ABSTRACT

Several HIV-1 subtype C-specific gag- and/or nef-based vaccines are currently intended for clinical trial in southern Africa. Here we provide sequences of 64 gag and 45 nef genes sampled in Malawi, Zambia, Zimbabwe, and South Africa and assess the degree of southern African HIV-1 diversity that will confront these vaccines. Whereas reasonable phylogenetic evidence exists for geographical clustering of subtype C gag and nef sequences from various other parts of the world, there is little evidence of similar population founder effects in the southern African epidemic. The entire breadth of subtype C diversity is represented in the southern African genes suggesting there may be no advantage in producing region- or country-specific subtype C vaccines. We do not, however, find much evidence of intersubtype recombination in the Southern African genes, implying that the likelihood of vaccine failure due to the emergence of intersubtype recombinants is probably low.

SOUTHERN AFRICAN COUNTRIES are experiencing the greatest burden of the AIDS epidemic as a result of the devastating spread of human immunodeficiency virus type 1 (HIV-1) subtype C. Estimated southern African adult HIV prevalence rates of over 30% in some regions highlight the need for an effective vaccine (http://www.who.int/emc-hiv/fact_sheets/).

One of the most important challenges in vaccine design is dealing with the high degrees of HIV-1 genetic diversity that are a characteristic of the AIDS epidemic. Fueled by high viral mutation, recombination, and replication rates and driven by strong host immune pressures, the diversity of global HIV populations is continuing to increase at an astonishing rate. There is, however, considerable variability in the rate at which different regions of the HIV genome are diversifying. Differences in the degree of nucleotide sequence conservation across the HIV genome reflect in part the balance that exists between selection pressures exerted by host immune systems and the functional constraints of viral proteins.

Gag and the central region of Nef are reasonably conserved HIV proteins with high epitope densities that are among the most common targets of host cellular immune responses. Conflicting clinical data exist on the association between T cell responses and plasma viremia. While some studies showed no correlation between plasma viral load and CD8+ T cell responses specific against HIV-1 p24-Gag and Nef, other studies showed an inverse correlation between HIV-1 C Gag T cell responses and plasma viral load indicating that Gag-specific cytotoxic T lymphocyte (CTL) responses are possibly important in controlling viremia. Gag and nef are therefore attractive potential vaccine components and, accordingly, there are currently several gag- and/or nef-based vaccines intended for clinical trial in southern Africa (http://www.iavireport.org/trialsdb IAVI database of AIDS vaccines in human trials).

While there is presently a large amount of data available for South African and Botswanan subtype C gag and nef sequences, the same is not true for other southern African countries. There-
for, to better assess the degree of HIV genetic diversity confronting proposed southern African subtype C-specific gag-and/or nef-based vaccines, 64 gag and 45 nef sequences were determined from 71 HIV-1-infected individuals recruited as part of the HIVNET028 study\textsuperscript{9,11} from five clinical sites in southern Africa: Johannesburg (n = 10; gag = 8; nef = 6) and Durban (n = 22; gag = 22; nef = 14) in South Africa; Lusaka in Zambia (n = 20; gag = 18; nef = 13); Harare in Zimbabwe (n = 9; gag = 6; nef = 5); and Blantyre in Malawi (n = 10; gag = 10; nef = 7).

Viral RNA was extracted from plasma using the QIAamp Viral RNA Kit (QIAGEN, Chatsworth, CA) or the NASBA RNA extractor kit (Life Sciences, Inc.). The gag and nef genes were reverse transcribed using the ThermoScript RT-PCR System (Invitrogen Corp., San Diego, CA), and an oligo(dT)\textsubscript{20} primer, and the primer Gag D reverse (\textsuperscript{5}' ATTCTCTTTATCT-ATTTTTGG3', at position 2382–2402 of the HXB2 genome) and the primer H11032 for the Gag D reverse (5'-ATTCTCTTTATCT-ATTTTTGG3', at position 9608–9625 in HXB2), respectively. A 1673-bp full-length gag gene product was amplified from cDNA by nested polymerase chain reaction (PCR) using the outer primers Gag D forward (\textsuperscript{5}' TTCTTCAGACTGGGCAGCGG3', at position 626–644) and Gag D reverse. Three fragments of \textasciitilde{}600 bp, overlapping by 77 and 99 nucleotides, respectively, were generated using inner primers: Gag A forward (\textsuperscript{5}' CCATTCACTGCCTTGCAG3', 683–704) and Gag A reverse (\textsuperscript{5}' AGGCAAGCTTTATTGAGG3', at position 9626–9646 in HXB2), respectively. A 1673-bp full-length gag gene product was amplified from cDNA by nested polymerase chain reaction (PCR) using the outer primers Gag D forward (\textsuperscript{5}' TTCTTCAGACTGGGCAGCGG3', at position 626–644) and Gag D reverse. Three fragments of \textasciitilde{}600 bp, overlapping by 77 and 99 nucleotides, respectively, were generated using inner primers: Gag A forward (\textsuperscript{5}' CCATTCACTGCCTTGCAG3', 683–704) and Gag A reverse (\textsuperscript{5}' AGGCAAGCTTTATTGAGG3', at position 9626–9646 in HXB2), respectively.

The PCR products were purified (QiAquick PCR Purification kit, QIAGEN, Chatsworth, CA) and sequenced with the forward and reverse Gag A, B, and C or Nef IF and IR primers using the BigDye Terminator Cycle sequencing kit version 3.1 (Applied Biosystems Inc., Foster City, CA) and an automated ABI PRISM 3100 genetic analyzer (Applied Biosystems Inc.).

Full length gag and nef sequences were respectively aligned using POA (gap open penalty = 12, gap extension penalty = 4)\textsuperscript{12} with all nearly full length gag (n = 1236) and nef (n = 1923) sequences available in the Los Alamos National Laboratory HIV-1 sequence database (http://hiv-web.lanl.gov/content/hiv-db) on 4 August 2005 (for gag) and 10 February 2006 (for nef). Except for subtype and circulating recombinant form (CRF) references and one HIV-1 N sequence, all sequences that did not cluster among the subtype C sequences in a neighbor-joining tree (transition/transversion ratio = 2.0, constructed with PHYLIP v3.5c)\textsuperscript{13} were discarded to obtain gag and nef datasets containing 558 and 677 sequences, respectively. These datasets were realigned using POA and potential inter- and intrasubtype recombinants among the HIVNET028 sequences were identified using six recombination detection methods implemented in RDP3 (default settings for all methods and a Bonferroni corrected p-value cutoff = 0.05).\textsuperscript{14}

None of the HIVNET028 gag or nef sequences were identified as being intersubtype recombinants. However, eight HIVNET028 gag and three HIVNET028 nef sequences were identified by two or more recombination detection methods as being intrasubtype C recombinants. Recombination can seriously impact the validity of phylogenetic inferences\textsuperscript{15,16} and we therefore removed the pieces of these HIVNET028 sequences that appeared to have a recombinant origin. Sequences between the approximate recombination breakpoints identified for the 11 detected events were replaced in the alignment with gap characters. In no case was more than 39% of a sequence replaced by gap characters. To further minimize the impact of recombination on subsequent phylogenetic reconstructions, 163 potentially recombinant gag sequences and 245 potentially recombinant nef sequences, identified by any one of the six recombination detection methods with the same Bonferroni p-value correction used to identify HIVNET028 recombinants, were completely removed from the gag and nef alignments. An additional 14 gag and 8 nef sequences identified elsewhere as being intersubtype recombinants but that were not identified as such during the recombination screen were also removed from the gag and nef datasets.

None of the HIVNET028 sequences was more closely related to previously identified intersubtype recombinants with subtype C-like gag or nef sequences than to other apparently “pure” subtype C sequences. Also, during the initial recombination screen we noted that only three Zambian, one Botswanan, and two South African sequences and one Botswanan and three South African nef sequences were probable intersubtype recombinants. All of these sequences are previously identified recombinants, but none is a recognized CRF. Our analyses indicated, however, that A2C.ZM.x.ZAM716 and A2C.ZM.89.ZAM174 share nearly identical breakpoints in gag and may be examples of a recombinant form currently circulating in Zambia. That only 7.8% of the 64 potentially recombiant southern African gag sequences and 3.1% of the 127 potentially recombinant Southern African nef sequences are intersubtype recombinants indicates that there is evidence of only limited gene flow from other subtypes or CRFs into the region’s overwhelmingly subtype C HIV population. This implies, first, that most all non-C subtypes and all recognized CRFs are either very rare or absent from southern Africa, and second, that there may be a degree of epidemiological isolation between subtype C and the other HIV-1 subtypes that have been identified in the region (such as A, B, and D).

The 373 gag and 421 nef sequences retained following elimination of potential recombinants were realigned with POA and used to construct neighbor-joining trees (Fig. 1; with transition/transversion ratios estimated from the alignments, pairwise deletion of indels, and 1000 bootstrap and branch length test replicates) with MEGA 3.0.\textsuperscript{17}

Both the gag and nef phylogenetic analyses provide clear evidence of geographical clustering of subtype C isolates in various parts of the world. South American isolates (mostly from Brazil) form single well-supported monophyletic groups (>70% bootstrap and > 95% branch length test support) in both the gag and nef trees. Also in both trees, the Indian isolates mostly fall into one of two groups. Whereas both groups are well supported (>70% bootstrap and 99% branch length test support) in the nef tree, only one is well supported in the
FIG. 1. Neighbor-joining trees indicating the phylogenetic relationships between 382 gag (A) and 430 nef (B) HIV-1 subtype C sequences. The trees were constructed using Kimura, 1980 distance matrices with transition:transversion ratios estimated from the data and pairwise deletion of indels (Mega 3.0 Kumar et al., 2004). Colored symbols at the ends of branches represent sequences sampled from different regions of the world. Open symbols with a red border represent the 64 gag and 44 nef sequences sampled from Southern Africa during the present study. Both trees were rooted using the HIV-1 subtype B isolate, HXB2, as an out-group and the significance of their interior branches tested using 1000 bootstrap and interior branch length test replicates. Solid and open black circles indicate branches with greater than 70% and 50% bootstrap support, respectively. All branches with less than 50% bootstrap support and less than 70% interior branch length test support have been collapsed. The arrow indicates a divergent subtype C-like gag sequence from Zambia. While the trees were constructed using an additional HIV-1 group N sequence and nine HIV-1 M subtype and subsubtype reference sequences, only the HIV-1 M subtype B reference sequence is shown.
gene diversity will increase the difficulty of constructing lineage early after the emergence of the ancestral subtype C presents a lineage that apparently split from the main subtype C 541–178 sequence is clearly subtype C-like (Fig. 1A), it represents support (50% bootstrap support) in both trees for an Ethiopian origin of Middle Eastern subtype C viruses. 

While there are small, well-supported exclusively South African or Botswanan groups in both trees, these may be the result of sampling bias since the overwhelming majority of southern African sequences have been obtained from these two countries. That sampling bias is responsible for at least some of the apparent clustering of South African sequences is supported by the observation that only 3 of the 30 South African HIVNET028 gag sequences and 2 of the 20 South African HIVNET028 nef sequences exclusively group with other South African sequences with greater than 50% bootstrap support. Conversely, of the South African HIVNET028 sequences five gag sequences and one nef sequence are probably (>50% bootstrap support) most closely related to non-South African sequences. Lack of clear geographical clustering of isolates obtained from different southern Africa countries is possibly evidence of the largely unrestricted movement of subtype C genotypes across the entire continent. 

The most striking feature of the gag and nef sequence phylogenies is that the southern African viruses represent the entire breadth of subtype C diversity. The high diversity of the southern African sequences in general, and those sampled in the HIVNET028 study in particular, is evidenced by the fact that in addition to all of them obviously being subtype C sequences, only 11 of 64 gag sequences and 11 of 45 nef sequences are clearly (>50% bootstrap support) most closely related to other individual subtype C sequences. 

One of the HIVNET028 gag sequences from Zambia (541–178 indicated by an arrow in Fig. 1A) is actually the most divergent subtype C gag sequence yet identified. Sequence 541–178 is also potentially an intrasubtype C recombinant. Five of the six methods used to detect recombination indicated that 541–178 had most likely inherited the sequence between nucleotide positions 959 and 1221 from a Botswanan subtype C virus, although the confidence in its recombinant nature was lower. The 541–178 sequence is also potentially an intrasubtype C recombinant. Five of the six methods used to detect recombination indicated that 541–178 had most likely inherited the sequence between nucleotide positions 959 and 1221 from a Botswanan subtype C virus, although the confidence in its recombinant nature was lower. 

ACKNOWLEDGMENTS

Support for this work was partly through NIH grant N01-AI-45202 as well as the South African AIDS Vaccine Initiative. DPM is supported by the South African National Bioinformatics Network. We would also like to thank the clinical staff at each of the clinical sites for their invaluable help in collecting and shipping samples. The HIVNET028 Study team comprised of: Clive Gray (co-chair), Haynes Sheppard (co-chair), Zambia University Teaching Hospital; Rosemary Musonda, Susan Allen and Michelle Klautzman, University of Alabama; Newton Kumwenda, Malawi College of Medicine; Taha Taha, Johns Hopkins University; Lynn Zijenah, Victoria Aquino, Michael Chirenje, Mike Mbizo, and Ocean Tobaiwa, University of Zimbabwe; David Katzenstein, Stanford University; Glenda Gray, James McIntyre and Armstrong Mafhandu, Perinatal HIV Research Unit, Chris Hani Baragwanath Hospital, Johannesburg; Efthyia Vardas, Mark Colvin, Dudu Msweli, Wendy Dlamin, Gita Ramjee, Salim Abdool Karim, Quarraisha Abdool Karim, Medical Research Council, Durban; Lynn Morris and Natasha Taylor, National Institute for Communicable Diseases; Carolyn Williamson, Helba Bredefeld and Celia Rademeyer, University of Cape Town; Jorge Flores, Division of AIDS, NIH (DAIDS); Ward Cates, Linda McNeil and Missie Allen, Family Health International.

REFERENCES

5. Kellerer AD, Long C, Holmes EC, et al.: Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27-re-
SOUTHERN AFRICAN HIV-1C gag AND nef DIVERSITY


Address reprint requests to:

Helba Bredell
Institute of Infectious Disease and Molecular Medicine
Faculty of Health Sciences
University of Cape Town
Observatory
Cape Town, South Africa 7925

E-mail: hbredell@curie.uct.ac.za
This article has been cited by:

1. Marion Vermeulen, Charl Coleman, Josephine Mitchel, Ravi Reddy, Harry van Drimmelen, Tracy Fickett, Michael Busch, Nico Lelie. 2013. Comparison of human immunodeficiency virus assays in window phase and elite controller samples: viral load distribution and implications for transmission risk. *Transfusion* n/a-n/a. [CrossRef]


11. Nobuhelo G. Ngandu, Helba Bredell, Clive M. Gray, Carolyn Williamson, Cathal Seoighe, And the HIVNET028 Study Team. 2007. CTL Response to HIV Type 1 Subtype C Is Poorly Predicted by Known Epitope Motifs. *AIDS Research and Human Retroviruses* 23:8, 1033-1041. [Abstract] [Full Text PDF] [Full Text PDF with Links]