

# Sequencing and Phylogenetic Analysis of Near Full-Length HIV-1 Subtypes A, B, G and Unique Recombinant AC and AD Viral Strains Identified in South Africa

Eduan Wilkinson,<sup>1,2</sup> Vera Holzmayer,<sup>3</sup> Graeme B. Jacobs,<sup>1</sup> Tulio de Oliveira,<sup>2</sup>  
Catherine A. Brennan,<sup>3</sup> John Hackett Jr.,<sup>3</sup> Estrelita Janse van Rensburg,<sup>4</sup> and Susan Engelbrecht<sup>1,5</sup>

## Abstract

By the end of 2012, more than 6.1 million people were infected with HIV-1 in South Africa. Subtype C was responsible for the majority of these infections and more than 300 near full-length genomes (NFLGs) have been published. Currently very few non-subtype C isolates have been identified and characterized within the country, particularly full genome non-C isolates. Seven patients from the Tygerberg Virology (TV) cohort were previously identified as possible non-C subtypes and were selected for further analyses. RNA was isolated from five individuals (TV047, TV096, TV101, TV218, and TV546) and DNA from TV016 and TV1057. The NFLGs of these samples were amplified in overlapping fragments and sequenced. Online subtyping tools REGA version 3 and jpHMM were used to screen for subtypes and recombinants. Maximum likelihood (ML) phylogenetic analysis (phyML) was used to infer subtypes and SimPlot was used to confirm possible intersubtype recombinants. We identified three subtype B (TV016, TV047, and TV1057) isolates, one subtype A1 (TV096), one subtype G (TV546), one unique AD (TV101), and one unique AC (TV218) recombinant form. This is the first NFLG of subtype G that has been described in South Africa. The subtype B sequences described also increased the NFLG subtype B sequences in Africa from three to six. There is a need for more NFLG sequences, as partial HIV-1 sequences may underrepresent viral recombinant forms. It is also necessary to continue monitoring the evolution and spread of HIV-1 in South Africa, because understanding viral diversity may play an important role in HIV-1 prevention strategies.

## Introduction

**S**OUTH AFRICA HAS THE HIGHEST NUMBER of people infected with HIV-1 worldwide, estimated at 6.1 million in 2012.<sup>1</sup> A major feature of HIV-1 is the extreme genetic diversity of the viral genome, which may have an impact on viral diagnostics, transmission, disease progression, and clinical management.<sup>2</sup> HIV consists of two types, HIV-1 and HIV-2, and HIV-1 can be further divided into four groups: M (Major), O (Outlier), N (Non-M, Non-N), and P. Group M is responsible for the pandemic and can be divided into nine

subtypes and subsubtypes, as well as into recombinant forms, which can be divided into circulating recombinant forms (CRFs) and unique recombinant forms (URFs). Currently, there are more than 65 CRFs and numerous URFs identified in the Los Alamos HIV Database [[www.hiv.lanl.gov/content/index](http://www.hiv.lanl.gov/content/index)]. The HIV-1 pandemic is not uniform, but complex and dynamic with different regional distributions of subtypes and CRFs. Subtype C is the most prevalent form in South Africa and accounts for nearly 50% of all HIV infections worldwide.<sup>3,4</sup> It is essential to continuously monitor the diversity and spread of HIV-1 worldwide as the pandemic matures.

<sup>1</sup>Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg Academic Hospital, Tygerberg, Cape Town, South Africa.

<sup>2</sup>Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa.

<sup>3</sup>Abbott Laboratories, Abbott Park, Illinois.

<sup>4</sup>Quintiles Central Laboratories, Livingston, United Kingdom.

<sup>5</sup>National Health Laboratory Services (NHLS), Western Cape Region, Tygerberg Hospital (Coastal), Cape Town, South Africa.

A total of 309 full or near full-length unique HIV-1 genomes from South Africa have been characterized in various studies, 296 (95.78 %) of which are subtype C isolates.<sup>5–10</sup> Other South African near full-length HIV-1 genomes include two subtype A1 isolates,<sup>8,11</sup> two subtype B isolates,<sup>8,11</sup> five subtype D isolates,<sup>12,13</sup> and four viral recombinant forms, which included three different URF\_AC recombinant forms and one complex URF.<sup>6,8,14</sup>

We describe the near full-length genome (NFLG) sequencing and phylogenetic analysis of seven additional South African viral strains, including HIV-1 subtypes A, B, G, and two URFs.

## Materials and Methods

### Ethics statement

This study was approved by the Health Research Ethics Committee (HREC) of Stellenbosch University (IRB0005239) and all study participants provided written informed consent for the collection of samples and subsequent analyses.

### Patients and RNA/DNA isolation

Plasma and peripheral blood mononuclear cell (PBMC) samples from the Tygerberg Virology (TV) cohort were obtained between 1998 and 2004. The TV cohort, which was previously described in Jacobs *et al.*,<sup>15</sup> is a rich sample repository containing specimens from patients from wide and diverse backgrounds based on race, socioeconomic status, and sexual orientation. Viral genotyping was performed on the envelope region of 410 sequences<sup>15</sup> and the partial *gag*, *pol*, and *env* regions of a further 10 sequences.<sup>11</sup> A total of 35 (8.53%) non-C isolates were identified among the 410 samples from the TV cohort. Of these, seven non-C strains were selected for further characterization based on the availability and quantity of samples.

RNA was extracted from 1 ml of the plasma samples (TV047, TV096, TV101, TV218, and TV546) using the QIAamp Ultrasense Virus kit. High-molecular-weight DNA was extracted from cultured TV016 using the Qiagen DNAeasy Blood and Tissue kit and from uncultured TV1057 using the QIAamp DNA Mini kit (Qiagen, GmbH, Hilden, Germany). These two samples were genotyped from proviral DNA due to the lack of plasma in the case of TV016, the difficulty in amplifying from RNA in the case of TV1057, and the fact that these two patients were on treatment at the time of sampling.

### RT-PCR and sequencing of TV047, TV096, TV101, TV218, and TV546

Between four and six overlapping fragments spanning the genome of HIV-1 were amplified using a nested long-range reverse transcriptase polymerase chain reaction (RT-PCR) method. Reverse transcription was performed with Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA) for cDNA synthesis, as described previously.<sup>16</sup> Primer sequences, not provided here, can be found in previous publications.<sup>17–19</sup> Primers p24-7, poli8R, JH38R, and 9131R-2 were used for cDNA synthesis. For PCR amplification, primer combinations RE9737F/p24-7 (first round) and RE9745/MOp24-6 (nested PCR) were used to amplify a 1.2-kb fragment from the 5' long terminal repeat (LTR) to *gag* p24. Primer combinations VFgag8/LPpoli-2 and VFgag9/LPpoli-4 (nested PCR) were used to amplify a 3.8-kb fragment spanning *gag* p24–*pol* IN

from four specimens and VFgag8/ppr4b and VFgag9/ppr10 from specimen TV218. Primer pairs 4274F-2 (5' ACAGCAG TACAAATGGCAGTATTCATTC)/LP7728R and 4277F-2 (5' GCAGTACAAATGGCAGTATTCAT)/LP7725R were utilized to amplify a 3.4-kb *pol–env* region from TV096, TV101, and TV546. Primer pairs ppf4/IDRps3R and ppf5B/LP7632R (5' TATCCATTGCAGCCAGGTCAT) were used to amplify a 3.2-kb *pol* IN–*env* IDR from TV218.

The region spanning *pol–env* in TV047 was amplified in two overlapping fragments: *pol–env* V3 (2.3 kb) using primers ppf4b/V3vh2R (5' AAAAATCCCCCTCCACA) and ppf5b/V3vh4R (5' GTGCRTTACAATTTTCYGGGTCC) and an *env*V3- IDR fragment using primer pairs V3vh1F (5' TAG GCCAGYAGTRTCAAC)/JH38R and V3vh3F (5' GCAGT CTRGCAGAARAAGAGGTARTA)/IDRps3R. To complete coverage of the *pol–env* region for TV047 and TV218, an additional *pol*-IN product was amplified using primers and conditions as previously described.<sup>17</sup>

A 1.6-kb *env* IDR-3' LTR product was generated with the primer combinations 7496F/9131R-2 and 7542F/9110R-2 from TV096 and TV218. Primer pairs JH41/9131R-2 and env-27F/9110R-2 amplified this region from TV047 and TV101, and a combination of primers env-27F/9131R-2 and 7542F/9110R-2 was used for amplification from specimen TV546.

PCR amplifications were performed with Advantage-2 polymerase Mix (Clontech, Palo Alto, CA) at cycling conditions described earlier.<sup>16</sup> PCR products were purified with the QIAquick PCR purification kit (Qiagen Inc.) and both strands were sequenced directly using the ABI Prism Big Dye Terminator Cycle Sequencing Reactions kit v. 1.0 (Applied Biosystems) and the ABI PRISM 3100 Genetic Analyser (Applied Biosystems). Sequence data were assembled and edited using Sequencher software Version 4.0.5 (Gene Code Corporation, Ann Arbor, MI). Positions with sequence ambiguities were assigned the appropriate IUPAC designations.

### PCR and sequencing of TV016 and TV1057

Four overlapping fragments, LTR–*gag* (1.09 kb), *gag–pol* (3.89 kb), *pol–env* (3.94 kb), and *env–LTR* (1.64 kb), were amplified using GoTaq DNA polymerase (Promega, Madison, WI) as described previously.<sup>11</sup> Sequencing reactions were done with the ABI Prism BigDye Terminator Cycle sequencing kit v. 1.0 and run on the ABI 3130xl automated DNA sequencer (Applied Biosystems, Foster City, CA). Sequenced data were assembled into contiguous fragments and edited in Sequencher Version 4.8 (Gene Codes Corporation, Ann Arbor, MI).

### Sequence quality analysis and preliminary subtyping using online tools

The HIV-1 sequence quality analysis tool was run before further analysis ([www.hiv.lanl.gov/content/sequence/QC/index.html](http://www.hiv.lanl.gov/content/sequence/QC/index.html)).

Sequences were then screened with online HIV-1 viral subtyping and recombinant detection tools: jpHMM (<http://jpymm.gobics.de>)<sup>20,21</sup> and REGA v. 3.0 (<http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/>).<sup>22</sup>

### SimPlot bootscan analyses

A multiple alignment was done with the HIV-1 reference subtypes ([www.hiv.lanl.gov/content/sequence/NEWALIGN/](http://www.hiv.lanl.gov/content/sequence/NEWALIGN/))

TABLE 1. DEMOGRAPHIC AND CLINICAL DATA OF THE SEVEN HIV-1-INFECTED PATIENTS

Patient	Sample date	Sex, age, ethnicity <sup>a</sup>	Mode of infection	Viral load (copies/ml blood)	CD4 cell count (cells/mm <sup>3</sup> blood)	ART	Origin
TV016	04-1998	M, 39, MR	Heterosexual	No data	421	Yes	Western Cape
TV047	08-2000	M, 29, MR	Heterosexual	29,800	265	No	Western Cape
TV096	09-2000	F, 36, Af	Heterosexual	204,020	96	No	Western Cape
TV101	09-2000	F, 23, Af	Heterosexual	88,558	2000	No	South Africa
TV218	11-2000	F, 25, Af	Heterosexual	76,117	No data	No	KwaZulu-Natal
TV546	08-2001	F, 31, Af	Heterosexual	15,985	No data	No	Eastern Cape
TV1057	04-2002	M, 53, Ca	Homosexual	140,000	108	Yes	Western Cape

<sup>a</sup>M, male; MR, mixed race; F, female; Af, African; Ca, caucasian  
ART, antiretroviral treatment.

align.html) and the new TV sequences using Clustal W.<sup>23</sup> The alignment was edited in Se-Al v. 2.0 (<http://tree.bio.ed.ac.uk/software/seal>) and then used for bootscan analysis in SimPlot v. 3.5.1. All of the bootscans were performed with the Kimura two-parameter nucleotide substitution, a window size of 400 bp, and a step size of 50 bp.

#### Maximum likelihood (ML) phylogenetic tree inference

We compiled a dataset that included the HIV-1 subtype reference dataset from the Los Alamos Database, as well as randomly selected additional sequences. Multiple sequence alignments were done with Clustal W.<sup>23</sup> Thereafter, all of the genes of HIV-1 [with the exception of the long terminal repeats (LTR) and the *nef*-coding region] were concatenated in Se-Al v 2.0 (<http://tree.bio.ed.ac.uk/software/seal>). Gene fragments were excised and overlapping gene regions were deleted from the structural genes (*gag*, *pol*, and *env*), while still conserving the open reading frames. This concatenated alignment was manually edited to obtain a codon alignment.

A modeltest was performed in jModelTest 1.0 using the Akaike Information Criterion (AIC) method to estimate the best-fitting model of nucleotide substitution. Maximum likelihood tree topology was inferred in phyML v. 3.0.<sup>24</sup> The maximum likelihood tree topology was inferred with the GTR, an estimated Gamma shape parameter, and the subtree pruning and regrafting (SPR) method of tree rearrangement. Branch support was calculated with the implementation of bootstrap resampling totaling 100 bootstrap replicates.

#### Phylogenetic analyses of possible recombinants, TV101 and TV218

Based on the breakpoints identified with online tools, ML tree topologies were inferred for each of the recombinant

fragments. Each fragment was aligned with the HIV-1 subtype reference alignment in Clustal W and edited in Se-Al v 2.0 as described before (<http://tree.bio.ed.ac.uk/software/seal>). ML tree topologies were inferred for each of the recombinant fragment alignments in phyML v. 3.0,<sup>24</sup> with the HKY85 model of nucleotide substitution and an estimated Gamma shape parameter and bootstrap resampling totaling 100 replicates.

The AC recombinant NFLG sequences described previously in South Africa (AF411956, GU201611, and DQ093606) were investigated with TV218, using jpHMM to determine if the breakpoints were similar and if we had a possible new CRF.

## Results

### Patient demographics

All available clinical and demographic data of the seven patients are summarized in Table 1. The mean viral load was 61,608 RNA copies/ml (SD=71,246.257), while the mean CD4<sup>+</sup> cell count was 548 cells/mm<sup>3</sup> (SD=1,139.721). The dates of sampling collection predate the implementation of the public national HIV treatment campaign and thus only two of the patients were receiving antiretroviral therapy at the time of sampling.

### Near full-length genomic sequences

Assembly of the overlapping amplification products TV47 (9,083 nucleotides), TV96 (9,058 nucleotides), TV101 (9,038 nucleotides), TV218 (9,039 nucleotides), and TV546 (9,104 nucleotides) resulted in characterization of NFLG spanning from the 5' U5 region through the 3' U3 region. Open reading frames (ORFs) were identified for *gag*, *pol*, and *env* structural genes and for *vif*, *vpr*, *vpu*, *nef*, *tat*, and *rev*

TABLE 2. SUBTYPING OF ALL SEVEN NEW SOUTH AFRICAN HIV-1 NON-SUBTYPE C STRAINS

Sequence name	REGA 3.0	jpHMM	SimPlot	ML-tree	Subtype assigned	GenBank accession
TV016	B	B/K	B	B	B	KJ948656
TV047	B	B	B	B	B	KJ948657
TV096	A1	A1	A1	A1	A1	KJ948658
TV101	A1/D	A1/D	A1/D	A1	URF A1/D	KJ948659
TV218	A1/C	A1/C	A1/C	C outlier	URF A1/C	KJ948661
TV546	G	G	G	G	G	KJ948662
TV1057	B	B	B	B	B	KJ948660

ML, maximum likelihood; jpHMM, jumping profile hidden Markov model.



regulatory/accessory genes. TV016 (8,039 nucleotides) and TV1057 (8,084 nucleotides) spanned from the beginning of the *gag* to the end of the *env* region and excluded the *nef*ORF.

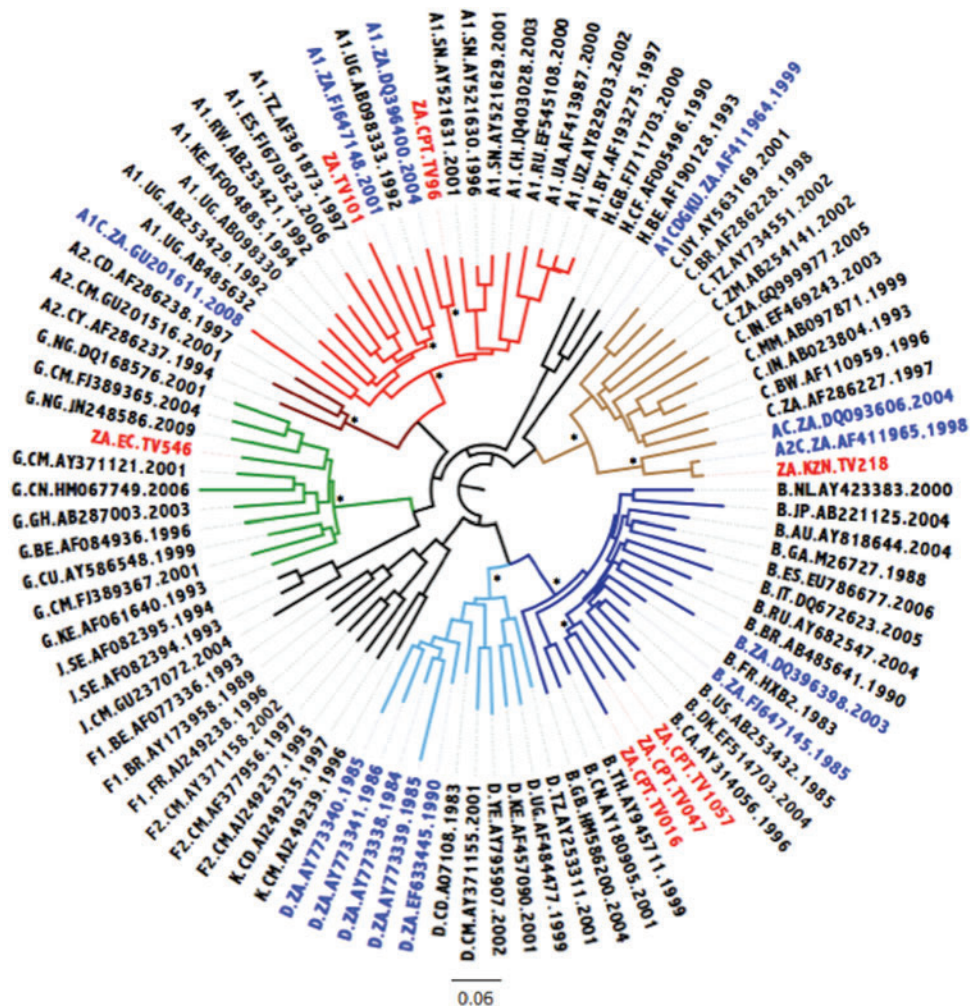
Online subtyping and bootscan analyses

Preliminary subtyping was done using the online REGA version 3 and jpHMM tools, summarized in Table 2. These tools provide a quick, preliminary analysis of the sequences before detailed manual phylogenetic inference. Three isolates (TV016, TV047, and TV1057) were identified with high confidence as HIV-1 subtype B. TV096 was identified as HIV-1 subtype A1 and TV546 as subtype G. The REGA and jpHMM tools identified two possible recombinants, TV101 and TV218. SimPlot bootscan analysis identified TV101 as an A1/D recombinant and TV218 as an A1/C recombinant. TV016, TV047, and TV1057 were confirmed as HIV-1 sub-

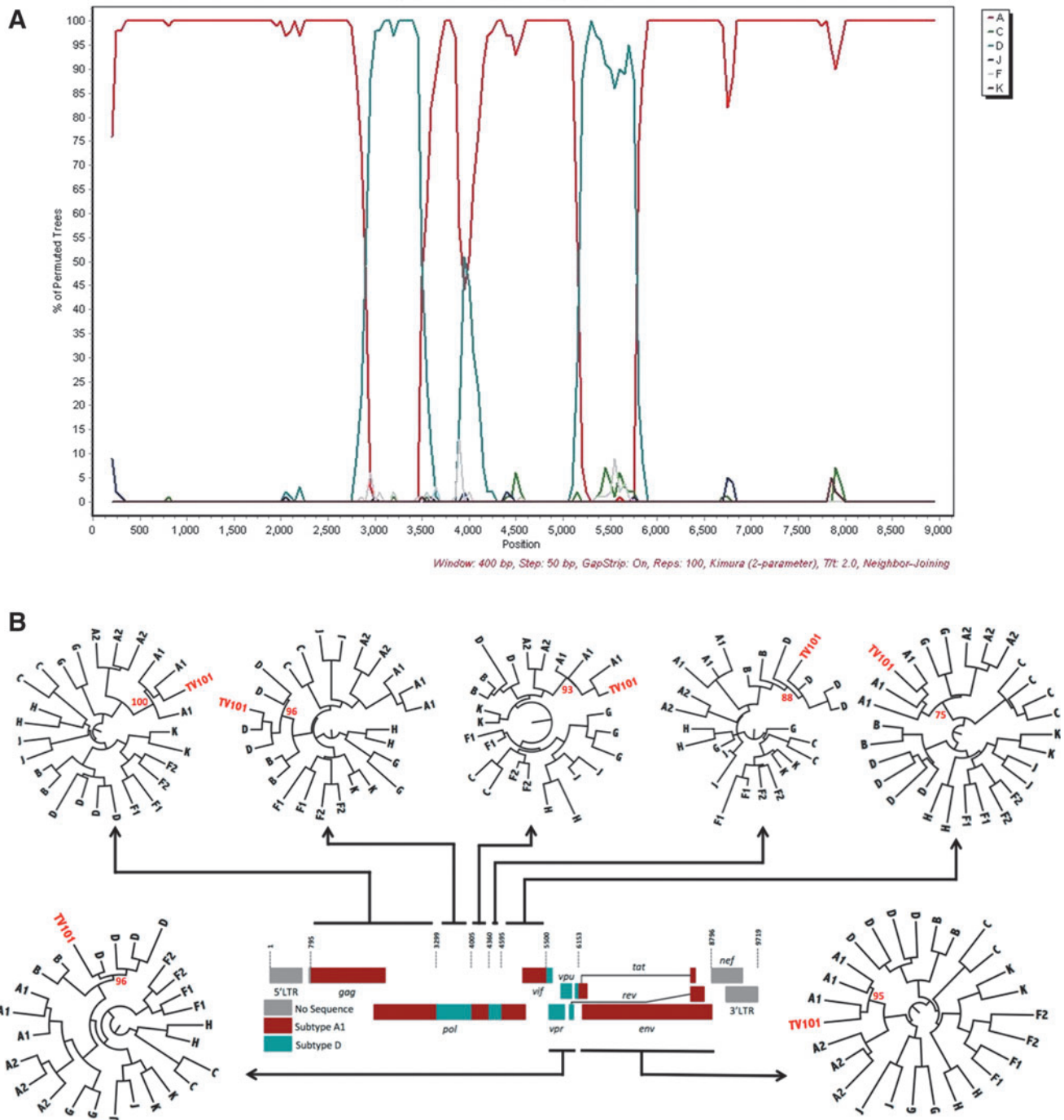
type B with high similarity scores using SimPlot and TV096 and TV546 as subtype A1 and subtype G, respectively.

ML phylogenetic tree inference

The ML tree is indicated in Fig. 1. Three isolates, TV016, TV047, and TV1057, clustered within HIV-1 subtype B with high bootstrap support. TV546 clustered within HIV-1 subtype G with high support, while TV218 clustered as an outlier to the subtype C cluster. The remaining two strains, TV096 and TV101, clustered within HIV-1 subtype A1 with high support. Based on the phylogenetic inference it would seem that TV218, as an outlier to the larger subtype C cluster, may represent a subtype C viral recombinant form. TV101 clustered within the A1 clade but this can possibly be explained by the small recombinant regions of subtype D that are interspersed within the larger subtype A1 sequence.



**FIG. 1.** Maximum likelihood (ML) tree inference of the near full-length genome (NFLG) concatenated sequence alignment. The ML tree inferred in phylogenetic analysis (phyML) contains the seven newly sequenced Tygerberg Virology (TV) isolates, HIV-1 reference strains, and 13 previously characterized non-subtype C strains from South Africa. The evolutionary distances were computed using the GTR model of nucleic acid substitution with an estimated Gamma shape parameter. The genetic distance is displaying in the scale bar at the bottom of the figure, while the major different clades of HIV-1 Group M have been highlighted. The sequence IDs of South African strains have been marked with the 7 newly genotyped isolates marked in red and the 13 previously genotyped isolates marked in blue. Bootstrap support values for the internal branches for each major clade are shown with an asterisk and indicate support higher than 90%. Color images available online at [www.liebertpub.com/aid](http://www.liebertpub.com/aid)

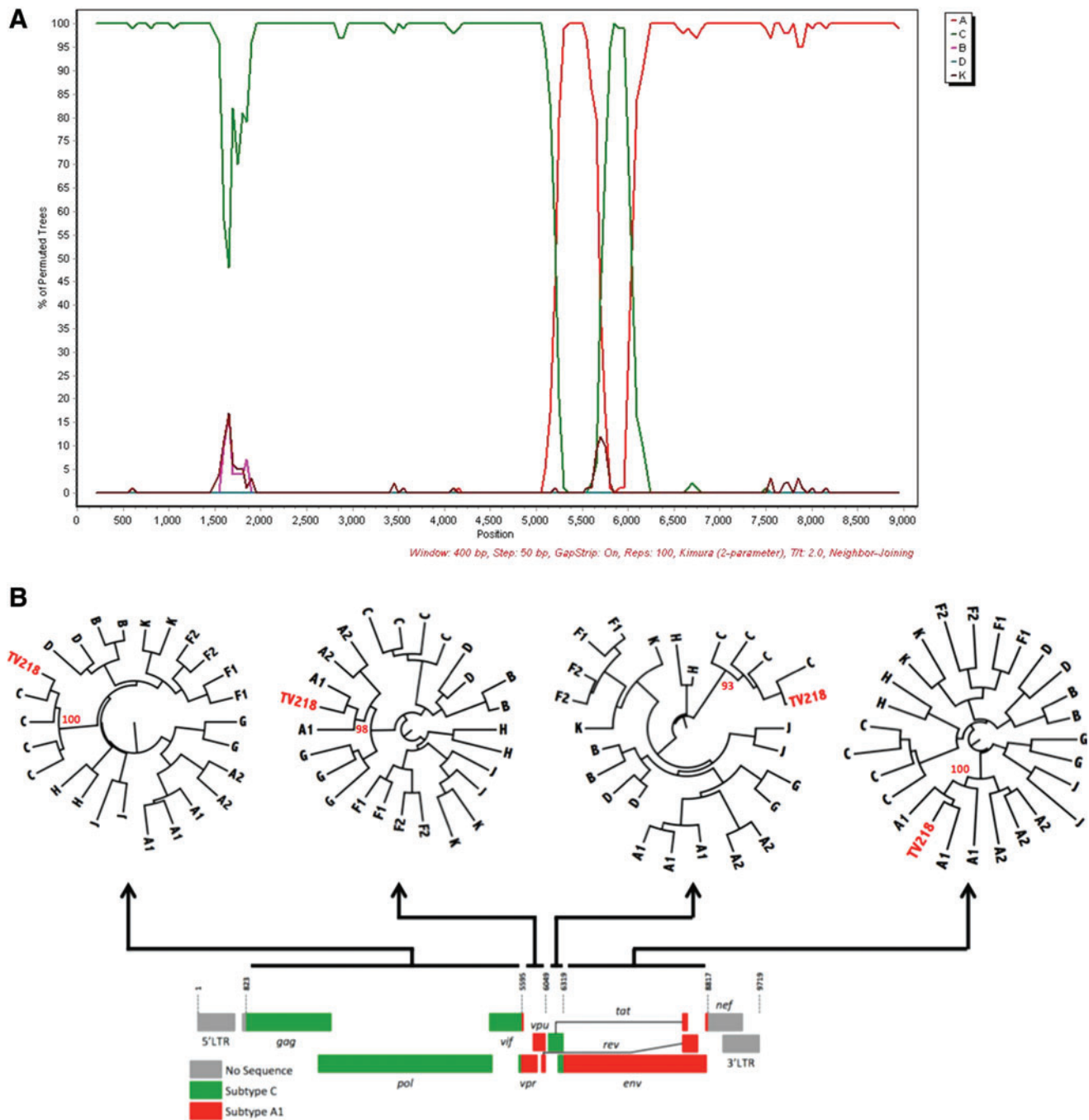


**FIG. 2.** Unique recombinant form of TV101. **(A)** Bootscan similarity plot constructed in SimPlot v. 3.5.1 shows the recombinant profile of this viral isolate. **(B)** Schematic diagram of A1 and D recombinant segments of TV101. ML trees of TV101 fragments with complimentary reference sequences were inferred in phyML with a total of 1,000 bootstrap replicates. The evolutionary distances were computed using the GTR model of nucleic acid substitution with an estimated Gamma shape parameter. Each *roman numeral* in the similarity plot at the *top* corresponds to a phylogeny in the *bottom* part of the schematic. TV101 is marked in *red* in each of the seven different phylogenies, while bootstrap support values for each of the clades in which the isolate clusters are also shown in *red*. Color images available online at [www.liebertpub.com/aid](http://www.liebertpub.com/aid)

*Phylogenetic analysis of TV101 and TV218 recombinants*

TV101 and TV218 were further investigated and phylogenies were also inferred from the recombinant breakpoints identified with online jpHMM, REGA version 3, and bootscan

analyses. Representations of the genome mosaic of TV101 and TV218 are illustrated in Fig. 2 and Fig. 3, respectively. TV101 is a recombinant between HIV-1 subtype A1 and D with six breakpoints and TV218 is a recombinant between HIV-1 subtype C and A1 with four breakpoints. The breakpoints of all the other South African AC sequences are illustrated in Fig. 4.



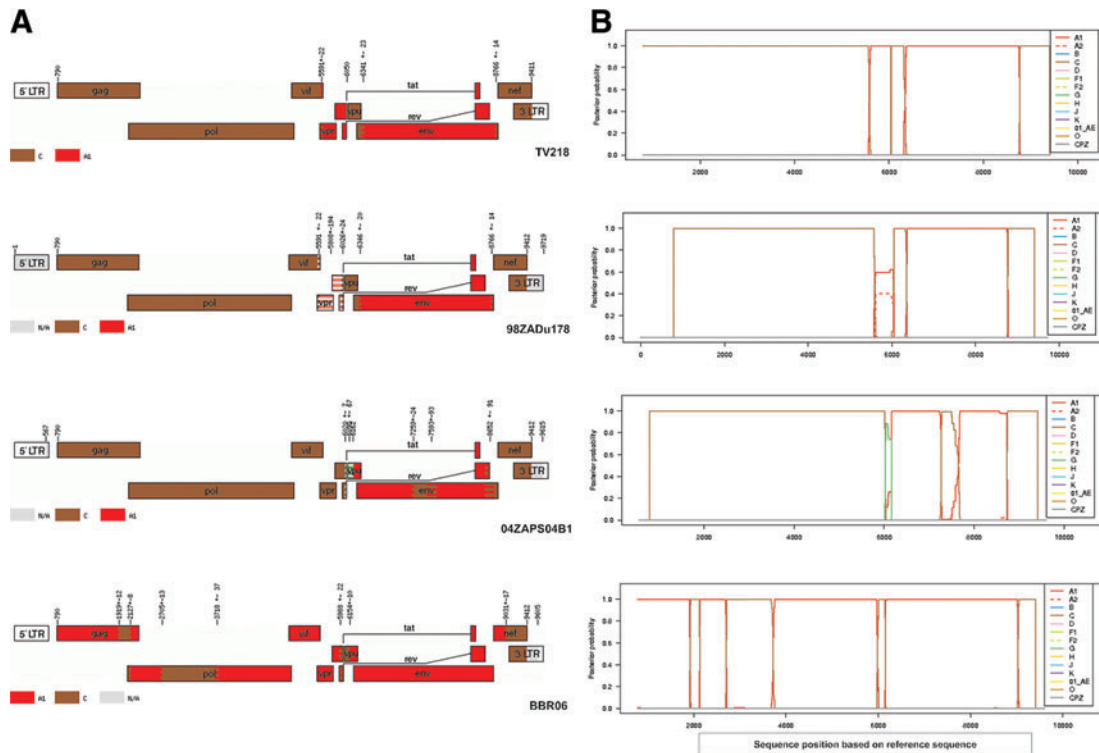
**FIG. 3.** Unique recombinant form of TV218. **(A)** Bootscan similarity plot constructed in SimPlot v. 3.5.1 shows the recombinant profile of this viral isolate. **(B)** Schematic diagram of A1 and C recombinant segments of TV218. Radiation maximum trees of TV218 fragments with complimentary reference sequences were implemented in phyML with a total of 1,000 bootstrap replicates. The evolutionary distances were computed using the GTR model of nucleic acid substitution with an estimated Gamma shape parameter. Each *roman numeral* in the similarity plot in the *top* corresponds to a phylogeny in the *bottom* part of the schematic. TV218 is marked in *red* in each of the four different phylogenies, while bootstrap support values for each of the clades in which the isolate clusters are also shown in *red*. Color images available online at [www.liebertpub.com/aid](http://www.liebertpub.com/aid)

**Discussion**

Limited information is available for HIV-1 non-subtype C sequences in South Africa and currently only 13 non-subtype C NFLG sequences are listed in the LANL HIV database. In this study, seven new NFLG sequences were characterized: three

subtype Bs (TV016, TV047, and TV1057), one subtype A1 (TV096), one subtype G (TV546), one unique AD (TV101), and one unique AC (TV218) recombinant form. This is the first NFLG of subtype G that has been characterized from South Africa. Only three subtype B NFLGs have been described from Africa and with this article we increase this number to six.





**FIG. 4.** jpHMM analysis of South African AC recombinants. (A) Indication of the recombinant breakpoints based on HXB2 numbering. The *uncolored* regions denote missing information due to the input fragment sequence and the *gray* regions denote missing information due to uninformative subtype models. (B) Posterior probabilities of the subtypes at each sequence position (original sequence positions) calculated by jpHMM. Color images available online at [www.liebertpub.com/aid](http://www.liebertpub.com/aid)

These NFLG subtype B sequences include virus strains isolated from both heterosexual and homosexual individuals. All of these NFLGs identified in the present study were characterized in HIV-1-infected South Africans. Traditionally, subtype C has been the predominant viral form of HIV-1 in South Africa and today the subtype still accounts for the majority of infections (>95.0%). However, in recent years we have seen an increase in the number of non-C HIV-1 isolates characterized among South African individuals (E. Wilkinson and G.B. Jacobs, unpublished observations). It is of the utmost importance to continue to monitor the genetic diversity of the HIV-1 epidemic within the country, as increasing heterogeneity can potentially impact the design of an effective vaccine, viral diagnostic assays, disease progression, and treatment and may lead to the rise of more recombinant forms. There are currently more than 65 CRFs identified in the Los Alamos Database and in 2011 they were responsible for at least 20% of HIV-1 infections worldwide.<sup>3</sup>

#### HIV-1 subtype B in South Africa

Previously, only two subtype B NFLGs have been characterized from South Africa<sup>8,11</sup> and another one from the African continent.<sup>25</sup> This isolate from Gabon shows no close phylogenetic relationship with any of the South African subtype B strains. Two of the subtype B strains (TV016 and TV047) were genotyped from patients who were heterosexually infected, while the other remaining patient, TV1057, became infected via homosexual contact in 1982. He is classified as a slow progressor and was receiving anti-

retroviral therapy when the sample was taken. Two different HIV epidemics have been described within South Africa: HIV-1 subtype B in homosexual men represented the early epidemic and accounted for the majority of HIV infections during the 1980s<sup>26</sup> and HIV-1 subtype C in the heterosexual population caused a later (or second) epidemic and is currently the most prevalent subtype.<sup>27</sup>

Although TV1057 was sampled in 2001, the infection occurred in 1982, providing an opportunity to analyze a subtype B strain that originated from the time of the early epidemic in the country. This strain was most closely related to GenBank accession number EF363124 from the United States as identified through a BLAST search. TV016 (infected heterosexually in 1989) and TV047 are characteristic of an emerging subtype B epidemic occurring in the heterosexual population, indicating a crossover of the two epidemics.<sup>15,28,29</sup> TV016 was most closely related to GenBank accession number EF363124 and TV047 to GenBank accession number AY835795.

#### HIV-1 subtype G in South Africa

TV546 is the first NFLG subtype G strain that has been characterized within South Africa. The first full-length HIV-1 subtype G sequences were described in 1998 originating from the Democratic Republic of the Congo.<sup>30</sup> The majority of HIV-1 subtype G isolates identified originate from several West Central and East African countries.<sup>31–33</sup> Subtype G is also included in nosocomial epidemic outbreaks in the former Soviet Union<sup>34</sup> and more recently has been linked to spread among infected drug users in the Iberian peninsula.<sup>35,36</sup> There

has been one report of the detection of subtype G *gag* sequences in South Africa that was generated from samples that were obtained from migrant workers from Nigeria, Kenya, Zambia, and Angola.<sup>37</sup> Compared to other subtypes, subtype G occurs infrequently and TV546 is only the second report of this subtype in South Africa. The strain was detected in a female patient residing in a rural area of the Eastern Cape Province and was most closely related to isolate 944-5 from Cameroon (FJ389366) as indicated by BLAST. This patient, who became infected via heterosexual contact, had no history of travel outside of South Africa.

#### *HIV-1 subtype A1 in South Africa*

TV096 is only the third NFLG of the subtype A1 isolate characterized from South Africa. The first full-length HIV-1 subtype A1 isolate from South Africa was characterized from a female patient of African descent,<sup>8</sup> while the second was characterized from an African male.<sup>11</sup> TV096 was sampled from a female patient of African descent in the late stages of HIV infection, as characterized by World Health Organization (WHO) criteria. Blast and phylogenetic inference indicated that TV096 was more closely related to African HIV-1 subtype A1 isolates from Senegal and Uganda. Phylogenetics of the previously characterized A1 sequences from South Africa also showed a close genetic relationship with other HIV-1 subtype A1 isolates from the East African region.

#### *Unique recombinant forms in South Africa*

Two unique recombinant forms were also identified in this study. TV101 is the second URF\_AD described from South Africa and is most closely related to AF457082 from Kenya. The first URF\_AD recombinant was characterized from a South African individual who became infected via heterosexual contact in Kenya.<sup>11</sup> These two URF\_AD sequences share no recombinant breakpoints. HIV-1 subtypes D and subtype A1 have been detected in South Africa in the past.<sup>8,11,12</sup>

In addition to the URF\_AD, TV218, a URF\_AC was also characterized in the present study. This is the fourth URF composed of subtypes A and C that has been identified in South Africa. Two A1/C recombinants, isolate 04ZAPS204B1 (GenBank accession DQ093606)<sup>8</sup> and isolate BBCR06 (GenBank accession GU201611),<sup>14</sup> and one A2/C recombinant, 98ZADu178 (GenBank accession AF411965),<sup>6</sup> have already been described. TV218, 98ZADu178, and 04ZAPS204B1 were sampled in Durban, KwaZulu-Natal, while the other A1/C isolate was sampled in the far northern part of South Africa.<sup>14</sup>

TV218 revealed a close genetic similarity to a previously described A2C subtype (AF411965, isolate 98ZADu178) from South Africa.<sup>6</sup> The 98ZADu178 sequence was derived from cultured cells from an asymptomatic sex worker in Durban (sample date 1998), while TV218 was directly amplified and sequenced from plasma obtained during 2000 from a 25-year-old female in Durban. These two A1C sequences are 97% similar and share similar breakpoints. The *vpr*, *tat*, and *rev* region of TV218 is subtype A1 with a high probability, whereas this region in 98ZADu178 is A1/A2 with a low probability.

## Conclusions

Phylogenetic inference of seven newly sequenced HIV-1 strains identified subtypes A1, B, and G as well as URF\_AC and URF\_AD. There is a need for more NFLG sequences because partial HIV-1 sequences may underrepresent viral recombinant forms. It is necessary to continue monitoring the evolution and spread of HIV-1 in South Africa and worldwide. Understanding HIV-1 diversity in South Africa will play an important role in HIV-1 prevention strategies.

## Sequence Data

The sequences analyzed during the study have been deposited in GenBank and are available under the following accession numbers: KJ948656 to KJ948662.

## Acknowledgments

This study was funded by the Poliomyelitis Research Foundation (PRF), the National Research Foundation (NRF), and the Medical Research Council (MRC) of South Africa. This research project was funded by the South African Medical Research Council (MRC) with funds from the National Treasury under its Economic Competitiveness and Support Package. This research and the publication thereof are the result of funding provided by the Medical Research Council of South Africa in terms of the MRC's Flagships Awards Project MRC-RFA-UFSP-01-2013/ UKZNHIVEPI.

## Author Disclosure Statement

No competing financial interests exist.

## References

1. Shisana O, Rehle T, Simbayi LC, *et al.*: *South African National HIV Prevalence, Incidence and Behaviour Survey, 2012*. HSRC Press, Cape Town.
2. Hemelaar J: Implications of HIV diversity for the HIV-1 pandemic. *J Infect Dis* 2013;66(5):391–400.
3. Hemelaar J, Gouws E, Ghys PD, *et al.*: WHO-UNAIDS Network for HIV Isolation and Characterization. Global trends in molecular epidemiology of HIV-1 during 2000–2007. *AIDS* 2011;25(5):679–689.
4. Santos AF and Soares MA: HIV genetic diversity and drug resistance: Review. *Viruses* 2010;2(2):503–531.
5. van Harmelen J, Williamson C, Kim B, *et al.*: Characterization of full-length HIV type 1 subtype C sequences from South Africa. *AIDS Res Hum Retroviruses* 2001;17(16):1527–1531.
6. Papathanasopoulos MA, Cilliers T, Morris L, *et al.*: Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and intersubtype recombinants isolated in South Africa. *AIDS Res Hum Retroviruses* 2002;18(12):879–886.
7. zur Megede J, Engelbrecht S, de Oliveira T, *et al.*: Novel evolutionary analyses of full-length HIV type 1 subtype C molecular clones from Cape Town, South Africa. *AIDS Res Hum Retroviruses* 2002;18(17):1327–1332.
8. Rousseau CM, Birditt BA, McKay AR, *et al.*: Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J Virol Methods* 2006;136(1–2):118–125.
9. Treurnicht FK, Seoighe C, Martin DP, *et al.*: Adaptive changes in HIV-1 subtype C proteins during early infection



- are driven by changes in HLA-associated immune pressure. *Virology* 2010;396(2):213–225.
10. Liu MK, Hawkins N, Richie AJ, *et al.*: Vertical T cell immunodominance and epitope entropy determine HIV-1 escape. *J Clin Invest* 2013;123(1):380–393.
  11. Wilkinson E and Engelbrecht S: Molecular characterization of non-subtype C and recombinant HIV-1 viruses from Cape Town, South Africa. *Infect Genet Evol* 2009;9(5):840–846.
  12. Loxton AG, Treurnicht F, Laten A, *et al.*: Sequence analysis of near full-length HIV type 1 subtype D primary strains isolated in Cape Town, South Africa, from 1984 to 1986. *AIDS Res Hum Retroviruses* 2005;21(5):410–413.
  13. Jacobs GB, Loxton AG, Laten A, and Engelbrecht S: Complete genome sequencing of a non-syncytium inducing HIV-1 subtype D strain from Cape Town, South Africa. *AIDS Res Hum Retroviruses* 2007;23(12):1575–1578.
  14. Iweriebor BC, Bessong PO, Mavhandu LG, *et al.*: Genetic analysis of the near full-length genome of an HIV type 1 A1/C unique recombinant form from Northern South Africa. *AIDS Res Hum Retroviruses* 2011;27(8):911–915.
  15. Jacobs GB, Loxton AG, Laten A, *et al.*: Emergence and diversity of different HIV-1 subtypes in South Africa, 2000–2001. *J Med Virol* 2009;81:1852–1859.
  16. Holzmayr V, Aitken C, Skinner C, *et al.*: Characterization of genetically diverse HIV type 1 from a London cohort: Near full-length genomic analysis of a subtype H strain. *AIDS Res Hum Retroviruses* 2009;25(7):721–726.
  17. Swanson P, Devare SG, and Hackett J Jr: Molecular characterization of 39 HIV-1 isolates representing group M (subtypes A-G) and group O: Sequence analysis of gag p24, pol, integrase, and env gp41. *AIDS Res Hum Retroviruses* 2003;19(7):625–629.
  18. Swanson P, Devare SG, and Hackett J Jr: Full-length sequence analysis of HIV-1 isolate CM237: A CRF01\_AE/B intersubtype recombinant from Thailand. *AIDS Res Hum Retroviruses* 2003;19(8):707–712.
  19. Holzmayr V, Zekeng L, Kaptué L, *et al.*: Near-full-length genomic sequence of a human immunodeficiency type 1 subtype G strain from Cameroon. *AIDS Res Hum Retroviruses* 2005;21(5):414–419.
  20. Spang R, Rehmsmeier M, and Stoye J: A novel approach to remote homology detection: Jumping alignments. *J Comp Biol* 2002;9:747–760.
  21. Zhang M, Schultz AK, Calef C, *et al.*: jpHMM at GOBICS: A web server to detect genomic recombinations in HIV-1. *Nucleic Acids Res* 2006;34:463–465.
  22. Peña ACP, Faria NR, Imbrechts S, *et al.*: Performance of the subtyping tools in the surveillance of HIV-1 epidemic: Comparison between Rega version 3 and six other automated tools to identify pure subtypes and circulating recombinant forms. *Infect Genet Evol* 2013;19:337–348.
  23. Thompson JD, Gibson TJ, Plewniak F, *et al.*: The CLUSTAL X windows interface: Flexible strategies for multiple-sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882.
  24. Guindon S, Dufayard J-F, Lefort V, *et al.*: New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of phyML 3.0. *Syst Biol* 2010;59(3):307–321.
  25. Huet T, Dazza MC, Brun-Vezinet F, *et al.*: A highly defective HIV-1 strain isolated from a healthy Gabonese individual presenting an atypical Western blot. *AIDS* 1989;3(11):707–715.
  26. Sher R: HIV infection in South Africa, 1982–1988—review. *South African Med J* 1989;76:314–318.
  27. van Harmelen J, Wood R, Lambrick M, *et al.*: An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* 1997;11:81–87.
  28. Jacobs GB, Wilkinson E, Isaacs S, *et al.*: HIV-1 subtypes B and C unique recombinant forms (URFs) and transmitted drug resistance identified in the Western Cape Province, South Africa. *PLoS One* 2014;9(3):e90845.
  29. Middelkoop K, Rademeyer C, Brown BB, *et al.*: Epidemiology of HIV-1 subtypes among men who have sex with men in Cape Town, South Africa. *J Acquir Immune Defic Syndr* 2014;65(4):473–480.
  30. Oelrichs RB, Vandamme AM, van Laethem K, *et al.*: Full-length genomic sequence of an HIV type 1 subtype G from Kinshasa. *AIDS Res Hum Retroviruses* 1999;15(6):585–589.
  31. Abimiku AG, Stern TL, Zwandor A, *et al.*: Subtype G HIV type 1 isolates from Nigeria. *AIDS Res Hum Retroviruses* 1994;10(11):1581–1583.
  32. Kaleebu P, Bobkov A, Cheingsong-Popov R, *et al.*: Identification of HIV-1 subtype G from Uganda. *AIDS Res Hum Retroviruses* 1995;11(5):657–659.
  33. 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(4):315–325.
  34. Bobkov A, Cheingsong-Popov R, Garaev M, *et al.*: Identification of an env G subtype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. *AIDS* 1994;8:1649–1655.
  35. Holguin A, Amparo A, and Vincent S: High prevalence of HIV-1 subtype G and natural polymorphisms at the protease gene among HIV infected immigrants in Madrid. *AIDS* 2002;16(8):1163–1170.
  36. Esteves A, Parreira R, Piedade J, *et al.*: Spreading of HIV-1 subtype G and envB/gag recombinant strains amongst injecting drug users in Lisbon, Portugal. *AIDS Res Hum Retroviruses* 2003;19(6):511–517.
  37. Bredell H, Hunt G, Casteling A, *et al.*: HIV-1 subtypes A, D, G, AG, and unclassified sequences identified in South Africa. *AIDS Res Hum Retroviruses* 2002;18(9):681–683.

Address correspondence to:  
 Susan Engelbrecht  
 Division of Medical Virology  
 P.O. Box 19063,  
 Tygerberg 7505  
 Western Cape Province  
 South Africa  
 E-mail: susanen@sun.ac.za