Meta-Analysis of Observational Studies in Epidemiology

Medicine®

OPEN

# Significant contribution of subtype G to HIV-1 genetic complexity in Nigeria identified by a newly developed subtyping assay specific for subtype G and CRF02_AG

Richard A. Heipertz Jr, PhD[a], Ojor Ayemoba, FMCPath[b], Eric Sanders-Buell, BS[a,c], Kultida Poltavee, MS[a,c], Phuc Pham, MS[a,c], Gustavo H. Kijak, PhD[a,c], Esther Lei, MS[a,c], Meera Bose, MS[a,c], Shana Howell, BS[a,c], Anne Marie O'Sullivan, BS[a,c], Adam Bates, BS[a,c], Taylor Cervenka, MS, BSN[a,c], Janelle Kuroiwa, BS[a,c], Akindiran Akintunde, MBBS[d], Onyekachukwu Ibezim, MPH[b], Abraham Alabi, PhD[d,e], Obumneke Okoye, MBBS[b], Mark Manak, PhD[a,c], Jennifer Malia, DrPH[a,f], Sheila Peel, MPH, PhD[a], Mohammed Maisaka, FWACP[g], Darrell Singer, MD[f,h], Robert J. O'Connell, MD[a], Merlin L. Robb, MD[a,c], Jerome H. Kim, MD[a], Nelson L. Michael, MD, PhD[a], Ogbonnaya Njoku, MD[d], Sodsai Tovanabutra, PhD[a,c,*]

**Abstract**

While abundant sequence information is available from human immunodeficiency virus type 1 (HIV-1) subtypes A, B, C and CRF01_AE for HIV-1 vaccine design, sequences from West Africa are less represented. We sought to augment our understanding of HIV-1 variants circulating in 6 Nigerian cities as a step to subsequent HIV-1 vaccine development.

The G/CRF02_AG multi-region hybridization assay (MHA) was developed to differentiate subtype G, CRF02_AG and their recombinants from other subtypes based on 7 HIV-1 segments. Plasma from 224 HIV-1 infected volunteers enrolled in a cohort examining HIV-1 prevalence, risk factor, and subtype from Makurdi (30), Abuja (18), Enugu (11), Kaduna (12), Tafa (95), and Ojo/Lagos (58) was analyzed using MHA. HIV-1 genomes from 42 samples were sequenced to validate the MHA and fully explore the recombinant structure of G and CRF02_AG variants.

The sensitivity and specificity of MHA varied between 73–100% and 90–100%, respectively. The subtype distribution as identified by MHA among 224 samples revealed 38% CRF02_AG, 28% G, and 26% G/CRF02_AG recombinants while 8% remained nontypeable strains. In envelope (*env*) gp120, 38.84% of the samples reacted to a G probe while 31.25% reacted to a CRF02 (subtype A) probe. Full genome characterization of 42 sequences revealed the complexity of Nigerian HIV-1 variants.

CRF02_AG, subtype G, and their recombinants were the major circulating HIV-1 variants in 6 Nigerian cities. High proportions of samples reacted to a G probe in *env* gp120 confirms that subtype G infections are abundant and should be considered in strategies for global HIV-1 vaccine development.

*Editor: Angela Amedee.*

*GenBank Accession Numbers: KX389609 to KX389649.*

*Current address: William Beaumont Army Medical Center, Department of Pathology, El Paso, Texas (Richard A. Heipertz Jr).*

*Current address: Armed Forces Research Institute of Medical Research, Bangkok, Thailand (Robert J. O'Connell).*

*Current address: International Vaccine Institute, Seoul, Republic of South Korea (Jerome H. Kim).*

*Disclosure Statements: Parts of this manuscript have been presented at the AIDS Vaccine Meeting in Atlanta, United States of America (2010) and Bangkok, Thailand (2011).*

*Disclaimer: The opinions expressed herein are those of the authors and should not be construed as official or representing the views of the United States Department of Defense or Department of the Army. Mention of trade names, commercial products, or organizations does not imply endorsement by the United States government.*

*These studies were supported by the cooperative agreements (W81XWH-07–2–0067 and W81XWH-11–2–0174) between the Henry M. Jackson Foundation for the Advancement of Military Medicine and the Department of Defense.*

*The authors have declared that no conflicts of interest exist.*

*Supplemental Digital Content is available for this article.*

*[a] US Military HIV Research Program, Walter Reed Army Institute of Research, Silver Spring, Maryland, [b] Emergency Plan Implementation Committee, Nigerian Ministry of Defense, Abuja, Nigeria, [c] Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, Maryland, [d] US Military HIV Research Program (HJF-MRI), Abuja, Nigeria, [e] Centre de Recherches Medicales de Lambarene (CERMEL) Fondation Internationale de l'Hôpital Albert Schweitzer Lambarene, Gabon, [f] United States Public Health Service, Rockville, Maryland, [g] 45 Nigerian Air Force Hospital, Makurdi, Nigeria, [h] Department of Defense HIV Program, Abuja, Nigeria.*

*[*] Correspondence: Sodsai Tovanabutra, Walter Reed Army Institute of Research, Attn: MHRP/Retro, 503 Robert Grant Avenue, Room 2N25, Silver Spring, MD 20910 (e-mail: stovanabutra@hivresearch.org).*



*Medicine (2016) 95:32(e4346)*

*Received: 13 April 2016 / Received in final form: 27 June 2016 / Accepted: 30 June 2016*

*http://dx.doi.org/10.1097/MD.0000000000004346*

**Abbreviations:** ART = anti-retroviral treatment, CRF = circulating recombinant form, env = envelope, GTR = general time reversible, HIVdb = HIV-1 sequence database, HMA = heteroduplex mobility assay, jpHMM = jumping profile hidden Markov model, LANL = Los Alamos National Laboratory, MEGA = molecular evolutionary genetics analysis, MHA = multi-region hybridization assay, ML = maximum likelihood, pol = polymerase, RT-PCR = reverse transcription polymerase chain reaction.

**Keywords:** Genetic complexity, HIV-1, Nigeria, Recombinant, Subtypes

## 1. Introduction

Twenty-five of the 35 million people globally living with human immunodeficiency virus type 1 (HIV-1) infection in 2013 were in Sub-Saharan Africa.[1] Nigeria, the most populous country in Africa, was ranked second in the world for the highest HIV-1 burden.[2] The United Nations Development Program reported a population of 174 million, with a prevalence of 3.1% HIV-1 among 15 to 49-year-old individuals in Nigeria in 2013.[3] While recent antiretroviral treatment (ART) programs in Nigeria have significantly increased access to ART among the general population, incidence has not decreased substantially. HIV-1 related stigma and discrimination still pose significant obstacles to ART uptake, a major concern given that members of the most-at-risk populations and their partners contributed to as much as 40% of the 227,518 new infections in Nigeria in 2014.[2]

Current HIV-1 vaccines includes only subtypes A, B, C, and circulating recombinant form (CRF) 01_AE.[4–12] Following the demonstration that an HIV-1 vaccine tailored to a specific subtype is possible,[5] new approaches to creating a globally effective vaccine will need to take into account West African HIV-1 variants.[13] Given its large population size and number of infected individuals, characterizing Nigerian HIV-1 variants is important to accurately inform global vaccine development. The Los Alamos National Laboratory (LANL) HIV-1 Sequence Database is biased towards subtype B. There were only 1.01% of subtype G and 2.54% of CRF02_AG, the main circulating strains in Nigeria, deposited in LANL on 10 November 2015.

HIV-1 group M is responsible for most of global infections, and can be further classified into 13 subtypes and 72 CRFs.[14,15] In Nigeria, subtypes A, B, C, D, F2, G, J, and group O have been identified along with several CRFs in varying proportions.[16–27] However, subtypes G and CRF02_AG have been the dominant subtypes in Nigeria and throughout West Africa within the past decade.[14,16,19] These studies identified HIV-1 subtypes using a variety of laboratory methods such as heteroduplex mobility assays (HMA) and partial genome sequencing of the *env*, *gag*, and polymerase (*pol*) genes.[16–27] The distribution of HIV-1 variants in Nigeria seems to differ based on geography, as subtype G is most prevalent in the north and CRF02_AG in the south.[20,26] Our group has also observed this general subtype distribution and has identified subtypes in complex combinations using MHA and full-length sequencing.[28–30] In addition, Leye et al[31] has developed an MHA assay for HIV-1 subtypes in West Africa using a different approach.

Since 2008, Nigeria's South-South geopolitical region has had a higher HIV-1 prevalence than the North-Central region.[2,19] These shifts in HIV-1 prevalence and subtype gradients suggest differences in the epidemiological origins and migration patterns for subtypes G and CRF02_AG.[32] Contemporary studies of HIV-1 sequence variation in Nigeria are needed to better inform HIV-1 vaccine design and HIV-1 prevention and treatment strategies in Nigeria. In this study, we exploit our expertise in designing, validating, and field-testing a real-time polymerase chain reaction (PCR) genotyping assay capable of distinguishing HIV-1 subtypes G, CRF02_AG, and their recombinant forms to more accurately characterize the HIV-1 strains circulating in Nigeria. The performance of this newly developed assay was validated by full-length HIV-1 genomic sequencing of a subset of samples.

## 2. Methods

### 2.1. Study populations

A prevalence, risk factor and HIV-1 subtype characterization cohort study was conducted in Nigeria from August 2009 to March 2012, jointly by the Emergency Plan Implementation Committee of the Nigerian Ministry of Defense HIV Programme, the United States Military HIV Research Program, Walter Reed Army Institute of Research, and the Department of Defense HIV Program-Nigeria. A total of 3229 adult volunteer community members were enrolled at 6 geographically dispersed sites: Makurdi, Abuja, Enugu, Kaduna, Tafa, and Ojo/Lagos. Blood samples from the 3187 participants who completed the study were submitted for HIV testing. The site-specific HIV prevalence ranged from 3.2% in the Enugu site to 22.8% in the Tafa site (Makurdi 9.4%; Abuja 7.3%; Enugu 3.2%; Kaduna 4.9%; Tafa 22.8%, and Ojo/Lagos 15.7%). The overall prevalence rate was 10.2%. From 324 identified HIV-1 positive volunteers, plasma samples from 224 were of sufficient viral titer and purity to allow subtype characterization. These included 30 samples from Makurdi, 18 from Abuja, 11 from Enugu, 12 from Kaduna, 95 from Tafa, and 58 from Ojo Lagos. Demographic characteristics and HIV risk factors of the study participants will be described elsewhere (Njoku et al, in preparation).

### 2.2. Subtype G, CRF02_AG probe and primer design and testing

HIV-1 subtype G and CRF02_AG sequences from the LANL HIV-1 sequence database (HIVdb) were aligned.[14] The highly conserved nucleotide regions within a subtype and the most distinct regions between subtype G and CRF02_AG were examined across the genome to identify the sequences that would maximize the discrimination between subtypes G and CRF02_AG. The probes were designed to hybridize to the target sequences based on visual inspection and using the QuickAlign tool from the HIVdb.[14] The primers were designed to amplify all HIV-1 subtypes and their recombinants using the same tools. The specifics on the initial testing of primers and probes for each region have been tested using a panel of previously characterized HIV-1 full-genome deoxyribonucleic acid (DNA) as described previously.[30,33–35]

### 2.3. HIV-1 subtype G, CRF02-AG MHA

Viral ribonucleic acid (RNA) was extracted from the plasma of volunteers using the MagNA Pure robotic nucleic acid extraction procedure (Roche Diagnostics, Indianapolis, IN) following the

manufacturer's instructions. Extracted RNA was then used in reverse transcriptase polymerase chain reaction (reverse transcriptase (RT)-PCR). The RT-PCR mix contained QIAGEN OneStep RT-PCR 5× Buffer (Qiagen), 200 μM each dNTP, 400 nM of each outer primer, 4 units of RNasin Ribonuclease Inhibitor (Promega), 0.5 μL of QIAGEN OneStep RT-PCR Enzyme Mix (Qiagen), and 10 μL of the extracted RNA, resulting in a final volume of 25 μL per reaction. The thermocycle routine was an initial hold at 50 °C for 30 minutes, then 15 minute at 95 °C, followed by 30 cycles of 95 °C for 10 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute with a final extension at 72 °C for 10 minutes in an ABI 9600 thermocycler (Applied Biosystems Inc., Forest City, CA). An aliquot of the first-round RT-PCR product was then used for second-round PCR, also conducted in a Taqman real-time format, incorporating universal primers and fluorescent probes specific for subtype G and CRF02_AG. Probes were 5′-end labeled with FAM (6-carboxyfluorescein) and 3′-end labeled with black hole quencher 1. The second-round RT-PCR mixes contained TaqMan 2× Universal PCR Master Mix (Applied Biosystems Inc.), 400 nM of each inner primer, 250 nM of fluorescent probe, and 2.5 μL of the first-round PCR product, resulting in a final volume of 12.5 μL. RT-PCR amplification was performed in a 384-well ABI PRISM 7900HT Sequence Detection System (Applied Biosystems Inc.) using the following routine: hold at 95 °C for 10 minutes, 40 cycles of 95 °C for 15 seconds, and 60 °C for 1 minute.

Fluorescence intensity was monitored during the reaction and analyzed using the SDS v2.1 software (Applied Biosystems Inc.). Results were considered positive when the threshold cycle was lower than 35 cycles and if an exponential increase in the normalized fluorescence intensity for over 5 consecutive cycles was observed. For each sample, the PCR positive control of each reaction was conducted in parallel with the other reactions to assess whether negative RT-PCR results were due to the lack of sample amplification or probe hybridization. This control reaction contained 2× SybrGreen PCR Master Mix (Applied Biosystems Inc.) and similar volumes of inner primers and template. The identity of the amplicons was verified using melting curves (95 °C for 15 seconds, 30 °C for 15 seconds, and 95 °C for 15 seconds at 10% ramp rate) that distinguished the expected PCR products from primer-dimers of lower thermal stability.

### 2.4. Sequencing

Complementary DNA was synthesized from extracted plasma RNA and subjected to a nested PCR using the method previously described.[36] Sequence data were manually edited using Sequencher 5.0 (Gene Codes Corporation, Ann Arbor, MI).

### 2.5. Phylogenetic analysis

The HIV-1 genotype was first assigned for each sequence in this cohort using the HIV-1 Genotyping Tool at the National Center for Biotechnology Information.[37] Phylogenetic analyses of the sequences were performed using Molecular Evolutionary Genetics Analysis (MEGA) 5.0.[38] Reference sequences consisting of all subtype G and CRF02_AG along with relevant CRFs circulating in Africa were retrieved from the HIVdb. Cohort and reference sequences were aligned and manually edited. Maximum likelihood (ML) trees were constructed with the general time reversible (GTR) substitution model and gaps treated as partial deletion with a cutoff of 95% of site coverage. Trees were visualized using CLC Genomics Workbench 7.0.4 (CLC Bio, Cambridge, MA). Nucleotide distances among sequences were calculated based on the GTR substitution model. Recombinant analysis was performed using a combination of jumping profile hidden Markov model,[39] visual inspections, and construction of sub-region phylogenetic trees with bootstrap analyses to confirm breakpoint assignments.

### 2.6. Ethics statement

This study was reviewed and approved by the Walter Reed Army Institute of Research Institutional Review Board and termed RV230.

## 3. Results

### 3.1. Development of the MHA subtype G, CRF02_AG (MHA G, 02_AG)

The MHA was developed in two steps: probe/primer design and testing/evaluation of the specificity of probes and primers using HIV-1 full genome DNA derived from 42 samples in the cohort. Seven regions between 203 and 348 nucleotides long were chosen across the HIV-1 genome to allow for amplification of all HIV-1 subtypes and CRFs (Fig. 1). Probes for three regions, *pol* (RT), *pol* (integrase) and *tat*, were designed to react with both subtype G and CRF02_AG. The other four probes that were located in *gag*, *pol* (RT2), *vpr*, and *env* (gp120) differentiated between subtype G or CRF02_AG. Our own panel of 30 HIV-1 genomes corresponding to subtypes A (n = 4), B (n = 4), C (n = 4), D (n = 4), G (n = 6), CRF01_AE (n = 4), and CRF02_AG (n = 4) was used to validate probes and primers. Primers and probes are described in Supplemental Table 1, http://links.lww.com/MD/B198.

Overall performance of the assay was evaluated against 37 HIV-1 full-genome identified in this cohort: 12 subtype G, 14 CRF02_AG, 9 G/CRF02_AG recombinants, 1 G/K recombinant, and 1 CRF01_AE/CRF02_AG recombinant. Each region was analyzed individually for specificity and sensitivity (Table 1). The sensitivity to detect subtypes in the seven regions varied between 73% and 100% while the specificity was between 90% and 100%. The number of MHA reactive regions varied from 1 to 7 with the mode of 6 regions. The sensitivity of the assay with the number of reactive regions at 4, 5, 6, and 7 were 91.07%, 82.14%, 59.38%, and 28.57%, respectively. The number of MHA reactive regions at 4 was used as a cut-off point for subtype designation to maximize the sensitivity of the assay.

### 3.2. Distribution of HIV-1 strains in Nigeria by MHA

The MHA identified 84 (38%) CRF02_AG, 62 (28%) subtype G, 58 (26%) G/CRF02_AG recombinants and 20 (8%) nontypeable strains out of 224 samples from 6 study sites in Nigeria. The HIV-1 subtype distributions differed among the study sites as shown in Fig. 2. Among the 30 samples from Makurdi, 12 (40%) were CRF02_AG with 7 (23%) each of subtype G and G/CRF02_AG, and 4 (14%) were nontypeable. Abuja had 18 samples of which were 7 (39%) CRF02_AG, 5 (28%) subtype G, 4 (22%) G/CRF02_AG recombinants, and 2 (11%) nontypeable. Eleven samples collected in Enugu contained 3 (27%) CRF02_AG, 4 (37%) subtype G, 2 (18%) G/CRF02_AG recombinants and 2 (18%) nontypeable. Kaduna had the highest proportion of CRF02_AG at 8 (67%) as compared with the other study sites, 2 (17%) subtype G, and 1 each (8%) of G/CRF02_AG and nontypeable. Tafa, the study site with the highest HIV-1

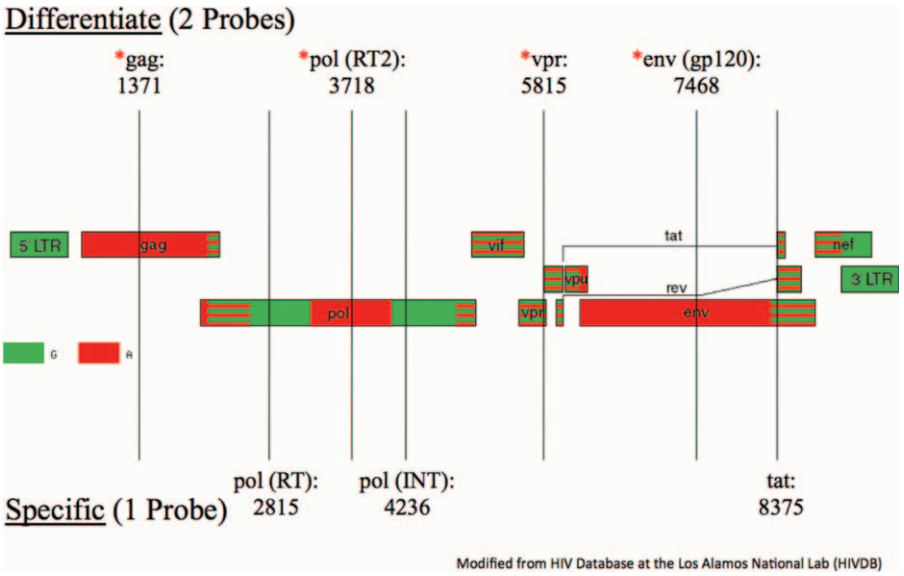



**Figure 1.** MHA G, 02_AG primer/probe region locations. The 7 regions of the MHA were chosen using the LANL HIVdb. Potential regions were visually identified and testing occurred on samples of known sequence stored at the MHRP. Regions that differentiate between subtype G and CRF02_AG contain two specific probes for each and are designated with a red *. Specific regions indicate those where a single probe identifies both subtype G and CRF02_AG. Numbers under the identified regions indicate the center of the probe hybridization HXB2 location.

prevalence and the largest number of samples collected (95), had 27 (28%) CRF02_AG, 34 (36%) subtype G, 29 (31%) G/CRF02_AG recombinants, and 5 (5%) nontypeable. Fifty-eight samples from Ojo/Lagos comprised of 27 (47%) CRF02_AG, 10 (17%) subtype G, 15 (26%) G/CRF02_AG strains, and 6 (10%) nontypeable. Logistic regression model analysis of our data did not reveal a link between the HIV-1 prevalence rate and subtype distribution at each site. Interestingly, among 224 samples, 87 (38.84%) were reactive with the subtype G probe while 70 (31.25%) were reactive to the CRF02_AG (subtype A) probe in *env* gp120 region.

### 3.3. Analysis of HIV-1 full genome

Forty-two of the 224 plasma samples were subjected to full-genome sequencing to validate the MHA and resolve some nontypeable samples. The demographic characteristics, risk factor, genetic subtypes, and sample collection dates are shown in Supplemental Table 2, http://links.lww.com/MD/B198. These infections came from all study sites: 18 from Makurdi, 11 from Abuja, 2 from Enugu, 1 from Kaduna, 6 from Tafa, and 4 from Ojo/Lagos. These included 9 men and 33 women all of whom had sexual exposure as their primary risk. Their ages ranged from 20 to 40 years. The median HIV-1 plasma viral load of these HIV-infected individuals was 37,851 copies/mL and ranged from 2203 to 316,940 copies/mL.

An ML phylogenetic tree of the 42 sequences of interest and 94 reference sequences from subtype A, C, G, CRF02_AG, CRF06_cpx, and CRF11_cpx identified from different parts of the world was constructed as shown in Fig. 3. The results revealed that these strains were 14 CRF02_AG, 12 subtype G, 2 subtype A, 1 subtype C, 9G/CRF02 unique recombinants, 1 CRF06_cpx, 1 CRF11_cpx, 1 G/K unique recombinant, and 1 CRF02_AG/CRF01_AE recombinant. CRF02_AG strains from Nigeria were dispersed among all CRF02_AG strains mostly identified from Nigeria, Cameroon, Ghana, and Senegal, except for 2 strains (010170 and 020382) which formed their own cluster. Subtype G strains from this cohort also mixed with other references previously found in Nigeria and Cameroon within the G cluster. Subtype A strains, 010499 and 040248, formed a monophyletic group that clustered with subtype A references from Kenya, Tanzania, and Rwanda. One subtype C identified from Abuja (strain 020523) formed a cluster with other subtype C references from India, South Africa, Kenya, Botswana, Senegal, and Tanzania. Strain 030498 clustered with 6 previously identified CRF06_cpx references and 1 CRF11_cpx was identified, strain 010131, which clustered with other CRF11_cpx from Cyprus. Breakpoint analysis and subtype assignment of 9 unique G/CRF02_AG recombinant strains (Fig. 4) revealed that the genome structures of these recombinants were either simple or very complex with breakpoint(s) ranging from 1 to 8. Strains 010073 and 010145 shared two common breakpoints and their genome structures were quite similar but not identical. Strains 010129 and 010483 also shared one breakpoint in the envelope region. Strain 010067 was a unique recombinant containing mostly subtype G except for a fragment of subtype K in the *pol* gene (partial p51 RT and partial p15 RT) corresponding to HXB2 number 3479 to 4066. The last unique recombinant identified by sequencing was a recombinant between CRF02_AG and CRF01_AE that had the last part of the genome as CRF01_AE spanning from the middle of *nef* gene through 3′LTR corresponding to HXB2 number 9023 to 9496. Supplemental Figures 1–11, http://links.lww.com/MD/B198 show the sub-trees generated for each of the segments of the 11 recombinant species that were analyzed.

We compared the mean nucleotide distances among the subtype G and CRF02_AG sequences we identified to a set of contemporary sequences from the HIVdb (Supplemental Table 3, http://links.lww.com/MD/B198). For both subtype G and CRF02_AG, the mean diversity calculated in *gag*, *pol*, *env*, and *nef* genes among our sequences did not differ from that calculated among previously sampled sequences in the HIVdb, showing that our dataset was representative of contemporary circulating sequences in Nigeria.

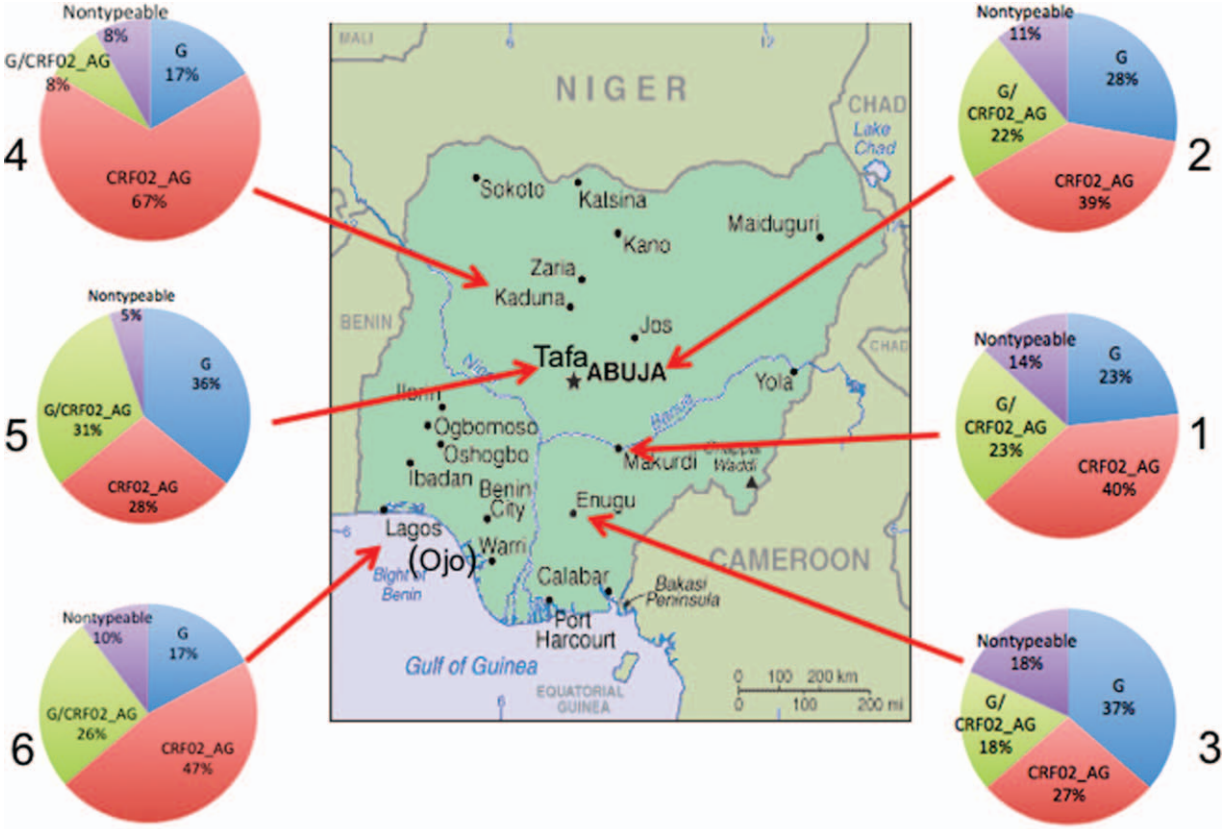
**Table 1**

**Sensitivity and specificity of MHA G, 02_AG on known sequence cohort samples.**

| Sample name | FL subtype | GAG: G | GAG: 02_AG | RT | RT2: G | RT2: 02_AG | INT | VPR: G | VPR: 02_AG | GP120: G | GP120: 02_AG | TAT | Multi-region hybridization assay (MHA) subtype |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 010079 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 010157 | G | G | | G+CRF02 | G | | G+CRF02 | Nonreactive | | G | | G+CRF02 | G |
| 010261 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 020133 | G | G | | G+CRF02 | G | | G+CRF02 | Nonreactive | | G | | G+CRF02 | G |
| 020303 | G | G | | G+CRF02 | Nonreactive | | G+CRF02 | Nonreactive | | G | | G+CRF02 | G |
| 010315 | G | G | | Nonreactive | G | | G+CRF02 | G | | G | | Nonreactive | G |
| 020134 | G | G | | Nonreactive | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 050158 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 050272 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 050489 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 060248 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 060409 | G | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 010060 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 010170 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | Nonreactive | G+CRF02 | CRF02_AG |
| 010181 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 020065 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | Nonreactive | | CRF02 | G+CRF02 | CRF02_AG |
| 020307 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 020382 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | Nonreactive | G+CRF02 | CRF02_AG |
| 020437 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | Nonreactive | G+CRF02 | CRF02_AG |
| 010325 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 030161 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 050135 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 050356 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 050366 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 060304 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 060418 | CRF02_AG | | CRF02 | G+CRF02 | | CRF02 | G+CRF02 | | CRF02 | | CRF02 | G+CRF02 | CRF02_AG |
| 010105 | G/CRF02_AG Recomb. | G | | G+CRF02 | G | | G+CRF02 | Miscalled G | Nonreactive | G | | G+CRF02 | G |
| 010205 | G/CRF02_AG Recomb. | G | | G+CRF02 | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 020420 | G/CRF02_AG Recomb. | G | | Nonreactive | G | | G+CRF02 | G | | G | | G+CRF02 | G |
| 010073 | G/CRF02_AG Recomb. | G | | G+CRF02 | G | | G+CRF02 | | CRF02 | G | | G+CRF02 | G/CRF02_AG Recomb. |
| 010145 | G/CRF02_AG Recomb. | G | | G+CRF02 | G | | G+CRF02 | | CRF02 | G | | G+CRF02 | G/CRF02_AG Recomb. |
| 010067 | G/K Recomb. | G | | G+CRF02 | Not called* | Not called* | G+CRF02 | G | | G | | G+CRF02 | G |
| 010208 | G/CRF02_AG Recomb. | G | | G+CRF02 | | CRF02 | G+CRF02 | G | | G | | G+CRF02 | G/CRF02_AG Recomb. |
| 020545 | CRF02_AG/CRF01_AE Recomb. | | CRF02 | Nonreactive | | CRF02 | G+CRF02 | | CRF02 | | Nonreactive | G+CRF02 | CRF02_AG |
| 010129 | G/CRF02_AG Recomb. | | CRF02 | G+CRF02 | | Nonreactive | G+CRF02 | Miscalled G | Nonreactive | G | | G+CRF02 | G/CRF02_AG Recomb. |
| 020468 | G/CRF02_AG Recomb. | | CRF02 | Nonreactive | G | | G+CRF02 | | Nonreactive | G | | G+CRF02 | G/CRF02_AG Recomb. |
| 010483 | G/CRF02_AG Recomb. | | CRF02 | G+CRF02 | G | | G+CRF02 | Nonreactive | | Nonreactive | | G+CRF02 | G/CRF02_AG Recomb. |
| | % Sensitivity | 100.00 | 100.00 | 86.49 | 94.74 | 94.12 | 100.00 | 76.47 | 80.00 | 95.45 | 73.33 | 97.30 | |
| | % Specificity | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 90.00 | 100.00 | 100.00 | 100.00 | 100.00 | |

*RT2 portion belongs to subtype K and was not amplified.

FL = full-length, GAG:G = group-specific antigen subtype G, RT = reverse transcriptase, RT2:G = reverse transcriptase second region subtype G, TAT = trans-activator of transcription, VPR:G = viral protein R subtype G.




| Site : Name | Prevalence Rate | HIV-1 Subtype | | | | Total |
|---|---|---|---|---|---|---|
| | | G | CRF02_AG | G/CRF02_AG | Nontypeable | |
| 1 : Makurdi | 9.4% | 7 | 12 | 7 | 4 | 30 |
| 2 : Abuja | 7.3% | 5 | 7 | 4 | 2 | 18 |
| 3 : Enugu | 3.2% | 4 | 3 | 2 | 2 | 11 |
| 4 : Kaduna | 4.9% | 2 | 8 | 1 | 1 | 12 |
| 5: Tafa | 22.8% | 34 | 27 | 29 | 5 | 95 |
| 6 : Ojo Lagos | 15.7% | 10 | 27 | 15 | 6 | 58 |
| Total | 10.2% | 62 | 84 | 58 | 20 | 224 |

**Figure 2.** Distribution of the HIV-1 subtype in Nigeria. HIV-1 subtypes were identified using the MHA G, CRF02_AG. Total number of each subtype and the percentages for each site are noted in the pie charts surrounding the map and identified by color: Blue = subtype G; Red = CRF02_AG; Green = subtype G/CRF02 recombinant; Purple = nontypeable by MHA. HIV-1 prevalence rates and actual sample numbers per subtype, per site are shown in the chart below the map.

## 4. Discussion

In this study, we have developed a rapid, RT-PCR based MHA to identify and distinguish between subtype G, CRF02_AG, and their recombinants. This subtyping assay is very effective in screening large-scale populations in the geographic regions where CRF02_AG, subtype G, and their recombinant strains

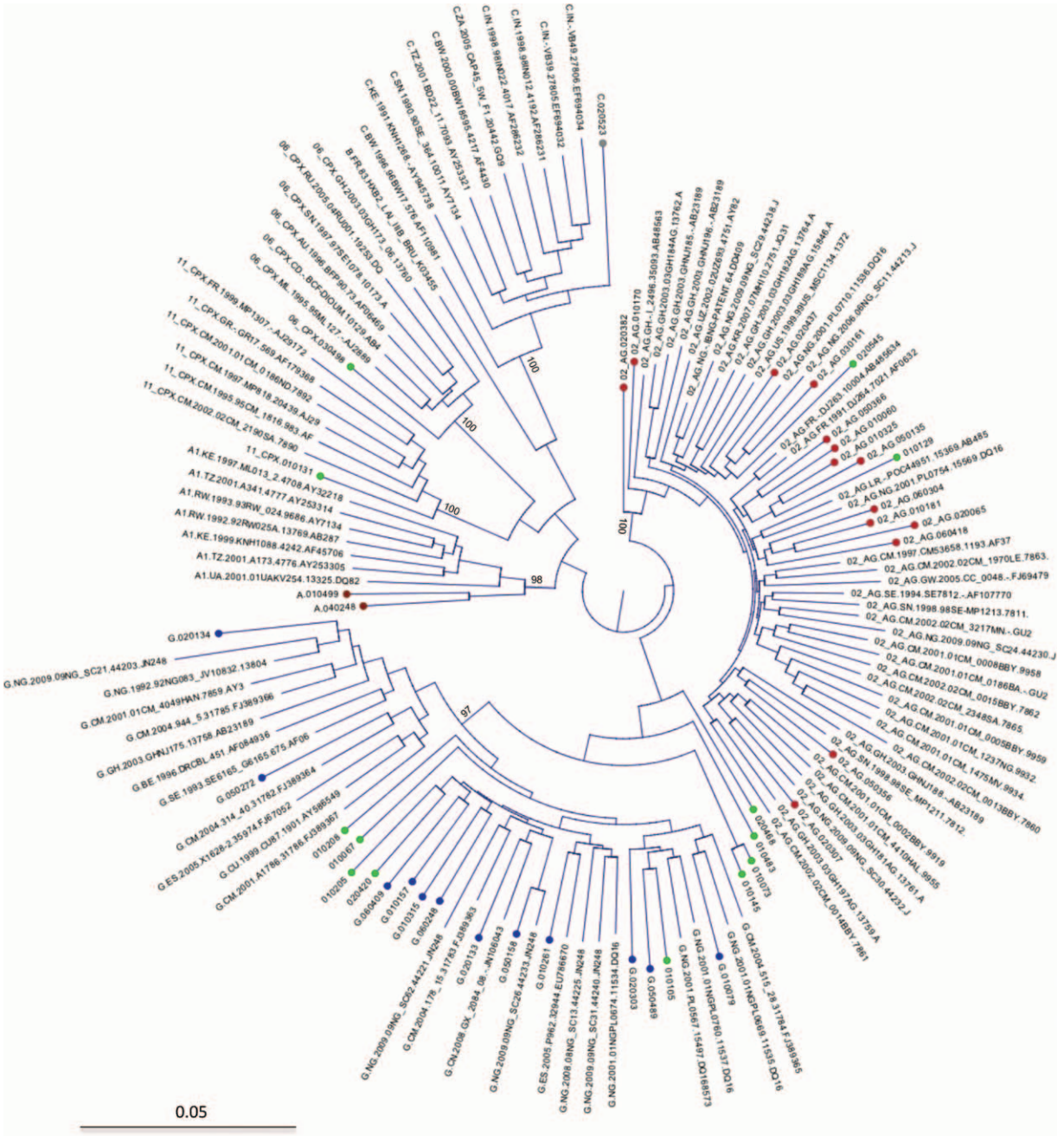



**Figure 3.** Full-length HIV-1 ML phylogenetic tree containing the 42 Nigerian sequences from this cohort. The sequences from this cohort along with 94 reference sequences from subtype A, C, G, CRF02_AG, CRF06_cpx, and CRF11_cpx identified from different parts of the world was constructed. Cohort sequences are identified by colored dot: Blue = subtype G; Red = CRF02_AG; Green = subtype G/CRF02 recombinants; Maroon = subtype A; Gray = subtype C; Green in the top left portion of the tree = CRF06_cpx and CRF11_cpx.

co-circulate. The assay was validated using known full-genome sequences from various pure subtypes and revealed 100% probe specificity on subtype G and CRF02_AG. The results from MHA and full genome sequencing of CRF02_AG, subtype G, and CRF02_AG/G recombinants were in agreement indicating that the performance of the MHA G/CRF02_AG assay is as accurate as the other MHA assays previously developed for Thailand and East Africa by our group.[30,33] In comparison to HMA and other partial genome sequencing techniques used in previous studies,[16–20,26,27] this MHA G/CRF02_AG assay is more powerful due to its ability to screen for the genetic subtype across the entire HIV-1 genome.

Subtype G was reported by Peeters et al[26] to be dominant in the north (Kano) and co-circulated at the similar proportion with CRF02_AG in the northeastern region of Maiduguri during 1994 to 1996. In 2002, Agwale et al,[17] reported that the overall proportion of HIV-1 strains from 36 Nigerian states characterized in gp41 were 54% subtype G, 45% subtype A, 0.4% subtype C, 0.4% subtype J, and 0.4% was unclassifiable. Sankale et al,[27] characterized the C2V3 region of *env* and partial *gag* genes from

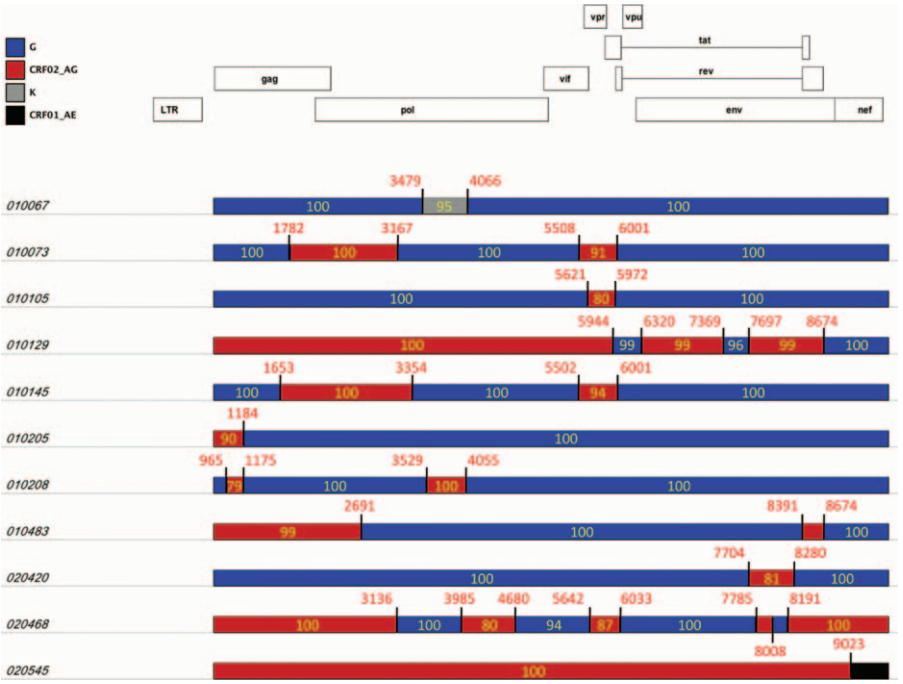



**Figure 4.** Analysis of cohort recombinants identified by full genome sequencing. Eleven of the 42 sequenced samples from this cohort were unique recombinant viruses. Breakpoint analysis of the recombinants was conducted and ranged from 2 to 9 different recombinant pieces. Bootstrap values greater than 70 are labeled inside their corresponding sequence regions. Region subtype is based on color: Blue = subtype G; Red = CRF02_AG; Gray = subtype K; Black = CRF01_AE.

samples collected in Ibadan and Saki, between the years 2002 and 2003 and found subtype G and CRF02_AG circulated at equal proportions of 32% as well as 6% CRF06_cpx, 2.3% subtype A, 1% subtype C, and 1% unclassified. Recombinants involving CRF02_AG and/or subtype G (20.7%) or other subtypes (2.3%) were also reported. Additionally, it was observed that the predominant strain in these two locations were different: subtype G in Ibadan and CRF02_AG in Saki which are only 100 miles apart.[27] The findings from our study using the MHA assay and full-genome sequencing show that CRF02_AG and subtype G were the main circulating strains in Nigeria, along with a substantial number of their recombinants, a few subtype A and C, and rare CRFs were not different from the previous findings during 1994 to 2012.

The relative proportion of subtypes reported from this cohort, 28% subtype G, 38% CRF02_AG, and 26% of their recombinants along with 2 subtype A, 1 subtype C, 1 CRF06_cpx, 1 CRF11_cpx, and 1 unique recombinant form G/K are quite different from previous reports.[16,17,19,20,26–28] In contrast to most previous studies, our samples were collected from a larger geographic area covering South and Central Nigeria, and thus, are more representative of the Nigerian epidemic. In addition, our subtype identification by full-genome MHA can more accurately characterize subtype G/CRF02_AG recombinants than analyses restricted to a specific gene region which miss these recombinants. Our data also demonstrated multiple recombinants circulating in the HIV-positive population in Nigeria. Recombinant genomes resulted from a single recombination event to greater than 8 events, highlighting the global HIV epidemic complexity as indicated by the growing numbers of CRF and unique recombinant forms in the LANL HIVdb.[14] The similar genetic distances of subtype G and CRF02_AG identified in this cohort and previously identified sequences from the LANL HIVdb suggested that our sampling was reflective of the diversity found in Nigeria. One of our MHA probes was located in the *env* gp120 region of the HIV genome. Fortunately, there was enough conserved nucleotide sequence in the region allowing PCR amplification of all HIV-1 subtypes and CRFs tested, while containing enough variance in sequence to allow for the accurate identification of subtype G or CRF02_AG in that region. Probe reactivity in that region indicated that only 31.25% (70 out of 224) of samples were reactive to the CRF02 (subtype A) probe in the *env* gp120 region. This suggests that almost 70% of Nigerian sequences in our study are of non-subtype A origin within *env*. This proves problematic as global vaccine concepts either include the subtypes A, B, C and CRF01_AE for insert design or mosaic inserts derived from published sequences in databases with low West African subtype representation.[4–13] The conventional view that West African HIV variants are largely CRF02_AG would justify the inclusion of subtype A sequences in a global vaccine design. However, our findings that subtype G and unique recombinant forms of subtype G/CRF02_AG are a significant burden in the Nigerian epidemic and the neighboring West African countries requires consideration for future global vaccine development.

## Acknowledgments

We are grateful to the study volunteers for their participation, Michelle Lazaro and Andrea Bradfield for their technical assistance, Silvana Tasca-Sina for her helpful discussions with analytical design and conclusions, and Erik Billings for submitting the nucleotide sequences to GenBank. We also thank the study team in Nigeria for collecting samples and data from the participants. We also like to thank Morgane Rolland for her critical reading of the manuscript and input.

## References


[1] United Nations AIDS. Fact Sheet 2015. http://www.unaids.org/sites/default/files/media_asset/20150901_FactSheet_2015_en.pdf. Accessed December 28, 2015.

[2] National Agency for the Control of AIDS. Nigeria Global AIDS Response Country Progress Report, 2015. http://www.unaids.org/sites/default/files/country/documents/NGA_narrative_report_2015.pdf. Accessed December 28, 2015.

[3] Malik, K. The Rise of the South: Human Progress in a Diverse World. http://hdr.undp.org/sites/default/files/reports/14/hdr2013_en_complete.pdf. Accessed December 28, 2015.

[4] Excler JL, Robb ML, Kim JH. Prospects for a globally effective HIV-1 vaccine. Vaccine 2015;33(suppl 4):D4–12.

[5] Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med 2009;361:2209–20.

[6] Barouch DH, Stephenson KE, Borducchi EN, et al. Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. Cell 2013;155:531–9.

[7] Pitisuttithum P, Gilbert P, Gurwith M, et al. Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. J Infect Dis 2006;194:1661–71.

[8] Flynn NM, Forthal DN, Harro CD, et al. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J Infect Dis 2005;191:654–65.

[9] Guenaga J, de Val N, Tran K, et al. Well-ordered trimeric HIV-1 subtype B and C soluble spike mimetics generated by negative selection display native-like properties. PLoS Pathog 2015;11:e1004570.

[10] Pugach P, Ozorowski G, Cupo A, et al. A native-like SOSIP.664 trimer based on an HIV-1 subtype B env gene. J Virol 2015;89:3380–95.

[11] Buchbinder SP, Mehrotra DV, Duerr A, et al. Efficacy assessment of a cell-mediated immunity HIV-1 vaccine (the Step Study): a double-blind, randomised, placebo-controlled, test-of-concept trial. Lancet (London, England) 2008;372:1881–93.

[12] Gray GE, Allen M, Moodie Z, et al. Safety and efficacy of the HVTN 503/Phambili study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. Lancet Infect Dis 2011;11:507–15.

[13] Barouch DH, Michael NL. Accelerating HIV-1 vaccine efficacy trials. Cell 2014;159:969–72.

[14] Los Alamos National Laboratory. HIV Sequence Database. http://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html. Accessed December 28, 2015.

[15] Taylor BS, Sobieszczyk ME, McCutchan FE, et al. The challenge of HIV-1 subtype diversity. New Engl J Med 2008;358:1590–602.

[16] Diallo K, Zheng DP, Rottinghaus EK, et al. Viral genetic diversity and polymorphisms in a Cohort of HIV-1-infected patients eligible for initiation of antiretroviral therapy in Abuja, Nigeria. AIDS Res Hum Retroviruses 2015;31:564–75.

[17] Agwale SM, Zeh C, Robbins KE, et al. Molecular surveillance of HIV-1 field strains in Nigeria in preparation for vaccine trials. Vaccine 2002;20:2131–9.

[18] Ajoge HO, Gordon ML, de Oliveira T, et al. Genetic characteristics, coreceptor usage potential and evolution of Nigerian HIV-1 subtype G and CRF02_AG isolates. PLoS ONE 2011;6:e17865.

[19] Ajoge HO, Gordon ML, Ibrahim S, et al. Drug resistance pattern of HIV type 1 isolates sampled in 2007 from therapy-naive pregnant women in North-Central Nigeria. AIDS Res Hum Retroviruses 2012;28:115–8.

[20] Chaplin B, Eisen G, Idoko J, et al. Impact of HIV type 1 subtype on drug resistance mutations in Nigerian patients failing first-line therapy. AIDS Res Hum Retroviruses 2011;27:71–80.

[21] Charurat M, Nasidi A, Delaney K, et al. Characterization of acute HIV-1 infection in high-risk Nigerian populations. J Infect Dis 2012;205:1239–47.

[22] Hawkins CA, Chaplin B, Idoko J, et al. Clinical and genotypic findings in HIV-infected patients with the K65R mutation failing first-line antiretroviral therapy in Nigeria. J Acquir Immun Defic Syndr 2009;52:228–34.

[23] Montavon C, Toure-Kane C, Nkengasong JN, et al. CRF06-cpx: a new circulating recombinant form of HIV-1 in West Africa involving subtypes A, G, K, and J. J Acquir Immun Defic Syndr 2002;29:522–30.

[24] Ojesina AI, Mullins C, Imade G, et al. Characterization of HIV type 1 reverse transcriptase mutations in infants infected by mothers who received peripartum nevirapine prophylaxis in Jos, Nigeria. AIDS Res Hum Retroviruses 2007;23:1587–92.

[25] Ojesina AI, Sankale JL, Odaibo G, et al. Subtype-specific patterns in HIV Type 1 reverse transcriptase and protease in Oyo State, Nigeria: implications for drug resistance and host response. AIDS Res Hum Retroviruses 2006;22:770–9.

[26] Peeters M, Esu-Williams E, Vergne L, et al. Predominance of subtype A and G HIV type 1 in Nigeria, with geographical differences in their distribution. AIDS Res Hum Retroviruses 2000;16:315–25.

[27] Sankale JL, Langevin S, Odaibo G, et al. The complexity of circulating HIV type 1 strains in Oyo state, Nigeria. AIDS Res Hum Retroviruses 2007;23:1020–5.

[28] McCutchan FE, Carr JK, Bajani M, et al. Subtype G and multiple forms of A/G intersubtype recombinant human immunodeficiency virus type 1 in Nigeria. Virology 1999;254:226–34.

[29] Carr JK, Salminen MO, Albert J, et al. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology 1998;247:22–31.

[30] Kijak GH, Sanders-Buell E, Wolfe ND, et al. Development and application of a high-throughput HIV type 1 genotyping assay to identify CRF02_AG in West/West Central Africa. AIDS Res Hum Retroviruses 2004;20:521–30.

[31] Vidal N, Diop H, Montavon C, et al. A novel multiregion hybridization assay reveals high frequency of dual inter-subtype infections among HIV-positive individuals in Cameroon, West Central Africa. Infect Genet Evol 2013;14:73–82.

[32] Delatorre E, Mir D, Bello G. Spatiotemporal dynamics of the HIV-1 subtype G epidemic in West and Central Africa. PLoS ONE 2014;9:e98908.

[33] Hoelscher M, Dowling WE, Sanders-Buell E, et al. Detection of HIV-1 subtypes, recombinants, and dual infections in east Africa by a multi-region hybridization assay. AIDS (London, England) 2002;16:2055–64.

[34] Hierholzer J, Montano S, Hoelscher M, et al. Molecular epidemiology of HIV Type 1 in Ecuador, Peru, Bolivia, Uruguay, and Argentina. AIDS Res Hum Retroviruses 2002;18:1339–50.

[35] Kijak GH, Tovanabutra S, Sanders-Buell E, et al. Distinguishing molecular forms of HIV-1 in Asia with a high-throughput, fluorescent genotyping assay, MHAbce v.2. Virology 2007;358:178–91.

[36] Heipertz RAJr, Sanders-Buell E, Kijak G, et al. Molecular epidemiology of early and acute HIV type 1 infections in the United States Navy and Marine Corps, 2005-2010. AIDS Res Hum Retroviruses 2013;29:1310–20.

[37] Rozanov M, Plikat U, Chappey C, et al. A web-based genotyping resource for viral sequences. Nucleic Acids Res 2004;32:W654–9.

[38] Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28:2731–9.

[39] Schultz AK, Zhang M, Bulla I, et al. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res 2009;37:W647–651.