AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 22, Number 8, 2006, pp. 770–779

Subtype-Specific Patterns in HIV Type 1 Reverse Transcriptase and Protease in Oyo State, Nigeria: Implications for Drug Resistance and Host Response

AKINYEMI I. OJESINA,[1] JEAN-LOUIS SANKALÉ,[1] GEORGINA ODAIBO,[2] STANLEY LANGEVIN,[1] SEEMA T. MELONI,[1] ABDOULAYE DIENG SARR,[1] DAVID OLALEYE,[2] and PHYLLIS J. KANKI[1]

ABSTRACT

As the use of antiretroviral therapy becomes more widespread across Africa, it is imperative to characterize baseline molecular variability and subtype-specific peculiarities of drug targets in non-subtype B HIV-1 infection. We sequenced and analyzed 35 reverse transcriptase (RT) and 43 protease (PR) sequences from 50 therapy-naive HIV-1-infected Nigerians. Phylogenetic analyses of RT revealed that the predominant viruses were CRF02_AG (57%), subtype G (26%), and CRF06_cpx (11%). Six of 35 (17%) individuals harbored primary mutations for RT inhibitors, including M41L, V118I, Y188H, P236L, and Y318F, and curiously three of the six were infected with CRF06_cpx. Therefore, CRF06_cpx drug-naive individuals had significantly more drug resistance mutations than the other subtypes ($p = 0.011$). By combining data on quasisynonymous codon bias with the influence of the differential genetic cost of mutations, we were able to predict some mutations, which are likely to predominate by subtype, under drug pressure. Some subtype-specific polymorphisms occurred within epitopes for HLA B7 and B35 in the RT, and HLA A2 and A*6802 in PR, at positions implicated in immune evasion. Balanced polymorphism was also observed at predicted serine-threonine phosphorylation sites in the RT of subtype G viruses. The subtype-specific codon usage and polymorphisms observed suggest the involvement of differential pathways for drug resistance and host-driven viral evolution in HIV-1 CRF02_AG, subtype G, and CRF06_cpx, compared to subtype B. Subtype-specific responses to HIV therapy may have significant consequences for efforts to provide effective therapy to the populations infected with these HIV-1 subtypes.

INTRODUCTION

THE GENETIC DIVERSITY OF HIV-1 continues to challenge our efforts at combating the HIV/AIDS pandemic. This variability has been attributed to the lack of proofreading ability of the reverse transcriptase enzyme,[1,2] the rapid turnover of virions,[3,4] recombination of viruses,[5] and selective immunologic pressures.[6] To date, 9 subtypes and 16 circulating recombinant forms (CRFs) have been described.[7] In North America and Europe, the most prevalent form is the subtype B. However, in Africa, where the disease is the most prevalent, non-B subtypes predominate, with all 9 subtypes and many CRFs described across the continent. In West Africa, the most prevalent form is CRF02_AG, originally described in 1994 in Nigeria.[8] Since then, studies involving Nigerian subjects have shown that subtype G and various forms of a recombinant of subtypes A and G are prevalent in the country.[9–12]

Widespread use of highly active antiretroviral therapy (HAART) has greatly reduced morbidity and mortality in industrialized nations.[13] Drugs most commonly used to target the enzymes reverse transcriptase (RT) and protease (PR) belong to three classes—nucleoside analogue RT inhibitors (NRTIs), non-NRTIs (NNRTIs), and protease inhibitors (PIs). Antiretroviral (ARV) therapy is increasingly playing a major role in controlling the pandemic as many countries in which non-B subtypes abound have recently commenced treatment programs,[14–16] with more people having access to drugs. There is a need to generate data on the efficacy of ARV, especially in the treatment of non-B infections.

One of the major obstacles to effective treatment is the de-

[1]Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115.
[2]Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria.

velopment of resistance mutations to various antiretroviral drugs. It is imperative to understand if mutations known to confer resistance to ARV drugs in subtype B infection would also confer resistance to non-B subtypes. Various studies have described genotypic and phenotypic resistance in non-subtype B infections.[16–23] Most of the studies described the lack of major drug resistance-conferring mutations, but a high prevalence of secondary mutations, although a recent study[16] reported considerable overlap between subtype B resistance mutations and mutations associated with at least one non-B subtype.

Genetic variation within the RT and PR may greatly influence viral replication and fitness, as well as susceptibility to therapy and the development of drug resistance.[24] It is conceivable that the widespread use of ARVs will exert extra selective pressure on the RT and PR and that this pressure will express itself as mutations differing by subtype. It is also possible that preexisting polymorphisms may affect the pathways to drug resistance. Thus, it is important to understand the baseline genetic variability of these genes, especially in non-subtype B viruses in order to elucidate the complex relationships between intersubtype polymorphisms and antiretroviral drug susceptibility.

This study was carried out to characterize the molecular diversity of HIV-1 PR and RT sequences in a Nigerian population, to identify the baseline drug resistance-related mutations (DRMs) and polymorphisms that exist in a population of ARV-naive individuals, and to define subtype-specific patterns with potential functional significance.

MATERIALS AND METHODS

Study population

The AIDS Prevention Initiative in Nigeria (APIN), at the Harvard School of Public Health, has been supporting community-based HIV-1 surveillance in Saki and Ibadan, a town and city, respectively, in Oyo State, Nigeria. The survey identified subpopulations engaging in high-risk behaviors within some organized artisans and trade groups in the communities (David Olaleye, unpublished data). The study was initiated prior to the commencement of the national ARV therapy program in Nigeria and received ethical approval from the Institutional Review Boards of both the Harvard School of Public Health and the University of Ibadan/University College Hospital, Ibadan in accordance with 45 CFR 46. The subjects were recruited and enrolled following written informed consent. Fifty high-risk HIV-1-infected subjects previously determined by gag and env gene amplification to be HIV-1 positive without known prior antiretroviral therapy were selected for the analysis in this study.

DNA amplification and sequencing by PCR

Proviral DNA was extracted from peripheral blood mononuclear cells with the Qiagen QIAmp Blood Midi kit (Qiagen, Inc., Chatsworth, CA). The entire 297 bp of PR and the first 1320 bp of RT (p51RT) were amplified by nested PCR using the following primers: OJ1: 5′-AAATGATGACAGCATGTCAGGGAG-3′, OJ2: 5′-TATCTACTTGTTCATTTCCTCCAAT-3′, OJ3: 5′-AGACAGGCTAATTTTTTAGGGA-3′, OJ4: 5′-CATTCCTGGCTTTAATTTTACTGG-3′, OJ5: 5′-CCAAAAATGATAGGGGGAATTGG-3′, and OJ6: 5′-CCTTTGTGTGCTGGTACCCATG-3′.

Primers OJ1 and OJ2 were used for first round amplification of a portion of the pol gene encoding both PR and the p51RT regions. Primers OJ3 and OJ4 were used for second round amplification of the PR region, while OJ5 and OJ6 were used to amplify the p51RT region. Polymerase chain reaction (PCR) reactions were carried out in a 100-μl volume reaction mixture containing 1 μg of DNA, 1× PCR buffer II (Roche Diagnostics, Alameda, CA), 2.5 units of Taq polymerase (Applied Biosystems, Foster City, CA), 0.2 mM of each dNTP, 2.5mM $MgCl_2$. Each reaction was subjected to 30 cycles of denaturation (1 min at 94°C), annealing (1 min at 46°C and 45°C, for the first and second round, respectively), and extension (1 min at 72°C) followed by a final extension (10 min at 72°C). Deionized water was used as the negative control for all reactions.

The PCR products were purified by 1.5% agarose gel electrophoresis and by using purification columns (gel extraction kit; Qiagen, Valencia, CA). The purified RT fragment products were subsequently cloned into the pCR2.1 vector (T/A Cloning; Invitrogen, San Diego, CA). The plasmid preparation for the double-stranded DNA sequencing was performed by alkaline lysis with silica columns (S.N.A.P. Mini-Prep kit; Invitrogen). The purified PR and plasmid-encoded RT samples were sequenced by dye terminator cycle sequencing using *Taq* polymerase (Perkin-Elmer Applied Biosystem Division, Foster City, CA) and an automated sequencer (ABI 377; Perkin-Elmer Applied Biosystem Division, Foster City, CA).

Sequence analysis

For all generated RT and PR consensus sequences, a multiple alignment was performed using Clustal X.[25] Reference sequences from the HIV sequence database of the Los Alamos National Laboratory were included. We used the Modeltest 3.6 program[26] to determine the appropriate likelihood model by computing the likelihood score of a tree under various models using the hierarchical likelihood ratio tests and the Akaike Information Criterion (AIC). Of all 56 substitution models evaluated, K81uf + I + G was identified as the best fitting one for the PR data while GTR + I + G was the best model for the RT data. The transition-transversion (TiTv) ratios and nucleotide substitution rate matrices (Table 1) were also evaluated. The parameters derived were then used in PAUP* to obtain the optimal maximum likelihood reconstruction[27] of phylogenetic relationships for the RT and PR DNA sequences, with HXB2 designated as the outgroup. Bootstrap values from 100 resamplings using heuristic searches with simple taxon addition[28] were calculated for the data. The sequences were gap stripped prior to tree reconstructions.

The aligned codons were visually compared across subtypes to identify both synonymous and nonsynonymous substitutions at every codon position in both p51RT and PR genes. Nucleotide sequences were translated to amino acid sequences using the MacClade 3.08a software.[29] Drug resistance mutations (DRMs) were identified as described in the Stanford HIV Drug Resistance Database (http://hivdb.stanford.edu/). The whole protease and p51RT were analyzed to identify potential DRMs, polymorphisms at DRM sites, and subtype-specific polymor-

TABLE 1. GENETIC DIVERSITY OF HIV-1 REVERSE TRANSCRIPTASE AND PROTEASE

	RT	*PR*
Subtypes[a]		
CRF02_AG	20 (57.1%)	21 (48.9%)
Subtype G	9 (25.7%)	16 (37.2%)
CRF06_cpx	4 (11.4%)	5 (11.6%)
Others	2 (5.7%)	1 (2.3%)
Phylogenetic parameters[b]		
Transition-transversion ratio	4.3763	3.9352
Best-fit model[c]	GTR+I+G	K81uf+I+G
Nucleotide substitution rate matrix		
A-C	2.1301	1.0000
A-G	11.0303	4.8929
A-T	1.0140	0.3479
C-G	1.3411	0.3479
C-T	14.1003	4.8929
G-T	1.0000	1.0000

[a]Maximum likelihood analysis by PAUP*.
[b]Parameters derived using Modeltest 3.6 program.
[c]GTR, general time reversible; K81uf, unequal-frequency Kimura three-parameter; I, invariable sites; G, gamma distribution.

phisms. Subtype-specific polymorphisms were defined, with respect to subtype B consensus, as specific mutations occurring in the majority of sequences from a particular subtype. The HIV subtype B consensus sequence was also used as the reference sequence for drug resistance analyses.

Prediction of phosphorylation sites and motifs

The phosphorylation sites in the amino acid sequences of PR and p51RT were predicted by NetPhos 2.0 and NetPhosK 1.0 (www.cbs.dtu.dk/services/). These programs produce neural network predictions for serine, threonine, or tyrosine phosphorylation sites and, in the case of NetPhosK, predict the likely kinase involved based on phosphorylation motifs.[30,31] Likely sites for phosphorylation by the serine-threonine kinases MAP kinase (MAPK), casein kinase II, and protein kinases A, G, and C (AGC) were also determined using previously described consensus recognition sequences.[32]

Identification of CTL epitopes

CTL epitopes within the PR and p51RT were identified from the Los Alamos Molecular Immunology Database.[33]

RESULTS

Genetic diversity and subtype

Maximum likelihood phylogenetic reconstruction was performed for 35 RT and 43 PR sequences from 50 HAART-naive subjects using PAUP (Fig. 1). The amplification of the samples from the remaining subjects was unsuccessful. For the RT gene, there were 20 (57%) CRF02_AG, 9 (26%) subtype G, and 4 (11%) CRF06_cpx sequences, while in the PR, there were 21 (49%) CRF02_AG, 16 (37%) subtype G, and 5 (12%) CRF06_cpx sequences (Table 1). The remaining sequences were subtype A (Table 2). Modeltest estimated the average TiTv ratios of RT and PR as 4.38 and 3.94, respectively. The rates of nucleotide substitution were C/T > G/A >> A/C > G/C > A/T > G/T for RT and G/A = C/T >> A/C = G/T >> G/C = A/T for PR (Table 1).

Amino acid diversity and subtype-specific polymorphisms

The entire PR and the first 1320 bp of RT were translated and analyzed. Two levels of amino acid diversity were examined: variation of each subtype with respect to HIV-1 subtype B and the level of variation within each subtype.

Figure 2 shows the alignments of consensus PR and p51RT amino acid sequences for CRF02_AG, subtype G, and CRF06_cpx, respectively, within our cohort, compared with subtype B consensus sequences. Within the RT, the consensus sequences for CRF02_AG, subtype G, and CRF06_cpx differed from subtype B consensus in 27, 32, and 29 positions, respectively. Within the PR, consensus sequences for CRF02_AG, subtype G, and CRF06_cpx differed from subtype B in 7, 10, and 9 positions, respectively. In addition, most of the sequences in the subtype G had the amino acids Q, K, E, and I at positions 35, 57, 67, and 82. This pattern was not seen in any other subtype.

The degree of amino acid variation within each subtype was also examined by identifying positions at which the consensus amino acid for the subtype was not conserved. Within p51RT, there were 121 (27.5%) and 76 (17.3%) nonconserved positions among the CRF02_AG and subtype G sequences, respectively. This difference was statistically significant (two-tailed; $p < 0.001$). Within PR, there were 22 (22.2%) nonconserved positions among the CRF02_AG sequences while the subtype G sequences had 14 (14.1%) nonconserved positions. This difference was not statistically significant (two-tailed; $p = 0.197$).

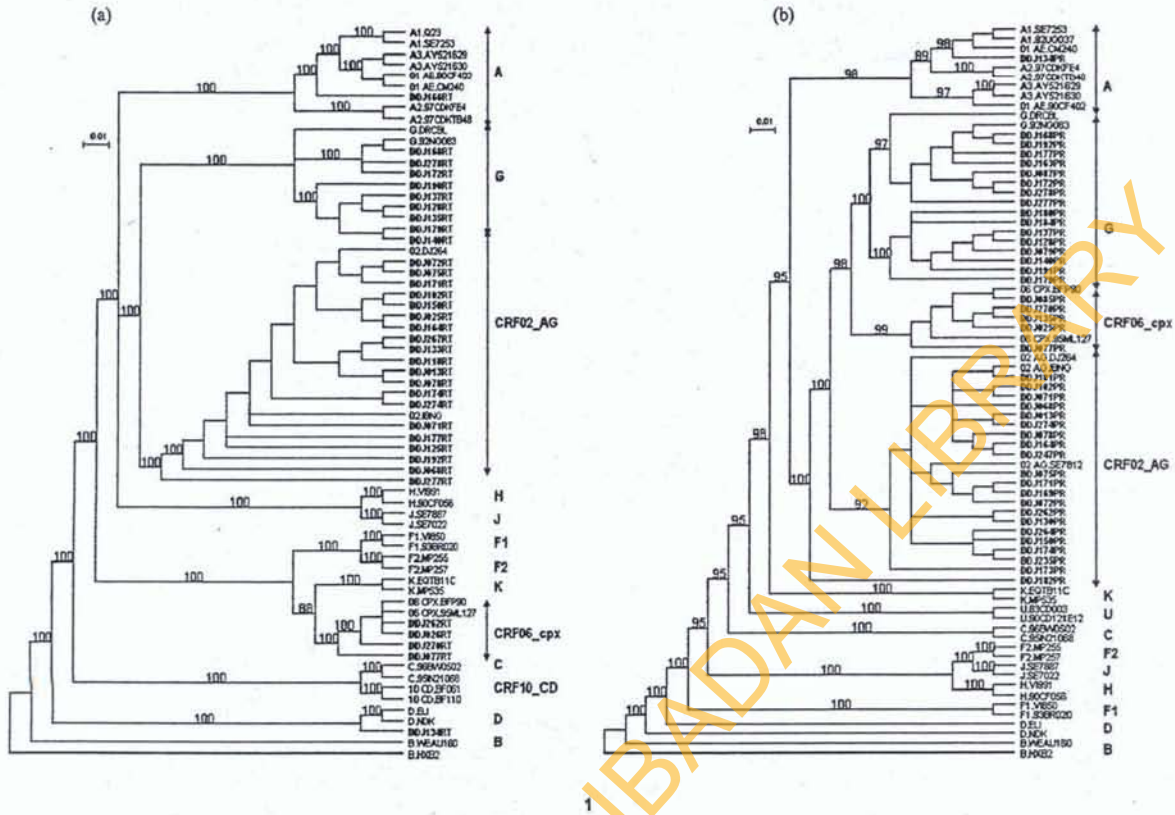

FIG. 1. Maximum likelihood trees for reverse transcriptase (a) and protease (b) sequences. The 35 RT and 43 PR sequences were each aligned with a set of 32 reference sequences from subtypes A–J and CRFs01, 02, and 06. The alignments were done by CLUSTAL X and the Modeltest 3.6 program and PAUP* (Swofford 2001) were used in tandem to determine the appropriate likelihood model by computing the likelihood score of a tree under various models using the hierarchical likelihood ratio tests and the Akaike Information Criterion (AIC). Of all substitution models evaluated, K81uf + I + G was identified as the best fitting one for the PR data while GTR+I+G was the best model for the RT data. The transition-transversion ratios were estimated as 4.38 and 3.94 for RT and PR, respectively. Bootstrap values from 100 resamplings using heuristic searches with simple taxon addition were calculated for the data; only values of >80 are indicated. HXB2 was the outgroup. The study sequences are in boldface.

Mutations at previously characterized drug resistance sites

We found primary and secondary DRMs as well as polymorphisms at positions of previously characterized DRMs (Table 2). In the RT, 22 of 35 sequences had mutations, with an average mutation number of 1.32 (range: 0–3). Interestingly, CRF06_cpx sequences had a higher average number of mutations (2.25 versus 0.6 and 0.63) than CRF02_AG and subtype G, respectively (two-tailed $p < 0.001$ with Bonferroni correction).

Primary drug resistance mutations were found at five RT sites (M41L, V118I, Y188H, P236L, Y318F), secondary mutations at two sites (V179E, R211K), and polymorphisms at three sites previously characterized for drug resistance (Table 2). The mutation M41L was found in three samples while the secondary NRTI mutation, R211K, was observed in 17 of 35 sequences. We found two polymorphisms (L210Q and T215A) at known primary drug resistance positions. These are distinct from described resistance mutations at those sites. In all, 6 (17.1%) of 35 RT samples harbored primary resistance mutations for NRTIs and NNRTIs. Notably, three of these individuals were infected with viruses that clustered with CRF06_cpx in the RT. Interestingly, three of the four CRF06_cpx RT samples harbored primary DRMs, compared with 11% of the other variants combined ($p = 0.011$)

Mutations were detected in all of the PR sequences, with an average mutation number of 2.72 (range: 2–4). No primary DRMs were observed. However, secondary DRMs were observed in all 43 PR sequences. Three secondary PI resistance mutations (L10I/V, M36I, L63P) were detected (Table 2). A polymorphism at a known primary mutation site (V82I) was found in all subtype G samples while a polymorphism at a known secondary mutation site (K20I) was also found in all samples.

TABLE 2. HIV-1 SUBTYPES AND GENOTYPES AT DRUG RESISTANCE-RELATED POSITIONS

Specific mutations in reverse transcriptase by subtype

HSPH No.	*Subtype*	*Mutation*[a]
DDJ 166	A	R211S
DDJ 013	CRF02_AG	R211K
DDJ 025	CRF02_AG	*L210Q*, R211K
DDJ 068	CRF02_AG	R211K
DDJ 071	CRF02_AG	R211K
DDJ 072	CRF02_AG	R211K
DDJ 075	CRF02_AG	—
DDJ 078	CRF02_AG	R211K
DDJ 102	CRF02_AG	—
DDJ 118	CRF02_AG	—
DDJ 125	CRF02_AG	**M41L**
DDJ 133	CRF02_AG	—
DDJ 150	CRF02_AG	**P236L**
DDJ 164	CRF02_AG	—
DDJ 171	CRF02_AG	—
DDJ 174	CRF02_AG	—
DDJ 177	CRF02_AG	R211K
DDJ 192	CRF02_AG	R211K
DDJ 267	CRF02_AG	—
DDJ 274	CRF02_AG	—
DDJ 277	CRF02_AG	R211K
DDJ 026	CRF06_cpx	**M41L, V118I,** *R211A*
DDJ 077	CRF06_cpx	R211K, **Y318F**
DDJ 262	CRF06_cpx	**M41L, V179E,** R211K
DDJ 270	CRF06_cpx	R211K
DDJ 134	D	**Y188H,** R211K
DDJ 128	G	R211K
DDJ 135	G	—
DDJ 137	G	—
DDJ 140	G	—
DDJ 168	G	R211K
DDJ 172	G	R211K
DDJ 179	G	
DDJ 190	G	*T215A*
DDJ 278	G	R211K

Frequency of mutations in protease by subtype

Position	*CRF02_AG*	*Subtype G*	*CRF06_cpx*
L10I/V	2/21	—	—
K20I	21/21	16/16	5/5
M36I	21/21	16/16	5/5
L63P	4/21	6/16	1/5
V77I	—	2/16	—
V82I	—	15/16	—
I93M	1/21	—	—

[a]Primary drug resistance mutations (DRM) are in boldface; polymorphisms at known DRM sites are in italics; secondary mutations are in regular type; absence of DRMs is represented by a dash.

Synonymy and subtype-specific codon usage

We hypothesized that subtype-specific codon usage would influence divergent mutational pathways for drug resistance and that both CRF02_AG and subtype G would have unique subtype-dependent synonymous mutations that would influence pathways of developing resistance. There was no significant preponderance of synonymous mutations at DRM sites by subtype.

We performed a codon-by-codon comparison of the nucleotide sequences of the subtype B consensus to the consensus sequences of CRF02_AG, subtype G, and CRF06_cpx samples of our cohort in both PR and RT in order to detect subtype-specific codon usage. We observed many examples of differential codon usage by subtype, mostly at positions that do not have any previously described DRMs. Figure 3 shows examples of how codon bias, by subtype, at DRM sites within PR and RT, may affect the rate and pathway of drug resistance development. At position 10 in PR, leucine is encoded for differentially in subtype G and CRF02_AG by the codons CTA and TTA, respectively. Based on this differential codon usage for leucine as well as the finding that C-A and T-G substitutions are more common than C-G and T-A substitutions (Table 1), the variants will preferentially develop L10I and L10V mutations, respectively, under drug pressure. Similarly, at position V179 in RT, subtype G and CRF02_AG viruses utilize the codon GTG to encode valine and more likely to develop V179M and V179E mutations, whereas subtype B viruses are biased toward V179I and V179D mutations. Similar effects of codon bias by subtype at DRM sites are present at other positions, especially L74 and L210 in RT and G73 in PR.

Effect of mutations on phosphorylation sites

Based on a recent report that some potential phosphorylation sites in RT were present in subtypes A, B, and D but absent in subtype C sequences,[19] we postulated that there would be differences in the number and sites for phosphorylation by subtype. We therefore analyzed the PR and p51RT for possible serine-threonine phosphorylation sites based on previously described phosphorylation motifs for MAP kinase, casein kinase II, and protein kinases A, G, and C. We also predicted the likelihood of phosphorylation with the NetPhos/NetPhosK programs. Two serine-threonine phosphorylation sites were found in the PR while 11 were found in the RT. All the sites predicted were within recognized phosphorylation motifs. Interestingly, the T386I/V polymorphism seen in all subtype G sequences led to the loss of a predicted phosphorylation site while a Q394R polymorphism in all these sequences created a protein kinase A/G/C site (as predicted by NetPhosK) at threonine 397 (Fig. 2).

Subtype-specific polymorphisms and MHC epitopes

The consensus sequences for CRF02_AG, subtype G, and CRF06_cpx PR and RT in the study were aligned to cytotoxic T lymphocyte (CTL), T-helper, or antibody epitope maps on the Los Alamos database. Then, a literature search was conducted to identify subtype-specific polymorphisms that occurred at positions previously implicated to modulate the immune response to PR and RT epitopes. No match was found

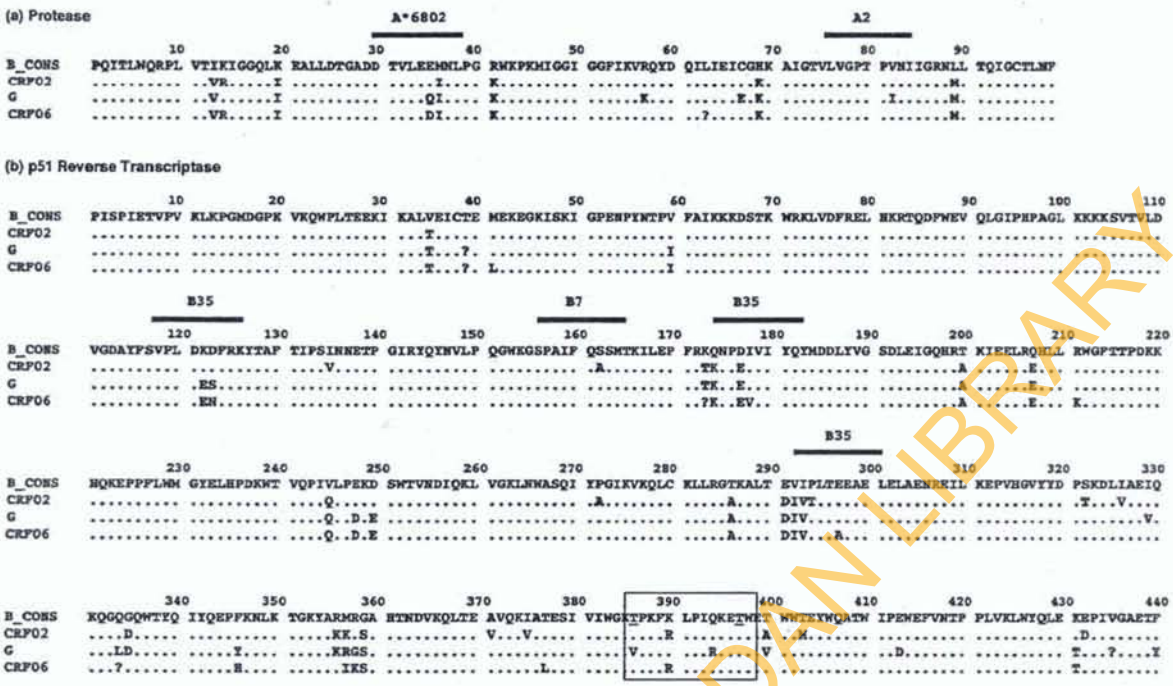

FIG. 2. Subtype-specific polymorphisms in HIV-1 protease and p51 RT sequences from Ibadan and Saki. Alignment of deduced amino acid sequences for protease (a) and reverse transcriptase (b) compared with the subtype B consensus sequence. The sequences were aligned with the CLUSTAL X software package and translated using the MacClade 3.08a software. Amino acids are represented by the single-letter amino acid code. Each amino acid residue not differing from the reference sequence is represented by a dot. The consensus amino acid was determined by simple majority from the sequences in the cohort, and a question mark (?) was used to denote polymorphic positions at which there was no clear consensus. HLA epitopes within which subtype-specific polymorphisms occur at sites previously described to be important for immune evasion are denoted by thick bands. The box denotes the region of balanced polymorphism of potential phosphorylation sites.

for T-helper and antibody epitopes. However, within the RT, we found three human leukocyte antigen (HLA) B35 epitopes (VPLDKDFRKY, NPDIVIYQY, and IPLTEEAEL), and one for HLA B7 (SPAIFQSSM), within which subtype-specific polymorphisms occur at positions that had been previously shown experimentally to have immune escape variants (Fig. 2). Within the protease, an A*6802 epitope DTVLEEMNL harbored polymorphisms in all our samples, while the A2 epitope VLVGPTPVNI contains a V82I polymorphism in all subtype G sequences (Fig. 2).

DISCUSSION

Although the pol gene is the most conserved region of HIV-1, with a variation of ~10%,[7,34] it possesses sufficient variability to allow for phylogenetic reconstruction.[35] Thus, we performed phylogenetic analyses of our RT and PR sequences for subtyping using optimal maximum likelihood models. We obtained transition-transversion (TiTv) ratios for RT and PR of 4.38 and 3.94. These values are higher than the average TiTv ratio of 1.42 described for the envelope V3 region,[36] and may be useful as default values for future analyses of RT and PR sequences. The complex diversity of HIV-1 viruses in Nigeria is becoming more evident. Although CRF02_AG is still a major subtype in the country, as previously reported,[10–12] subtype G and CRF06_cpx viruses appear to be increasing in prevalence.

We found a significant degree of primary and secondary DRMs in the RT and PR, as well as polymorphisms at positions of previously characterized drug-resistant mutations. Six (17%) of 35 RT sequences harbored primary NRTI/NNRTI mutations, including M41L, V118I, Y188H, P236L, and Y318F. This is rather high for a drug-naive cohort but it is noteworthy that half the sequences with drug-naive DRMs are CRF06_cpx. Our findings suggest the possibility that CRF06_cpx sequences harbor more drug-resistant-related mutations in drug-naive individuals. However, it is possible that individuals with these DRMs were infected with drug-resistant viruses, possibly from individuals not adherent to therapy, or on suboptimal ARV regimens. Indeed, the DRMs observed in our CRF06_cpx samples were not observed in the reference sequences on the Los Alamos database. In addition, the drug-naive status was self-reported and it is possible that the reliability of this self-report could be linked to other factors. The relationship between CRF06_cpx and drug resistance is not well characterized and further investigation is required.

The most common mutations in RT were L214F (83%) and R211K (49%). Previous studies from the West African region showed that some patients with RT mutations, R211K, L214F,

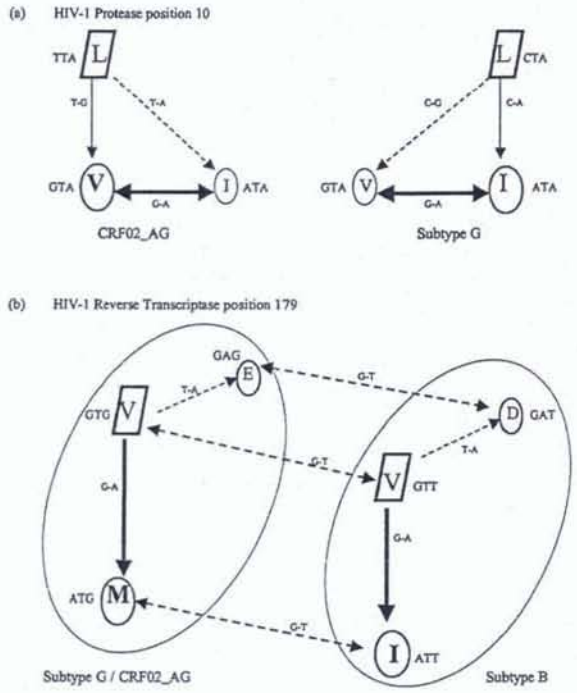

FIG. 3. Prediction of subtype-specific drug resistance mutations from codon bias and nucleotide substitution matrix data. Resistance mutations may be influenced by quasi-synonymy and genetic cost of mutations. Examples of predicted mutations at positions L10 in PR (**a**) and V179 in RT (**b**) are shown. The wild-type amino acids are enclosed within quadrilaterals, while mutant amino acids are enclosed with ovals. The predicted relative abundance of mutants is proportional to the size of the oval. The preferred codon bias is indicated adjacent to each amino acid. Codon usage was determined from consensus sequences from our samples and subtype B consensus. The thickness of the arrows indicates the ease of substitution. The nucleotide substitution matrix data for PR (G/A = C/T >> A/C = G/T >> G/C = A/T) is utilized to form three broad groups. The most likely substitutions (G/A and C/T) are represented by bold arrows, and the least likely substitutions (A/T and G/C) by arrows with broken lines. The intermediate level substitutions (A/C and G/T) are depicted with regular thickness arrows. A similar scheme is used for RT based on the matrix C/T > G/A >> A/C > G/C > A/T > G/T. CRF02_AG and subtype G viruses are predicted to preferentially develop L10V and L10I mutations, respectively, in PR (**a**). Both variants are predicted to preferentially develop V179M and V179E mutations in RT (b), while subtype B viruses will develop V179I and V179D mutations under drug pressure.

I135V, and L210M, also demonstrated phenotypic resistance.[17] The R211K mutation was also found in a majority of drug-naive Zambians with subtype C infection.[20] The K20I and M36I secondary PI mutations were found in all subjects, regardless of subtype, while all the subtype G viruses harbored the V82I polymorphism in the PR. These data raise concerns for second line and salvage therapy regimens, which usually contain PIs, although these regimens are infrequently used in Africa currently.

Subtype-specific polymorphisms may play functional roles in determining the path to drug resistance. Development of drug resistance is dependent upon the extent of viral replication during therapy, the ease of acquiring a particular mutation or set of mutations, and the effect of mutations on drug susceptibility and viral fitness.[24] Due to the degeneracy of the genetic code, it is possible for single nucleotide substitutions in any particular codon to result in synonymous or nonsynonymous mutations. However, it is possible that different mutational pathways will be employed by different codons even if they encode the same amino acids. This relationship between codons that specify the same amino acid, but can achieve a different pattern of amino acid substitution by a single nucleotide change, has recently been termed quasi-synonymy.[37] Quasi-synonymous mutations within the virus may dictate different mutational pathways, which may have bearing on the development of drug resistance and viral escape from the immune response. This phenomenon has been described in other cohorts,[38,39] and is likely responsible for the finding that the V106M mutation is more commonly associated with multi-NNRTI resistance with subtype C but not with subtype B infection.[40–42]

By combining codon bias data and the influence of the differential genetic cost of mutations, we were able to predict mutations that are likely to predominate by subtype, under drug pressure. Figure 3 shows two examples of subtype-specific codon differences in this study, and our predictions of how they may influence mutational routes to development of drug resistance. In PR, CRF02_AG and subtype G viruses are differentially more likely to develop L10V and L10I drug resistance mutations, respectively. In RT, both subtype G and CRF02_AG are predisposed to developing V179M and V179E mutations in contradistinction to V179I and V179D mutations in subtype B. Interestingly, recent data from Kantor *et al.*[16] reporting results from a global collaboration study on non-subtype B resistance mutations, support our predictions. The statistically significant treatment-associated mutations at position 10 in PR for subtype G and CRF02_AG were L10V and L10I, respectively. Likewise, the statistically significant treatment-associated mutations at position 179 in RT for subtype G and CRF02_AG were V179M and V179E, while V179I and V179D are commonly seen in subtype B infection. Kantor *et al.* reported considerable overlap between subtype B resistance mutations and mutations associated with at least one non-B variant,[16] and some clinical data show that polymorphisms do not affect response to HAART in non-subtype B infection.[41,42] However, based on our data, the possibility that subtype-specific quasi-synonymy and codon usage may have biochemical, phenotypic, genotypic, or clinical impact[7,37] on the routes to, and severity of resistance cannot be ruled out.

The activities of RT and PR are enhanced by phosphorylation[43–46] but the phosphorylation sites on RT have not been described. There was variation in the positions of predicted serine-threonine phosphorylation sites in the RT and PR by subtype, although there were equal numbers of sites in the three subtypes. In the RT, subtype G sequences lacked a predicted phosphorylation site that was present in subtype B and CRF02_AG due to a T386V/I polymorphism. However, a phosphorylation site that was absent in the other subtypes was predicted in subtype G because of a Q394R polymorphism. This suggests the possibility that under selective pressure, the number of phosphorylation sites may change, or that there may be

some degree of balanced polymorphism at play to ensure optimal enzymatic activity. It remains to be seen if these mutations will have any effect on therapy outcomes.

Polymorphisms embedded within CTL and T-helper epitopes have been described for other subtypes.[33] We found subtype-specific polymorphisms in RT within epitopes for HLA B7 and B35 (Fig. 2), which, interestingly, have relatively high allelic frequencies (11% and 7%, respectively) in the Southern part of Nigeria[47,48] Our previous studies have previously demonstrated that the CTL response is abrogated with the S162A mutation in subtype A individuals.[49] This polymorphism is found in a majority of CRF02_AG viruses within the B7 epitope. Kawana *et al.* showed that 80% of HLA $B35^+$ individuals but only 10% of HLA $B35^-$ individuals had NPDIVIYQY peptides with the D177E substitution in the RT.[50] These peptides had reduced binding affinity to B35 and the D177E substitution may be an escape mutant; 95% of all our samples had this substitution. In addition, the V179I substitution in the NPDIVIYQY epitope has been reported to reduce B35 recognition and binding affinity.[51] Therefore RT position 179 is likely to be under dual pressure from both antiretroviral therapy and cell-mediated immune response. Within PR, position V82 in the A2 epitope VLVGPTPVNI was demonstrated to be under dual immune and drug pressure;[52] all our subtype G sequences have a V82I polymorphism. The A*6802 epitope DTVLEDINL in PR is considered to be a "resistant epitope" because it was preferentially reactive in highly exposed but persistently seronegative women.[53] However, this exact epitope is not found in CRF02 and subtype G samples.

HIV-1 subtypes are predominant in different geographical regions of the world[54] and many epitopes have ethnogeographic bias.[55] It is conceivable that subtype-specific polymorphisms exist within CTL and T-helper epitopes for HLA proteins that are common in the regions where these subtypes predominate. HLA B35 and B7 occur at similarly high levels in whites,[47,56] and have been associated with increased rates of disease progression and higher viral loads, respectively.[54,57] We suggest that the relatively high frequencies of these antigens within various populations contribute to the selective pressures that lead to the development of immune escape variants, which then persist in these populations. For example, Fig. 2 shows that more diversity across subtypes occurs within B35 epitopes. It is also possible that differences in HLA frequencies between populations will contribute to this phenomenon. Our data support the hypothesis that some subtype-specific polymorphisms may be immune escape variants that have persisted in the populations where these epitopes and subtypes predominate.

There were several challenges during the study, including the failure of proviral DNA amplification of some samples, therefore a potential source of bias. This may have resulted from atypical nucleotides at primer sites or low viral loads. These samples were obtained as part of a larger survey of HIV prevalence in the community, and viral load and CD4 determinations were not part of the original study protocol. DNA proviral cloning rather than RT-PCR was utilized for the analysis because the samples available were more suitable for the former. RT-PCR has the advantage of detecting viral quasispecies that have most recently been selected by therapy, whereas proviral DNA cloning is beneficial for detecting archived mutations.[4,58] It would be interesting to compare the utility of both methods in tandem in future studies. Despite these various limitations, this subset of samples shows a similar spectrum of subtypes in the larger epidemiological study (Sankale et al., manuscript in preparation) and the resistance data obtained serve as critical baseline information for planning antiretroviral regimens.

In summary, we analyzed the PR and RT nucleotide and amino acid sequences in antiretroviral-naive Nigerians. All PR sequences had secondary mutations and 17% of the RT sequences had primary NRTI/NNRTI mutations. We also demonstrated that HIV-1 CRF02_AG and subtype G often utilize different synonymous codons, either in contrast with one another or with subtype B, in encoding amino acids, and have distinct functional motifs that may influence divergent pathways to drug resistance. Our data also suggest the possibility that subtype-specific polymorphisms arose in part as immune escape mutants. These findings emphasize the significance of subtype-specific determinants of susceptibility to antiretroviral therapy and immune response.

SEQUENCE DATA

All nucleotide sequences were submitted to GenBank. The accession numbers for the RT sequences are DQ273889–DQ273923 while the accession numbers for the protease sequences are DQ273924–DQ273966.

ACKNOWLEDGMENTS

This study was supported by the AIDS Prevention Initiative in Nigeria (APIN), a project based at the Harvard School of Public Health, and funded by the Bill and Melinda Gates Foundation. A.I.O is an APIN Fellow. The authors thank the University of Ibadan/University College Hospital, Ibadan, and all the subjects who participated in the study.

REFERENCES

1. Mansky LM: Retrovirus mutation rates and their role in genetic variation. J Gen Virol 1998;79:1337–1345.
2. Preston BD: Error-prone retrotransposition: Rime of the ancient mutators. Proc Natl Acad Sci USA 1996;93:7427–7431.
3. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, and Markowitz M: Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 1995;373:123–126.
4. Wei X, Ghosh SK, Taylor ME, *et al.*: Viral dynamics in human immunodeficiency virus type 1 infection. Nature 1995;373:117–122.
5. Robertson DL, Sharp PM, McCutchan FE, and Hahn BH: Recombination in HIV-1. Nature 1995;374:124–126.
6. Wolinsky SM, Korber BT, Neumann AU, *et al.*: Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. Science 1996;272:537–542.
7. Kantor R and Katzenstein D: Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. AIDS Rev 2003;5:25–35.
8. Howard TM, Olaleye DO, and Rasheed S: Sequence analysis of the glycoprotein 120 coding region of a new HIV type 1 subtype

A strain (HIV-1IbNg) from Nigeria. AIDS Res Hum Retroviruses 1994;10:1755–1757.
9. Agwale SM, Zeh C, Robbins KE, *et al.*: Molecular surveillance of HIV-1 field strains in Nigeria in preparation for vaccine trials. Vaccine 2002;20:2131–2139.
10. 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;254:226–234.
11. Montavon C, Toure-Kane C, Liegeois F, *et al.*: Most env and gag subtype A HIV-1 viruses circulating in West and West Central Africa are similar to the prototype AG recombinant virus IBNG. J Acquir Immune Defic Syndr 2000;23:363–374.
12. 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: 315–325.
13. Yeni PG, Hammer SM, Hirsch MS, *et al.*: Treatment for adult HIV infection: 2004 recommendations of the International AIDS Society-USA Panel. JAMA 2004;292:251–265.
14. Idigbe EO, Adewole TA, Eisen G, *et al.*: Management of HIV-1 infection with a combination of nevirapine, stavudine, and lamivudine: A preliminary report on the Nigerian antiretroviral program. J Acquir Immune Defic Syndr 2005;40:65–69.
15. Laurent C, Diakhate N, Gueye NF, *et al.*: The Senegalese government's highly active antiretroviral therapy initiative: An 18-month follow-up study. AIDS 2002;16:1363–1370.
16. Kantor R, Katzenstein DA, Efron B, *et al.*: Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: Results of a global collaboration. PLoS Med 2005;2:e112.
17. Adje-Toure C, Bile CE, Borget MY, *et al.*: Polymorphism in protease and reverse transcriptase and phenotypic drug resistance of HIV-1 recombinant CRF02_AG isolates from patients with no prior use of antiretroviral drugs in Abidjan, Cote d'Ivoire. J Acquir Immune Defic Syndr 2003;34:111–113.
18. Becker-Pergola G, Kataaha P, Johnston-Dow L, Fung S, Jackson JB, and Eshleman SH: Analysis of HIV type 1 protease and reverse transcriptase in antiretroviral drug-naive Ugandan adults. AIDS Res Hum Retroviruses 2000;16:807–813.
19. Gordon M, De Oliveira T, Bishop K, *et al.*: Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: Implications for vaccine and antiretroviral control strategies. J Virol 2003;77:2587–2599.
20. Handema R, Terunuma H, Kasolo F, *et al.*: Prevalence of drug-resistance-associated mutations in antiretroviral drug-naive Zambians infected with subtype C HIV-1. AIDS Res Hum Retroviruses 2003;19:151–160.
21. Konings FA, Zhong P, Agwara M, *et al.*: Protease mutations in HIV-1 non-B strains infecting drug-naive villagers in Cameroon. AIDS Res Hum Retroviruses 2004;20:105–109.
22. Pillay C, Bredell H, McIntyre J, Gray G, and Morris L: HIV-1 subtype C reverse transcriptase sequences from drug-naive pregnant women in South Africa. AIDS Res Hum Retroviruses 2002;18: 605–610.
23. Vicente AC, Agwale SM, Otsuki K, *et al.*: Genetic variability of HIV-1 protease from Nigeria and correlation with protease inhibitors drug resistance. Virus Genes 2001;22:181–186.
24. Shafer RW: Genotypic testing for HIV-1 drug resistance (November 30, 2003). http://hivdb.stanford.edu/modules/lookUpFiles/pdf/GenotypicResistance.pdf. Accessed February 17, 2004.
25. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, and Higgins DG: The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 1997;25:4876–4882.
26. Posada D and Crandall KA: MODELTEST: Testing the model of DNA substitution. Bioinformatics 1998;14:817–818.
27. Swofford DL: PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, MA, 2001.
28. McCormack GP and Clewley JP: The application of molecular phylogenetics to the analysis of viral genome diversity and evolution. Rev Med Virol 2002;12:221–238.
29. Maddison WP and Maddison DR: Interactive analysis of phylogeny and character evolution using the computer program MacClade. Folia Primatol (Basel) 1989;53:190–202.
30. Blom N, Gammeltoft S, and Brunak S: Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 1999;294:1351–1362.
31. Blom N, Sicheritz-Ponten T, Gupta R, Gammeltoft S, and Brunak S: Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics 2004;4:1633–1649.
32. Kennelly PJ and Krebs EG: Consensus sequences as substrate specificity determinants for protein kinases and protein phosphatases. J Biol Chem 1991;266:15555–15558.
33. Brander C and Goulder PJR: The identification of optimal HIV-derived CTL epitopes in diverse populations using HIV clade-specific consensus. In: *HIV Molecular Immunology 2001* (Korber BT, Brander C, Haynes BF, *et al.*, eds.). Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM, 2001, pp. I-1–20.
34. McClure MA, Johnson MS, Feng DF, and Doolittle RF: Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. Proc Natl Acad Sci USA 1988;85:2469–2473.
35. Hue S, Clewley JP, Cane PA, and Pillay D: HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS 2004;18:719–728.
36. Leitner T, Kumar S, and Albert J: Tempo and mode of nucleotide substitutions in gag and env gene fragments in human immunodeficiency virus type 1 populations with a known transmission history. J Virol 1997;71:4761–4770.
37. Kijak GH, Currier JR, Tovanabutra S, *et al.*: Lost in translation: Implications of HIV-1 codon usage for immune escape and drug resistance. AIDS Rev 2004;6:54–60.
38. Dumans AT, Soares MA, Machado ES, *et al.*: Synonymous genetic polymorphisms within Brazilian human immunodeficiency virus type 1 subtypes may influence mutational routes to drug resistance. J Infect Dis 2004;189:1232–1238.
39. Loemba H, Brenner B, Parniak MA, *et al.*: Genetic divergence of human immunodeficiency virus type 1 Ethiopian clade C reverse transcriptase (RT) and rapid development of resistance against non-nucleoside inhibitors of RT. Antimicrob Agents Chemother 2002;46:2087–2094.
40. Johnson VA, Brun-Vezinet F, Clotet B, *et al.*: Drug resistance mutations in HIV-1. Top HIV Med 2003;11:215–221.
41. Frater AJ, Beardall A, Ariyoshi K, *et al.*: Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. AIDS 2001;15:1493–1502.
42. Turner D, Brenner B, Moisi D, *et al.*: Nucleotide and amino acid polymorphisms at drug resistance sites in non-B-subtype variants of human immunodeficiency virus type 1. Antimicrob Agents Chemother 2004;48:2993–2998.
43. Haneda E, Furuya T, Asai S, Morikawa Y, and Ohtsuki K: Biochemical characterization of casein kinase II as a protein kinase responsible for stimulation of HIV-1 protease in vitro. Biochem Biophys Res Commun 2000;275:434–439.
44. Harada S, Haneda E, Maekawa T, *et al.*: Casein kinase II (CK-II)-mediated stimulation of HIV-1 reverse transcriptase activity and characterization of selective inhibitors in vitro. Biol Pharm Bull 1999;22:1122–1126.

45. Idriss H, Kawa S, Damuni Z, Thompson EB, and Wilson SH: HIV-1 reverse transcriptase is phosphorylated in vitro and in a cellular system. Int J Biochem Cell Biol 1999;31:1443–1452.
46. Lazaro JB, Boretto J, Selmi B, Capony JP, and Canard B: Phosphorylation of AZT-resistant human immunodeficiency virus type 1 reverse transcriptase by casein kinase II in vitro: Effects on inhibitor sensitivity. Biochem Biophys Res Commun 2000;275: 26–32.
47. Okoye RC, Williams E, Alonso A, *et al.*: HLA polymorphisms in Nigerians. Tissue Antigens 1985;25:142–155.
48. Uko GP, Asuquo MA, Dawkins RL, Udim ER, Banjoko SO, and Jaja O: HLA class I phenotype and gene frequencies in parts of southern Nigeria. Dis Markers 1997;13:123–129.
49. Cao H, Kanki P, Sankale JL, *et al.*: Cytotoxic T-lymphocyte cross-reactivity among different human immunodeficiency virus type 1 clades: Implications for vaccine development. J Virol 1997;71: 8615–8623.
50. Kawana A, Tomiyama H, Takiguchi M, Shioda T, Nakamura T, and Iwamoto A: Accumulation of specific amino acid substitutions in HLA-B35-restricted human immunodeficiency virus type 1 cytotoxic T lymphocyte epitopes. AIDS Res Hum Retroviruses 1999;15:1099–1107.
51. Tomiyama H, Miwa K, Shiga H, *et al.*: Evidence of presentation of multiple HIV-1 cytotoxic T lymphocyte epitopes by HLA-B*3501 molecules that are associated with the accelerated progression of AIDS. J Immunol 1997;158:5026–5034.
52. Karlsson AC, Deeks SG, Barbour JD, *et al.*: Dual pressure from antiretroviral therapy and cell-mediated immune response on the human immunodeficiency virus type 1 protease gene. J Virol 2003;77:6743–6752.
53. Kaul R, Dong T, Plummer FA, *et al.*: CD8(+) lymphocytes respond to different HIV epitopes in seronegative and infected subjects. J Clin Invest 2001;107:1303–1310.
54. Wainberg MA: HIV-1 subtype distribution and the problem of drug resistance. AIDS 2004;18:S63–68.
55. Winkler CA and O'Brien SJ: Effect of genetic variation on HIV transmission and progression to AIDS. In: *AIDS in Africa, 2nd ed.* (Essex M, Mboup S, Kanki PJ, Marlink RG, and Tlou SD, eds.). Kluwer Academic, New York, 2002, pp. 52–73.
56. Imanishi T, Akaza T, Kimura A, Tokunaga K, and Gojobori T: Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: *HLA 1991*, Vol. 1 (Tsuji K, Aizawa M, and Sasazuki T, eds). Oxford University Press, Oxford, 1992, pp. 1065–1220.
57. Moore CB, John M, James IR, Christiansen FT, Witt CS, and Mallal SA: Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. Science 2002;296:1439–1443.
58. Koch N, Yahi N, Ariasi F, Fantini J, and Tamalet C: Comparison of human immunodeficiency virus type 1 (HIV-1) protease mutations in HIV-1 genomes detected in plasma and in peripheral blood mononuclear cells from patients receiving combination drug therapy. J Clin Microbiol 1999;37:1595–1597.

Address reprint requests to:
Phyllis J. Kanki
Department of Immunology and Infectious Diseases
Harvard School of Public Health
651 Huntington Avenue, FXB 404
Boston, Massachusetts 02115

E-mail: pkanki@hsph.harvard.edu