

# SCIENTIFIC REPORTS

OPEN

## Near full-length HIV-1 subtype B sequences from the early South African epidemic, detecting a BD unique recombinant form (URF) from a sample in 1985

Adetayo Emmanuel Obasa<sup>1</sup>, Susan Engelbrecht<sup>1,2</sup> & Graeme Brendon Jacobs<sup>1</sup>

HIV-1 subtype C is the most prevalent subtype in South Africa. Although subtype B was previously detected in South Africa, there is limited sequence information available. We characterized near full-length HIV-1 subtype B sequences from samples collected at the start of the South African HIV-1 epidemic, in the 1980s. Five samples were analysed by PCR amplification, Sanger DNA sequencing and phylogenetic analyses. The viral genomes were amplified in two overlapping fragments of 5.5 kb and 3.7 kb. The sequences were subtyped using REGA version 3.0, RIP version 3.0 and jpHMM. Maximum Likelihood phylogenetic trees were inferred with MEGA version 6. Four HIV-1 patient sequences were subtyped as pure HIV-1 subtype B. One sequence was characterized as a novel HIV-1 subtype B and D recombinant. The sequences clustered phylogenetically with other HIV-1 subtype B sequences from South Africa, Europe and the USA. We report the presence of an HIV-1 subtype B and D recombinant strain detected in the beginning of the epidemic. This indicates that viral recombination events were already happening in 1985, but could have been missed as sequence analyses were often limited to small genomic regions of HIV-1.

Since the beginning of the HIV-1 pandemic more than 35.4 million (25.0–49.9 million) people have died of AIDS-related illnesses<sup>1</sup>. South Africa experiences the largest HIV-1 epidemic with 7.2 million infected individuals<sup>2</sup>. The first HIV-1 cases in South Africa were reported in 1983 and the infections were confined to men who have sex with men (MSM) and bi-sexual risk groups<sup>3</sup>. The viral sequences between 1984 and 1990 were characterized as HIV-1 subtypes B and D<sup>4–6</sup>. The spread of HIV-1 subtype C in the heterosexual population in South Africa has been on the increase since it was first identified in the late 1980s<sup>7</sup>.

There is a total of 182 short fragment HIV-1 subtype B sequences from South Africa available in the Los Alamos HIV database (<http://www.hiv.lanl.gov>), accessed in September 2018. The majority of these short fragment sequences belong to the envelope (*env*) coding region and the *gag-pol* region. Out of these, only five subtype B near full-length genome sequences from South Africa have been published: R84 (FJ647145)<sup>8</sup>, TV016 (KJ948656)<sup>9</sup>, TV047 (KJ948657)<sup>9</sup>, TV1057 (KJ948660)<sup>9</sup>, and 03ZAPS045MB2 (DQ396398)<sup>10</sup>. Four of these sequences were characterized in our laboratory. Patient R84 was a MSM sampled in 1985<sup>8</sup>, TV016 was a heterosexual male sampled in 1988, TV047 was a heterosexual male sampled in 2000 and TV1057 was an MSM sampled in 2002<sup>9</sup>. Wilkinson *et al.*<sup>9</sup> described these sequences, while Rousseau *et al.*<sup>10</sup> described patient sequence 03ZAPS045MB2, a female sampled in 2003. Thus, there are limited information available for HIV-1 subtype B near full-length sequences from the early epidemic in South Africa. In Africa, there is only one published near full-length HIV-1 subtype B sequence OYI 397 from Gabon, described by Huet *et al.*<sup>11</sup>. Patient sequence OYI\_397 was sampled in 1988 and the risk factor, gender and patient ethnicity were not recorded<sup>11</sup>.

<sup>1</sup>Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, 8000, South Africa. <sup>2</sup>National Health Laboratory Services (NHLS), Western Cape Region, Tygerberg Hospital, Cape Town, South Africa. Correspondence and requests for materials should be addressed to G.B.J. (email: [graeme@sun.ac.za](mailto:graeme@sun.ac.za))

Patient	Age	Risk factor	Sample date	Clinical symptoms	Viral phenotype	RT activity (cpm/ml)
R68	24	MSM	1985	Symptomatic	SI	569940
R459	34	MSM	1987	Symptomatic	SI	1607500
R526	32	Bisexual	1987	Symptomatic	SI	36227
R605	29	Bisexual	1985*	Asymptomatic	Uncultured	Uncultured
R1296	48	Bisexual	1986	AIDS	NSI	951000

**Table 1.** Demographics and clinical data of patients. \*Blood sample dated 1985, **MSM**: Men who have sex with men; **SI**: Syncytium Inducing; **NSI**: Non-syncytium Inducing; **RT**: Reverse Transcriptase.

The objective of this study was to characterize HIV-1 subtype B near full-length virus sequences from the early South African epidemic from 1985–1987. This study forms the basis for continued research in our attempt to reconstruct the epidemiology and evolutionary history of HIV-1 in South Africa.

## Results

**Patient demographics and virus isolation.** The demographic and clinical data of the patients are summarized in Table 1.

Virus isolation was identified by cytopathic effect in the co-cultures for both patients R68 and R459. The cytopathic effect consisted of large multinucleated giant cells with clear balloon-like cytoplasm for R68 and smaller slowly developing syncytia for R459. The co-cultures for R526 and R1296 were non-cytopathic and virus isolation was confirmed by the Reverse Transcriptase (RT) assay. The RT counts ranged from 36227 to 1607500 counts per minute per millilitre (cpm/mL). We received a blood sample from R605 and no virus isolation was done.

**PCR amplification and near full-length genomic sequences.** All PCR amplicons were positive for both fragments 1 and 2. Assembly of the overlapping sequence contigs resulted in the following sequences, relative to HXB2: R68 [832–9183 base pairs (bp)], R459 [849–9222 bp], R526 [814–9222 bp], R605 [850–9186 bp] and R1296 [1042–9219 bp]. The sequences resulted in the characterization of near full-length HIV-1 genomes spanning from the 5' *gag* region through the 3' U3 region. Open reading frames were identified for *gag*, *pol*, and *env* structural genes and for *vif*, *vpr*, *vpu*, *nef*, *tat*, and *rev* regulatory and accessory genes.

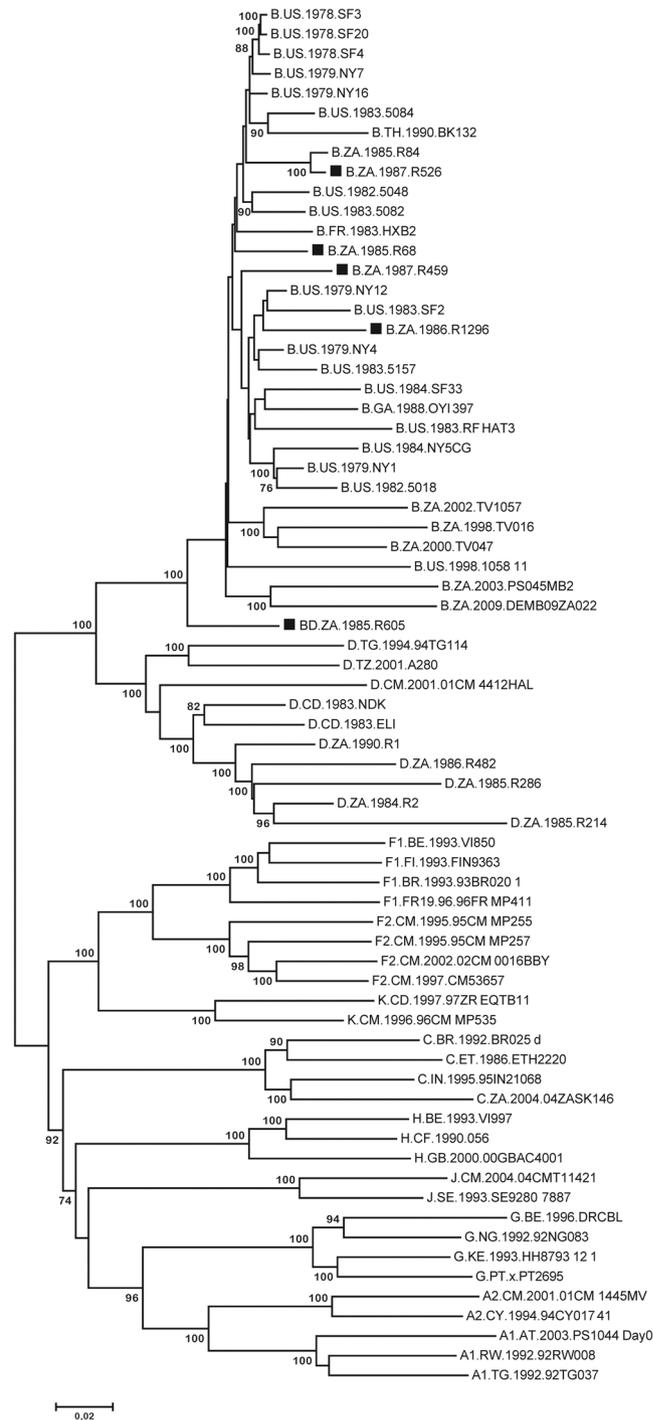
**Maximum Likelihood phylogenetic tree inference.** Maximum Likelihood (ML) tree topologies were inferred for the near full-length sequences [Fig. 1]. The best-fitting nucleotide substitution model was evaluated using MEGA version 6. Four out of five patient sequences clustered within subtype B reference sequences, while sequence R605 is seen as a HIV-1 subtype B outlier. Sequence R526 clustered with sequence R84 from South Africa with high bootstrap support. Sequence R68 and R459 clustered with strains from the USA. Sequence R1296 clustered with strains from San Francisco and New York.

**Recombinant analyses.** A representation of the unique recombinant form between subtypes B and D of R605 is illustrated in Fig. 2. The following breakpoints were observed from the jpHMM analyses: *gag* – *pol* (862–3918 bp HXB2) subtype B, *pol* – *vif* (3918–5974 bp HXB2) subtype D and *vpu* – *env* (5974–9187 bp HXB2) subtype B. In order to validate the recombinant breakpoints, ML phylogenetic trees were inferred for each fragment.

## Discussion

In 1982 the first two HIV-1 patients were reported in South Africa in MSM patients. Both had contacts in the USA before developing AIDS<sup>3</sup>. In South Africa, two independent epidemics have been described<sup>7</sup>. The first was initially spread by MSM. This was later identified as HIV-1 subtype B and D<sup>4–6</sup>. The second epidemic started to expand in the heterosexual population as far back as early 1988, with seroprevalence data in blood donor populations from different race groups that showed a characteristic pattern. These independent epidemics are consistent respectively to global epidemiological patterns<sup>12</sup>. In addition, human migration is a major process that shaped the origin and dissemination of HIV-1<sup>13</sup>. Data generated by Wilkinson *et al.*<sup>14</sup> indicate that in South Africa the bulk of viral introductions occurred during a period of socio-political change in the country (1980–2000)<sup>14</sup>. Based on our current knowledge South Africa experienced multiple introductions of HIV-1 from neighbouring countries towards the move to full independence in the 1990s, after the end of Rhodesian bush war<sup>15</sup>. The results generated from this study further strengthen the hypothesis of the presence of HIV-1 subtype B and HIV-1 subtype D amongst the MSM risk group at the beginning of the epidemic in the country. This may also explain the HIV-1 subtype BD URF detected and sequenced from the PBMC of patient R605.

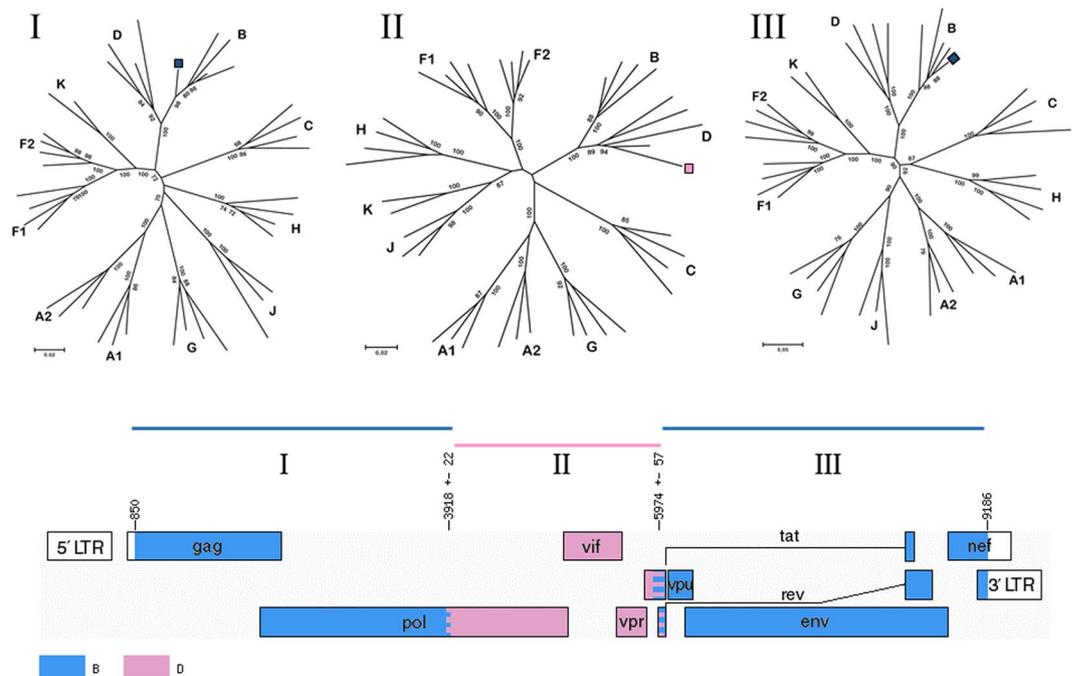
In 2014, an unusual high HIV-1 subtype diversity was described in Cape Town<sup>16</sup>. This suggests that international tourism and MSM migration into the country, seeking constitution protection, might play a vital role to the MSM network in Cape Town<sup>16</sup>. Furthermore, there has been an increase in the number of subtype B cases spread through heterosexual and mother-to-child transmission. In this study, five new near full-length HIV-1 sequences were characterized: four subtype B's (R68, R459, R526 and R1296) and one unique subtype B and D recombinant form (R605). The near full-length subtype B sequences included in this study were isolated from both MSM and heterosexual individuals. Patient R84 sampled in 1985 (GenBank FJ647145) is the only subtype B near full-length sequence which has been characterized from the early epidemic in South Africa<sup>8</sup>. This subtype B isolate, along with 5 subtype D viruses, were characterized in three separate studies, represents the only near full-length sequences of the early MSM HIV-1 epidemic within South Africa<sup>4–6</sup>. Three subtype B sequences have been described from the later epidemic: TV016 sampled in 1998 (GenBank KJ948656)<sup>9</sup>, TV047 sampled in 2000



**Figure 1.** HIV-1 near full-length genome (NFLG) molecular phylogenetic analysis inferred by the maximum likelihood (ML) method. The ML phylogenetic tree inferred in MEGA version 6 contains the five new HIV-1 sequences (■), HIV-1 subtype B and D reference sequences, previously characterized sequences from South Africa and the HIV-1 Group M reference subtypes. The evolutionary distances were inferred using the general time reverse (GTR) model of nucleic acid substitution with an estimated Gamma shape parameter and invariant sites. The genetic distance is displayed in the scale bar at the bottom of the figure. Bootstrap values higher than 70% are indicated.

(GenBank KJ948657)<sup>9</sup> and TV1057 sampled in 2002 (GenBank KJ48660)<sup>9</sup>. In Gabon, OYI\_397 sampled in 1988 represents the only other HIV-1 subtype B near full-length sequence published from Africa<sup>11</sup>. Although the infection occurred in the 1980s, analyses of early sequences are critical to understanding the time of origin of the early epidemic in the country<sup>14</sup>.

In the 1980s a minor subtype D epidemic was also present in South Africa<sup>4,5</sup>. According to the LANL HIV database (accessed 04 June 2018), there are 83 near full-length pure HIV-1 subtype D sequences described in the



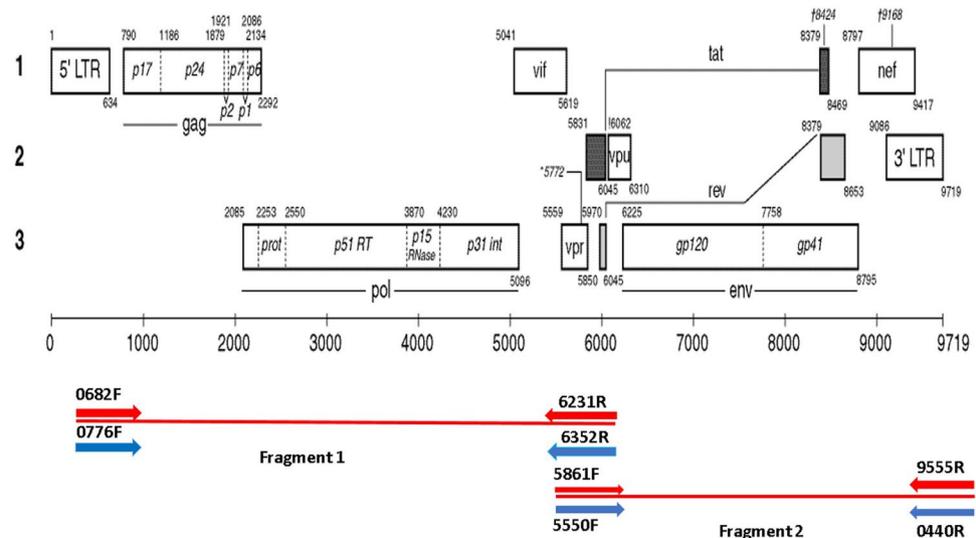
**Figure 2.** Subtyping analyses of the unique recombinant form (URF) of R605, a BD recombinant sequence according to jumping profile Hidden Markov Model (jpHMM), LANL and phylogenetic analyses. Phylogenetic analyses (ML tree) according to breakpoint located at position (862–3918 bp) in the *gag*–*pol* region clusters amongst subtype B sequences. ML tree according to breakpoint position (3918–5974 bp) in the *pol*–*vif* region clusters amongst subtype D sequences. ML tree according to breakpoint position (5974–9187 bp) in the *vpu*–*env* region clusters amongst subtype B sequences. In order to validate the sequences a total of 100 bootstrap replicates was carried to infer a ML phylogenetic tree.

world and only 10 (12.04%) of these sequences were not from Africa. Loxton *et al.*<sup>4</sup>, and Jacobs *et al.*<sup>5</sup>, described five HIV-1 subtype D viruses from South Africa. Limited focus has been placed on this HIV-1 subtype D in this country. In 1997, HIV-1 subtype D was identified in a male MSM and one heterosexual patient through partial *gag* analysis<sup>17</sup>. Bredell *et al.*<sup>18</sup> conducted a study where they identified a HIV-1 subtype C and D recombinant in a heterosexual through partial sequencing of *gag* and *env* gene. Furthermore, in the same study, they also identified HIV-1 subtype D in a male heterosexual through partial sequencing of the *gag* gene. There are currently 83 near full-length HIV-1 subtype D sequences in the Los Alamos Database, of which only ten were not from Africa. Before 1990 there was eight near-full-length subtype D sequences described. Four of these sequences, R2 (1984), R214 (1985), R286 (1985) and R482 (1986), were from South Africa<sup>4</sup>. The other four sequences were from the DRC: NDK (1983)<sup>19</sup>, ELI (1983)<sup>20</sup>, 84ZR085 (1984)<sup>21</sup> and CDC Z2 (1985)<sup>22</sup>.

In this study, we identified an HIV-1 subtype BD Unique Recombinant Form (URF) from a patient sample obtained in 1985 during the start of the epidemic in South Africa. To the best of our knowledge, R605 is the first URF<sub>BD</sub> described in the world. In South Africa, previous studies have identified the presence of URFs. In 2014, Jacobs and co-workers described the emergence of a possible URF of HIV-1 subtype B and C in South Africa<sup>23</sup>. In addition, the presence of recombinants AD (TV101) and AC (TV218) were also described. TV101 is the second AD recombinant described in 2015 and is closely related to AF457082 from Kenya<sup>24</sup>. The first AD recombinant characterized were from a South African patient that was infected via heterosexual contact in Kenya. Recombinant AC (TV218) was sampled from a 25-year-old female in Durban<sup>9</sup>. Two AC recombinant forms were also described by Rousseau *et al.*<sup>10</sup> and Papathanasopoulos *et al.*<sup>25</sup>. The breakpoints (recombination events) in all four of the AC recombinant fragments differ and thus each one represents URFs. In addition, previous studies have described the presence of unique inter-subtype recombinants such as BC, BF, and AC in both the MSM and heterosexual populations<sup>16,23,25</sup>.

## Conclusion

With the advancement of molecular techniques and the availability of stored samples, we managed to characterize an HIV-1 recombinant sequence, sampled in 1985 in South Africa. Phylogenetic inference of five newly sequenced HIV-1 strains, identified subtypes B and the novel BD recombinant. Characterization of HIV-1 sequences during the early years were often based on short sub genomic regions. This most likely limited the identification of recombinant viral strains. Only near full-length genome analyses of HIV-1 can enable us to make accurate assumptions on viral recombination and pure subtypes. It is necessary to continue monitoring the evolution and spread of HIV-1 as understanding HIV-1 diversity can help us understand specific transmission patterns and help us understand how the virus continues to spread in South Africa and worldwide.



**Figure 3.** The amplification strategy. The diagram shows amplification for whole genome. Full genome amplification was achieved by amplifying DNA in two fragments. Four primers were used for each fragment utilising pre-nested and nested PCR. Fragment 1 consisted of genes starting from *gag* to *vpu* and was approximately 5.5 Kb in length. The second fragment of genes starting from *vpr* until the 3'/LTR and was approximately 3.7 kb in length.

## Methods

**Ethics statement.** This study was approved by the Health Research Ethics Committee of Stellenbosch University, South Africa (N15/08/071). The study was conducted according to the ethical guidelines and principles of the international Declaration of Helsinki 2013, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research. A waiver of consent was awarded to conduct sequence analyses on these archived samples.

**Study design.** HIV-1 strains from infected patients were isolated routinely at the Division of Medical Virology, Tygerberg Academic Hospital in the Western Cape Province, between 1984 and 1992. Patient samples [plasma, serum and peripheral blood mononuclear cells (PBMCs)] were stored at  $-20^{\circ}\text{C}$ . In this study, we selected to sequence the NFLG of five previously identified HIV-1 subtype B samples from the start of the epidemic in South Africa.

**PBMC cultures.** PBMC cultures were performed as previously described<sup>26</sup>. Briefly, lymphocytes were separated from the buffy layer on a ficoll gradient and cultured at  $37^{\circ}\text{C}$  in an atmosphere of 5%  $\text{CO}_2$ . The RPMI-1640 medium was supplemented with 20% fetal calf serum, 0.2%  $\text{Na}(\text{HCO}_3)_2$ , 2% L-glutamine, antibiotics, and 10–20% Interleukin 2. These PBMCs were co-cultured using umbilical cord lymphocytes that were stimulated with Phytohaemagglutinin for 3 days and treated with Polybrene for 30 minutes before co-cultivation. The medium was replaced every 2 to 3 days and the cell cultures were examined daily for cytopathic effect (CPE).

**Reverse transcriptase assay.** The reverse transcriptase (RT) assays were performed as previously described<sup>27</sup>. Briefly, RT activities were assayed in cell-free supernatants of control and infected PBMC cultures. Aliquots of 5 to 20 ml of supernatant fluid were clarified by centrifugation at 10 000 g for 10 minutes, thereafter pelleted at 100 000 g for 50 minutes. The pellets were re-suspended in 60  $\mu\text{l}$  of lysis buffer (10 Mm Tris, 0.2% Triton X-100, 1 mM EDTA, 5 mM dithiothreitol, and 60 mM KCl, pH 7.3). The assays were performed in duplicate on 30  $\mu\text{l}$  amounts of pellet lysate. The template used was 0.01 mg/ml poly(rA) oligo(dT) (Pharmacia, Sweden) in 5 mM  $\text{MgCl}_2$ , 2.4 mM Dithiothreitol, 60 mM KCL and 0.0034 mM methyl [H]thymidine-5' -triphosphate (48 Ci/mMole) (Amersham, England).

**Nucleic acid extraction.** High molecular weight DNA was extracted from the HIV-1 positive cultures and PBMCs using conventional phenol-chloroform extraction methods and stored at  $4^{\circ}\text{C}$ .

**PCR amplification.** The NFLG amplification protocol was adapted from Grossman *et al.*<sup>28</sup>. Briefly, two overlapping fragments were amplified using the high fidelity KAPA HiFi HotStart ReadyMix (2X) (KAPA Biosystems, USA). The first fragment (F1) ranges from *gag* to *vpu*, position 0776–6231, approximately 5.5 kilobases [kb] in length, relative to HXB2 (Fig. 3). First round PCR was performed with the primer pair 0682F (5'-TCTCTCGACGCAGGACTCGGCTTGCTG-3') and 6352R (5'-GGTACCCATAATAGACTGTRACCCACAA-3'), followed by second round nested primer pair with 0776F (5'-CTAGAAGGAGAGAGAGATGGGTGCGAG-3') and 6231R (5'-CTCTCATTGCCACTGTCTTCTGCTC-3')<sup>28</sup>. The PCR cycling conditions were as follows: Initial denaturation of  $95^{\circ}\text{C}$  for 5 minutes, followed by 30 cycles of  $98^{\circ}\text{C}$  for 20 seconds,  $65^{\circ}\text{C}$  for 15 seconds and  $72^{\circ}\text{C}$  for 3 minutes and a final extension at  $72^{\circ}\text{C}$  for 5 minutes. The second

fragment (F2) starts from the *vif* to the 3′LTR, 5861 to 9555, approximately 3.7 kb in length, relative to HXB2. The F2 PCR was performed with first round primers 5550F (5′-AGARGAYAGATGGAACAAGCCCCAG-3′) and 9555R (5′-TCTACCTAGAGAGACCCAGTACA-3′), followed by second round nested PCR with primers 5831F (5′-TGGAAGCATCCRGGAAGTCAGCCT-3′) and 0440R (5′-CCAGAGCTCACCTAGCACCA TCCAAAGGTCAGTGGG-3′). The PCR cycling conditions were as follows: Initial denaturation of 95 °C for 5 minutes, followed by 30 cycles of 98 °C for 20 seconds, 65 °C for 15 seconds and 72 °C for 2 minutes and final extension at 72 °C for 5 minutes. Each PCR reaction contained 25 µl of KAPA HiFi HotStart ReadyMix (2X), 1.5 µl of each primer (10 µM), 5 µl of template DNA at a concentration of 20 ng/µl and 17 µl of nuclease-free water to a final volume of 50 µl.

**PCR purification.** PCR products were purified using the direct QIAquick PCR purification kit. The QIAquick gel extraction kit was used where multiple bands were visible on the agarose gel (Qiagen, Germany). Purification steps were performed according to the manufacturer's instructions.

**Sequencing.** Conventional Sanger DNA sequencing reactions were performed with the ABI Prism Big Dye Terminator Cycle sequencing kit version 3.1 and run on the ABI 3130xl automated DNA sequencer (Applied Biosystems, USA). The sequencing reactions were performed according to the manufacturer's instructions. Sequence chromatograms were assembled into contiguous fragments and confirmed in Sequencher version 5.0, <https://www.genecodes.com/> (Gene Codes Corporation, Ann Arbor, MI, USA). Sequencher® version 5.4.6 DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA <http://www.genecodes.com>.

**Sequence quality control.** The nucleotide sequences were verified for stop codons, insertion and deletions using an online quality control program on the HIV LANL database (<https://www.hiv.lanl.gov/content/sequence/QC/index.htm>). Each patient's sequences were named according to the 2-letter country code.year.name, for example ZA.85.R1296. HIV BLAST was used to find sequences similar to our sequences [https://www.hiv.lanl.gov/content/sequence/BASIC\\_BLAST/basic\\_blast.html](https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html). Thereafter, patient's HIV sequences were preliminary characterized and subtyped using the jumping profile Hidden Markov Model (jpHMM) (<http://jphmm.gobics.de/>)<sup>29</sup> and Recombinant Identification Program (RIP) version 3.0 <https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html><sup>30</sup>.

**Maximum likelihood (ML) phylogenetic tree inference.** The HIV-1 subtype reference sequence dataset from 2010 were acquired from e LANL <https://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>. Additional subtype B sequences from 1978–1985 were included in our analyses, as the BLAST results indicated close similarity to our new subtype B sequences. All near-complete genome sequences for subtype B and D from South Africa were also included.

Multiple sequence alignments were done with MAFFT version 7.388<sup>31,32</sup>, as implemented in Geneious version 11.1.5 (<https://www.geneious.com>). This multiple alignment was further refined by using the HMM option in HIVAlign (<https://www.hiv.lanl.gov/content/sequence/VIRALIGN/viralalign.html>) and manually checked. The best fitting evolutionary model of nucleotide substitution was estimated in Mega version 6.06<sup>33</sup>, using 24 different nucleotide substitution models. The model with the lowest Bayesian Information Criterion (BIC) score was used for further analysis. Subsequently the Maximum Likelihood tree was inferred with the general time reversible (GTR) model of nucleotide substitution<sup>34</sup>, using a discrete Gamma distribution (+G) with five rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+I)<sup>35</sup>. To test the reliability of the inferred ML tree, we used bootstrap analysis with a total of 100 replicates.

**Recombination analyses.** Based on the breakpoints identified with the jpHMM online tool, ML tree topologies were inferred for each of the recombinant fragments of R605. Each fragment was aligned with the HIV-1 subtype reference alignment in MAFFT 7.388<sup>33,34</sup>, as implemented in Geneious 11.1.5 (<https://www.geneious.com>). ML phylogenetic trees were inferred with MEGA 6.06 as described before.

**GenBank.** The sequences analysed from the study have been submitted to Genbank and are available under the following accession number: MH234639, MH234640, MH234641, MH234642 and MH234643.

## References

- UNAIDS Fact sheet – Latest statistics on the status of the AIDS epidemic, <http://www.unaids.org/en/resources/fact-sheet> (July 2018).
- UNAIDS data, <http://www.unaids.org/en/resources/documents/2018/unaids-data-2018> (2018).
- Ras, G. J., Simson, I. W., Anderson, R., Prozesky, O. W. & Hammersma, T. Acquired immune deficiency syndrome: a report of two South African cases. *South African Medical Journal* **64**, 140–142 (1983).
- Loxton, A. G., Treurnicht, F., Laten, A., van Rensburg, E. J. & Engelbrecht, S. Sequence analysis of near full-length HIV type 1 subtype D primary strains isolated in Cape Town, South Africa, from 1984 to 1986. *AIDS Research and Human Retroviruses* **21**, 410–413 (2005).
- Jacobs, G. B., Loxton, A. G., Laten, A. & Engelbrecht, S. Complete Genome Sequencing of a Nonsyncytium-Inducing HIV Type 1 Subtype D Strain from Cape Town, South Africa. *AIDS Research and Human Retroviruses* **23** (2007).
- Engelbrecht, S., Laten, J. D., Smith, T. L. & van Rensburg, E. J. Identification of env subtypes in fourteen HIV type 1 isolates from South Africa. *AIDS Research and Human Retroviruses* **11**, 1269–1271 (1995).
- Van Harmelen, J. *et al.* An association between HIV-1 subtypes and mode of transmission in Cape Town, South Africa. *AIDS* **11**, 81–87 (1997).
- Wilkinson, E. & Engelbrecht, S. Molecular characterization of non-subtype C and recombinant HIV-1 viruses from Cape Town, South Africa. *Infect. Genet. Evol.* **9**, 840–846 (2009).
- Wilkinson, E. *et al.* 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. *AIDS Research and Human Retroviruses* **31**, 412–420 (2015).

10. Rousseau, C. M. *et al.* Large-scale amplification, cloning and sequencing of near full-length HIV-1 subtype C genomes. *J. Virol. Methods* **136**(1–2), 118–125 (2006).
11. Huet, T., Dazza, M. C., Brun-Vézinet, F., Roelants, G. E. & Wain-Hobson, S. A highly defective HIV-1 strain isolated from a healthy Gabonese individual presenting an atypical Western blot. *AIDS* **3**(11), 707–715 (1989).
12. Schoub, B. D. *et al.* Considerations on the further expansion of the AIDS epidemic in South Africa - 1999. *South African Med. Journal AMJ* **77**, 613–618 (1990).
13. Junqueira, D. M. & Almeida, S. E. HIV-1 subtype B: Traces of a pandemic. *Virology* **495**, 173–184 (2016).
14. Wilkinson, E. *et al.* Origin, imports and exports of HIV-1 subtype C in South Africa: A historical perspective. *Infect Genet. Evol. MEEGID* **46**, 200–208 (2016).
15. Dalai, S. C. *et al.* Evolution and molecular epidemiology of subtype C HIV-1 in Zimbabwe. *AIDS* **18**, 2523–2532 (2009).
16. Middelkoop, K. *et al.* Epidemiology of HIV-1 Subtypes Among Men Who Have Sex With Men in Cape Town, South Africa. *J. Acquir. Immune Defic. Syndr.* **65**, 473–480 (2014).
17. Bredell, H., Williamson, C., Sonnenberg, P., Martin, D. J. & Morris, L. Genetic characterization of HIV type 1 from migrant workers in three South African gold mines. *AIDS Research and Human Retroviruses* **14**(8), 677–684 (1998).
18. Bredell, H. *et al.* HIV-1 Subtype A, D, G, AG and unclassified sequences identified in South Africa. *AIDS Research and Human Retroviruses* **18**, 681–683 (2002).
19. Spire, B. *et al.* Nucleotide sequence of HIV1-NDK: a highly cytopathic strain of the human immunodeficiency virus. *Gene* **81**(2), 275–284 (1989).
20. Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* **46**(1), 63–74 (1986).
21. Gao, F. *et al.* A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. *J. Virol.* **72**(7), 5680–5698 (1998).
22. Srinivasan, A. *et al.* Molecular characterization of human immunodeficiency virus from Zaire: nucleotide sequence analysis identifies conserved and variable domains in the envelope gene. *Gene* **52**(1), 71–82 (1987).
23. Jacobs, G. B. *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* **9** (2014).
24. Dowling, W. E. *et al.* Forty-one near full-length HIV-1 sequences from Kenya reveal an epidemic of subtype A and A-containing recombinants. *AIDS* **16**(13), 1809–1820 (2002).
25. Papanthanasopoulos, M. A. *et al.* Full-length genome analysis of HIV-1 subtype C utilizing CXCR4 and intersubtype recombinants isolated in South Africa. *AIDS Research and Human Retroviruses* **18**(12), 879–886 (2002).
26. Spracklen, F. H. *et al.* The acquired immune deficiency syndrome and related complex. A report of 2 confirmed cases in Cape Town with comments on human T-cell lymphotropic virus type III infections. *South African Medical Journal* **68**, 139–143 (1985).
27. Rübbsamen-Waigmann, H. *et al.* Isolation of variants of lymphocytotropic retroviruses from the peripheral blood and cerebrospinal fluid of patients with ARC or AIDS. *J. Med. Virol.* **19**(4), 335–344 (1986).
28. Grossmann, S., Nowak, P. & Neogi, U. Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management. *J. Int. AIDS Soc* **18**, 1–8 (2015).
29. Schultz, A. K. *et al.* jpHMM: Improving the reliability of recombination prediction in HIV-1. *Nucleic Acids Research* **37**, W647–651 (2009).
30. Siepel, A. C., Halpern, A. L., Macken, C. & Korber, B. T. A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Research and Human Retroviruses* **11**(11), 1413–1416 (1995).
31. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**(4), 772–780 (2013).
32. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**(14), 3059–3066 (2002).
33. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725–2729 (2013).
34. Tavaré, S. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *Lectures on Mathematics in the Life Sciences*. **17**, 57–86 (1986).
35. Yang, Z. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10**, 1396–1401 (1993).

## Acknowledgements

Grant support for this study were received from the National Research Foundation (NRF), Poliomyelitis Research Foundation (PRF) of South Africa, National Health Laboratory Service (NHLS) Research Trust Grant Ref-004\_94657, and the South African Medical Research Council (SAMRC) TB HIV Collaborating Centre for Laboratory Research (TygHIVLab), Ref-S004116.

## Author Contributions

G.B.J. and S.E. conceptualized the study. S.E. performed PBMC cultures, R.T. assays and DNA isolation. A.E.O. performed the PCR amplification and sequence analyses and wrote the first draft. All authors performed data analysis, edited, read and approved the final manuscript.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019