GENETIC CHARACTERIZATION OF HUMAN IMMUNODEFICIENCY VIRUS FROM NORTHERN SOUTH AFRICA

BY

Benson Chuks Iweriebor (BSc Honours, MSc)

A Thesis Submitted In Fulfilment of The Requirements For The Award of The Degree Of Doctor of Philosophy in Microbiology At the University of Venda

Supervisor
Dr. Pascal Obong Bessong
AIDS Virus Research Laboratory
Department of Microbiology
University of Venda, Thohoyandou
South Africa

Co-supervisors
Prof. Jeffrey Mphahlele
HIV/AIDS and Hepatitis Unit
Department of Virology
University of Limpopo, Medunsa Campus, Pretoria
South Africa

Prof. Sylvester Rodgers Moyo
School of Health Sciences,
National Polytechnic,
Windhoek, Namibia

January 2011

ABSTRACT

Globally, human immunodeficiency virus type 1 (HIV-1) is extraordinarily variable, and this diversity poses a major obstacle to AIDS vaccine development, diagnosis and therapy. Since HIV-1 M group began its expansion in humans roughly 70 years ago it has diversified rapidly now comprising a number of different subtypes and circulating recombinant forms (CRFs). Currently, strains belonging to the same subtype can differ by up to 20% in their envelope gene, and between subtype distances can soar to 35%. Moreover this diversity is continually growing. Although the scale of the HIV-1 pandemic makes action imperative, there is still much to learn about the extent and immunological implications of HIV-1 sequence diversity. As with other infectious agents, effective public health surveillance is essential to track the epidemic, guide research, and direct prevention activities. Significant challenges were highlighted by recent finding that some of the more divergent HIV-1 strains were not reliably detected by all antibody screening tests in current use.

HIV-1 subtype C is responsible for the vast majority of the estimated 5 million infected South Africans. Thus, while subtype C viruses dominate the epidemic, the exceptional high prevalence rates in South Africa could provide significant opportunities for the spread of new genetic subtypes and/or evolution of current subtypes. Furthermore, analysis should therefore focus on areas where little information is known, such as the Limpopo Province, and where the opportunity for the introduction of new variants exists, and where prevalence rates are relatively high. The strategic location of the Province makes it imperative for regular diversity studies to be carried out. The implications of HIV-1 diversity in diagnostics and vaccine development; and the formulation of treatment regimens necessitated this study. Thus, this study therefore identified and characterized circulating HIV-1 genetic variants in sero-positive, drug naïve populations from Limpopo Province where there is an estimated prevalence rate of about 19%.

The molecular epidemiology of HIV-1 in two highly endemic areas in Limpopo Province of South Africa; mutations associated with drug resistance to protease inhibitors (PIs), nucleotide reverse transcriptase inhibitors (NRTIs) and non

nucleoside reverse transcriptase inhibitor (NNRTs), co-receptor usage and epitope mapping of HIV-1 isolates from drug naive individuals were investigated. Subtyping was done by phylogenetic analysis making use of ClustalX2 software while drug resistance mutation analysis, co-receptor usage prediction epitope mapping and substitution of the functional motifs were all determined by interpretation algorithms.

Phylogenetic analysis of the test isolates revealed that subtype C is predominantly driving the epidemic in northern South Africa (Limpopo Province). Co-receptor usage prediction showed that majority of the isolates were R5 viruses as all had their tetrapeptide GPGQ motif characteristic of subtype C utilizing CCR5 viruses conserved. Genetic drug resistance mutation analysis of the 35 PR gene sequences did not reveal any major mutation associated with resistance but a high degree of minor mutations and polymorphisms were observed. Examination of the 44 RT genes showed a K103N substitution in two isolates. K103N change causes high level resistance to nevirapine. Epitope mapping of the *gag* p17 and p24 consensus sequences of the test isolates did not reveal any difference between them and the subtype C consensus sequences. They all had the same dissociation constant for the epitopes recognized by the HLA they were mapped against. Also, all the functional motifs in the PR and RT genes were conserved in majority of the test isolates.

Molecular characterization of the test isolates has helped to update the baseline data on the circulating strains of HIV -1 in northern South Africa. Since all the isolates are subtype C as in other regions of South Africa and result of epitope mapping compares very well with those of subtype C consensus sequences, vaccine based on subtype C viruses could be designed and evaluated in the Province. Also, it has shown that entry inhibitor- the new class of antiretroviral drug could be of significance should the current NRTIs and NNRTIs begins to fail as majority of the isolates had their GPGQ motifs conserved. Since no important resistance mutation to the PIs and RTIs was found among the test isolates, usage of these classes of drugs will continue to have positive impact in reducing morbidity and mortality due to AIDS in the studied area.

The second element of this study focused on the amplification and analysis of a unique recombinant form composed of subtypes A1 and C subgenomic regions. Recombination plays a key role in HIV-1 genetic diversity which on the long run has a grave implication on diagnosis, therapy and vaccine development. The URF recombinant that was analysed was isolated from a female patient residing at Bela Bela which has a high HIV-1 prevalence. The amplified near-full length genome was sequenced by the 454 Genome Sequencer FLX system. The sequence generated was delineated into respective gene by the sequence locator- a web-based online tool. Analysis of the recombinant virus was carried out in order to determine the subtypes that constituted the mosaic genome making use of jpHMM and REGA subtyping analytical tools.

The results obtained revealed that the mosaic genome is composed of A1/C with seven breakpoints of recombination partitioning the genome into eight segments alternating between sub-subtypes A1 and C viruses. Analysis of the near-full length genome by subtyping tools showed discordant assignment of some gene regions by the different tools. Further analysis of the accessory genes did not reveal any major changes like premature termination or loss of functional motifs but deletions and insertions were observed in the *tat, rev* and *nef* genes respectively. The isolation of a recombinant virus in a region where subtype C is the dominant variant shows the dynamic nature of this virus and calls for regular monitoring of the HIV-1 genetic landscape of the region.

The third component of this study also involved the analysis of a recombinant virus isolated from a female patient at Mankweng Hospital near Polokwane the provincial capital of Limpopo Province. Partial fragment of 5665 nucleotides was generated, sequenced and analysed by various subtype analytical tools. Presence of drug resistance mutations in the PR, RT and IN genes was determined as well as prediction of co-receptor usage.

Results revealed a mosaic recombination between subtype C and CRF11_cpx only at the RT gene of the isolate. All other gene regions analysed phylogenetically belong to pure subtype C virus. This is a novel strain as there is no known variant that has this genomic recombination in the HIV database. The epidemiological implication of this strain in Limpopo Province is not known.

The frequency and pattern of polymorphisms among HIV-1 subtypes associated with resistance or resulting to a faster emergence of drug resistance once under drug pressure has been evaluated extensively on subtype B and little information exist about other subtypes. Subtype C variants are responsible for more than 50% of the global epidemic, and it is the subtype that is driving the epidemic in Southern Africa, the region with the highest HIV prevalence in the world. Antiretroviral scale up in the Limpopo region is high and therefore, it is necessary to determine whether genetic subtype differences will influence therapy outcome. The investigative theme of chapter five of this thesis is on the implications of nucleotide polymorphisms on the genetic barrier to evolution of drug resistance mutations in subtype C viruses when compared to the global subtype B consensus sequence. The protease and reverse transcriptase nucleotide sequences generated in this study were compared with the global subtype B consensus sequence at codons known to code for drug resistance mutations according to the Stanford drug resistance algorithm.

The results revealed a reduced genetic barrier in subtype C viruses at codon V106M (GTA to GTG) and an increased barrier at codon L210W (TTA/CTG/CTA to TGG) when compared to subtype B global consensus sequence. From this analysis, these are the only codons where significant differences exist between the subtypes. Apparently, there are, no major genetic barriers existing between subtypes B and C at known positions that code for drug resistance mutations.

In conclusion, HIV-1 subtype C viruses is the predominant circulating variant in Limpopo Province as phylogenetic analyses of the partial *gag*, *pol*, and *env* C2-C3 gene fragments from HIV chronically infected patients showed that majority of the viruses are HIV-1 subtype C. The circulating HIV-1 viruses will be susceptible to the currently available protease inhibitors and reverse transcriptase inhibitors as drug resistance mutations in the naive population are very low (4.7%). Two unique recombinant forms HIV-1 A1/C and HIV-1 C/CRF11_cpx were each detected from two different individuals in the Waterberg and Capricorn districts respectively. There is no immunological difference between the HIV-1 subtype C viruses from Limpopo Province and the global consensus of subtype C viruses as epitope mapping using the consensus generated from test isolates had the same

dissociation constant as the subtype C consensus sequence. Also, there is apparently no significant difference on the impact of nucleotide polymorphisms on the genetic barrier to antiretroviral drug resistance between subtype C viruses and subtype B viruses.