: only the preliminary subtype in the patient (CRF02) dominated, instead of the recombinant CRF02/F2 as in the female patients. In this study, it was observed that subtypes F2 and CRF02 evolved independently in the female partners. Finding the genotype of infected HIV patients can help acquire and target new drugs and vaccines engineered to a specific genotype, such as F2 or the recombinant form F2/CRF02.

## IV. ACKNOWLEDGMENTS

S.Kanipakam wishes to acknowledge her mentor, Dr. Viswanath Ragupathy, the lab chief Dr. Indira Hewlett at the FDA, and Andrew Huang for his expertise in bioinformatics.

## V. REFERENCES

- [1] Magiorkinis, G., Angelis, K., Mamais, I., Katzourakis, A., Hatzakis, A., Albert, J., & ... Paraskevis, D. (2016). Research paper: The global spread of HIV-1 subtype B epidemic. *Infection, Genetics And Evolution*, doi:10.1016/j.meegid.2016.05.041