

SCIENTIFIC REPORTS



OPEN

Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and available of integrase inhibitors in Cape Town, South Africa

Dominik Brado^{1,2}, Adetayo Emmanuel Obasa^{2,3}, George Mondinde Ikomey⁴, Ruben Cloete⁵, Kamalendra Singh^{3,6,7}, Susan Engelbrecht², Ujjwal Neogi³  & Graeme Brendon Jacobs²

HIV-Integrase (IN) has proven to be a viable target for highly specific HIV-1 therapy. We aimed to characterize the HIV-1 IN gene in a South African context and identify resistance-associated mutations (RAMs) against available first and second generation Integrase strand-transfer inhibitors (INSTIs). We performed genetic analyses on 91 treatment-naïve HIV-1 infected patients, as well as 314 treatment-naïve South African HIV-1 IN-sequences, downloaded from Los Alamos HIV Sequence Database. Genotypic analyses revealed the absence of major RAMs in the cohort collected before the broad availability of combination antiretroviral therapy (cART) and INSTI in South Africa, however, occurred at a rate of 2.85% (9/314) in database derived sequences. RAMs were present at IN-positions 66, 92, 143, 147 and 148, all of which may confer resistance to Raltegravir (RAL) and Elvitegravir (EVG), but are unlikely to affect second-generation Dolutegravir (DTG), except mutations in the Q148 pathway. Furthermore, protein modeling showed, naturally occurring polymorphisms impact the stability of the intasome-complex and therefore may contribute to an overall potency against INSTIs. Our data suggest the prevalence of INSTI RAMs, against INSTIs, is low in South Africa, but natural polymorphisms and subtype-specific differences may influence the effect of individual treatment regimens.

Combination antiretroviral therapy (cART) has dramatically reduced HIV infection to a chronic and manageable disease resulting in a near-normal life expectancy^{1,2}. However, cART has also led to the development of resistance-associated mutations (RAMs) and transmitted drug resistances (TDRs), which are associated with a higher rate of virological failure³⁻⁵. With rising levels of drug resistance and first-line cART failure, more patients will require second-line and salvage cART, which shows treatment efficacy in terms of viral suppression in 60% of the cases, but may be up to three times more expensive than the first-line cART^{6,7}.

HIV-1 Integrase (IN) is responsible for the integration of the viral nucleic material into the host genomic DNA⁸. Integrase strand transfer inhibitors (INSTIs) are highly potent antiretroviral agents with durable efficacy, minimal toxicity and is internationally approved and used for both treatment-naïve and treatment-experienced patients⁹⁻¹¹.

¹Division of Virology, Institute for Virology and Immunobiology, Faculty of Medicine, University of Wuerzburg, 97080, Wuerzburg, Germany. ²Division of Medical Virology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, 7505, Cape Town, South Africa. ³Division of Clinical Microbiology, Department of Laboratory Medicine, Karolinska Institute, University of Stockholm, Stockholm, Sweden. ⁴CSCCD, Faculty of Medicine and Biomedical Sciences, University of Yaoundé, Yaoundé, Cameroon. ⁵South African Medical Research Council Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Western Cape, South Africa. ⁶Department of Molecular Microbiology and Immunology, Columbia, MO, 65211, USA. ⁷Christopher Bond Life Sciences Center, University of Missouri, Columbia, MO, 65211, USA. Dominik Brado and Adetayo Emmanuel Obasa contributed equally to the work. Correspondence and requests for materials should be addressed to A.E.O. (email: obasa@sun.ac.za)

There are currently three US-Food and Drug Administration (FDA)-approved INSTIs: Raltegravir (RAL), Elvitegravir (EVG) and Dolutegravir (DTG). Two newer INSTIs, bicittegravir (BIC) and cabotegravir (CAB), are presently under consideration¹². The use of higher genetic barrier drugs such as Dolutegravir (DTG) is crucial to the success of salvage therapy to mitigate the emergence of resistant variants¹³. In 2007, the first INSTI RAL was approved for the treatment of patients infected with HIV-1, followed by EVG in 2012. These first-generation INSTIs are highly effective in the treatment of HIV-1-infected patients, but have a low barrier to resistance, resulting in the rapid emergence of RAMs^{14,15}. DTG is a second-generation INSTI that was approved by the FDA in 2014¹⁶. It has a higher resistance barrier than that of RAL and EVG¹⁷. In the case of DTG, resistance is selected slowly *in vitro*, but has not emerged in studies of therapy-naïve patients until today^{18,19}. When compared to an EFV based first-line regimen, patients receiving DTG have shown to be superior regarding viral suppression rates and had stabilized CD4⁺ T-Cell counts²⁰. This is mainly attributed to better adherence and fewer discontinuation rates under treatment. The WHO, however, only names it as an alternative to the above-mentioned first-line regimen, as little research has been done on the use of DTG²¹.

Since its initiation in 2004, South Africa's national HIV treatment program has grown to become the biggest in the world, currently treating approximately 3.4 million people²². Being in concordance with the World Health Organisations (WHO) guidelines, the recommended first-line combination antiretroviral therapy (cART) in South Africa consists of a non-nucleoside reverse transcriptase inhibitor (NNRTI) backbone regimen of Efavirenz (EFV), combined with two nucleoside reverse transcriptase inhibitors (NRTIs), namely Lamivudine (3TC) and either Tenofovir Disoproxil Fumarate (TDF) for adults or Abacavir (ABC) for children, respectively. The recommended second-line cART consists of the nucleoside reverse transcriptase inhibitors (NRTIs) Zidovudine (AZT) and Lamivudine (3TC) and a Ritonavir-boosted (*r*) Protease Inhibitor (PI), usually Atazanavir²³.

In this study we aim to provide further information on the susceptibility and primary drug resistance mutations profile of INSTIs as well as to establish a protocol to screen for Integrase RAMs in an HIV-1 subtype C predominated setting in South Africa.

Results

Patient demographics. The patient demographics are summarized in Supplementary Table 1. We amplified and confirmed successful sequencing, containing all 288 amino acids of the IN gene, for 91 samples.

HIV-1 subtyping. Based on HIV-1 subtyping using online automated tools and phylogenetic analysis, 85 (92%) of the samples were identified as HIV-1 subtype C followed by five (5.4%) as HIV-1 subtype B (5.6%, TV122; TV431; TV404; TV420 and TV356) and one as HIV-1 subtype A1 strains (1.1%, TV412) (Fig. 1).

Resistance mutation analyses. Drug resistance analyses showed that no major INSTI RAMs were present in this study. One sample (TV367) carried the accessory drug mutation G140E, a non-polymorphic mutation, that has been selected *in vitro* before, but which alone does not seem to influence the susceptibility of the virus to INSTIs²⁴. Minor, polymorphic, mutations were present in 6/91 samples (6.6%), of which four samples contained the mutation L74I (TV122, TV128, TV173, TV405), one other sample contained the mutation L74M (TV366) and another the polymorphism S230N (TV364).

Of note is, that 55/91 (60.4%) samples carried the M50I polymorphism, all of which were classified as subtype C. M50I does not confer resistance to any of the currently available INSTIs and therefore is not listed as RAM in the Stanford University HIV Drug Resistance Database (<https://hivdb.stanford.edu/>). However, it has been selected *in vitro*, following a bicittegravir (BIC) resistance selection assay²⁵. In this M50I succeeded an R263K mutation, and only conferred low-level resistance to BIC (2.8-fold) in this combination. R263K was not present in our cohort.

Database derived IN sequence resistance analyses. After excluding multiple sequences from a patient to avoid overestimation of the variant calling and problematic sequences, we used 314 sequences collected between 1999 and 2007. These, we subsequently screened for the presence of RAMs, and identified 6.4% (20/314) sequences to contain RAMs with only 2.86% (9/314) having major INSTI resistance mutations: Q148H, T66S, E92G, S147G, T66A, Y143YF as well as Y143H. Q148H, T66S, E92G, Y143YF, and S147G were present in one sequence each (0.3%) whereas T66A and Y143H could be detected in two sequences respectively (0.6%). 3.5% (11/314) of the sequences contained four different IN accessory mutations, namely E157Q, T97A, G163, and S230R.

Generation of consensus South African HIV-1 subtype C sequence (HIV-1C_{ZA}). The consensus sequences generated using the database-derived HIV-1C_{ZA} sequences (n = 314) and cohort sequences (n = 85) identified 17 naturally occurring polymorphisms D25E, V31I, M50I, I72V, F100Y, L101I, T112V, T124A, T125A, K136Q, V201I, T218I, L234I, A265V, R269K, D278A, and S283G. (Fig. 2). Further profiling of sequences obtained through HIVseq Program, HIV-1B (n = 5278), HIV-1C (n = 1416), cohort sequences HIV-1C-ZA (n = 87) confirmed the findings. Among the 17 mutations 11 were further increased in our cohort.

Molecular modeling. The molecular models were created using the cryoEM structure of HIV-1B IN (Figure 3). The cryoEM structure has an active site mutation E152Q, in the modeled structure, this mutation was reverted to glutamate. Here, we focussed on our five naturally occurring polymorphisms: E25, I50, Y100, I101 and I201 in the modeled structure of HIV-1C_{ZA} IN (Figure A). Our model showed that I50 (M50I mutation) is in the proximity of two strands of substrate DNA from two different monomers and therefore appears important in stabilizing/binding with DNA substrate (Figure B). Residue E25 (D25E mutation) from one monomer forms an ion-pair with K188 of a different monomer in a symmetric fashion (Figure C). The two monomers directly bind DNA substrate

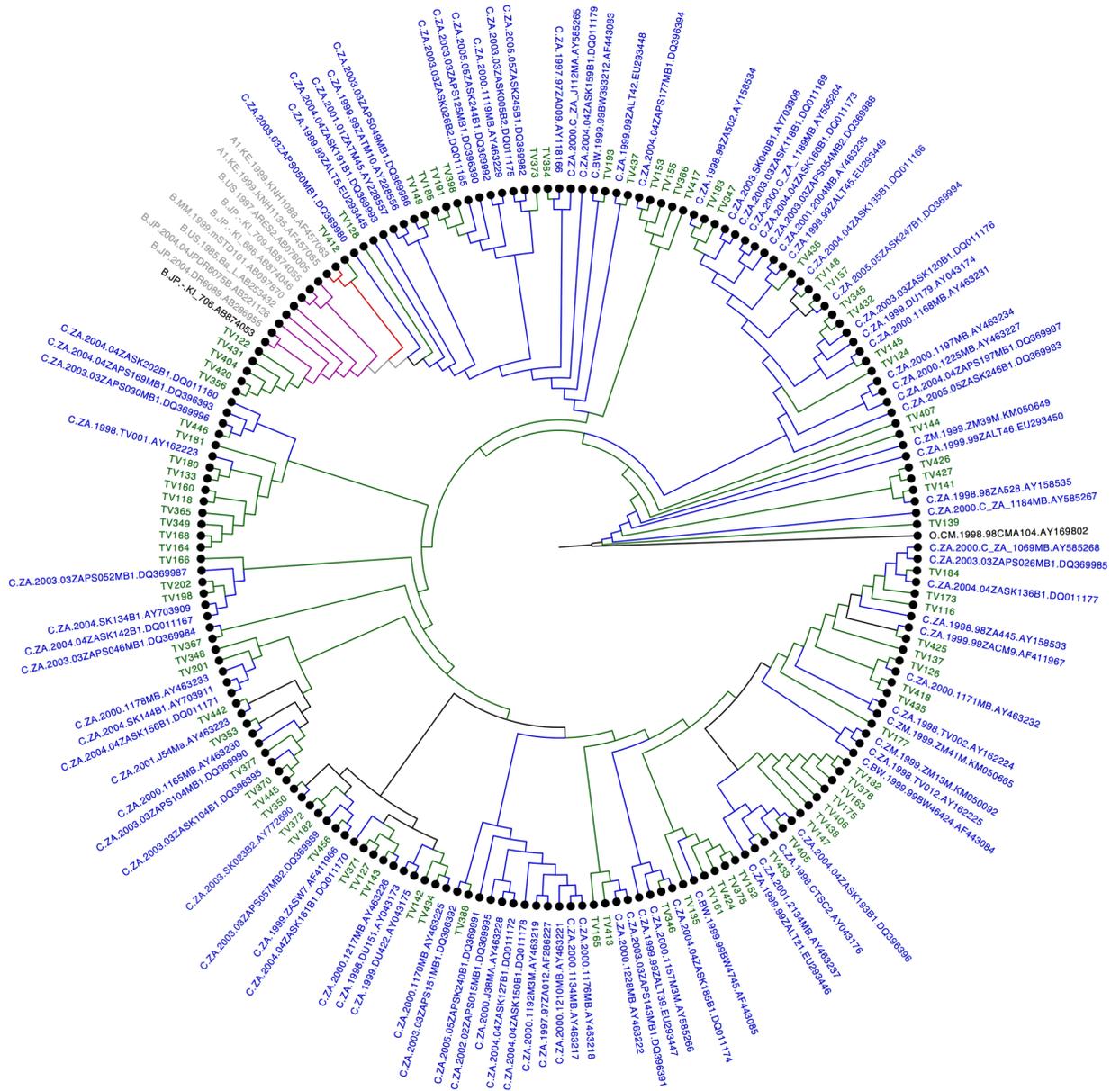


Figure 1. HIV-1 Integrase molecular phylogenetic analysis inferred by the maximum likelihood (ML) method. The ML phylogenetic tree inferred in RAXML contains 91 patient sequences and HIV-1 reference sequences dataset from 2010 were acquired from HIV-1 LANL database. The evolutionary distances were computed 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; while the majority of the sequences clusters with HIV-1 subtype C. 87 of the samples clustered with Subtype C reference strains (91.1%), five with Subtype B (5,6%) and one with Subtype A1 strains (1.1%).

and have their active site positioned close to DNA. This interaction is essential in maintaining the tetramer of IN. Mutations Y100 (F100Y) and I101 (L101I) are close to the active site of IN (D64, D116, and E152) in two monomers of which one subunit binds DNA through the active site, and the other one has an indirect interaction. Mutation I201 (V201I) is juxtaposed at the interface of two monomers of IN proteomer forming a hydrophobic interaction suggesting its importance in the maintenance of IN/DNA complex.

Discussion

InSTI containing regimens are considered a new and effective form of salvage therapy for cART-experienced patients failing first and/or second-line cART. South Africa, managing the most extensive HIV treatment program, has faced an increase in resistance rates against NNRTIs, NRTIs as well as PIs in the past years²⁶. With second-generation InSTIs not being readily available until recently and drug resistance rates in cART naïve patients in some cases exceeding 10%, these drugs could play an essential role in maintaining treatment options against multi-drug-resistant virus variants and preventing resistant viruses from further spreading^{5,27,28}

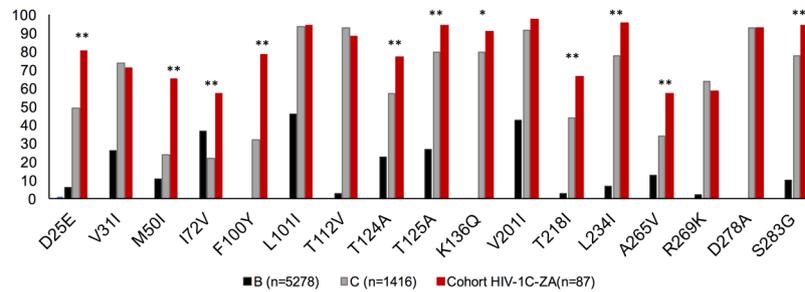


Figure 2. HIV-1C_{ZA} Integrase mutation profiling. Integrase mutation profiling of consensus sequences generated using the database-derived HIV-1C_{ZA} sequences HIV-1B (n = 5278), HIV-1C (n = 1416), cohort sequences HIV-1C-ZA (n = 87) identified 17 naturally occurring polymorphisms D25E, V31I, M50I, I72V, F100Y, L101I, T112V, T124A, T125A, K136Q, V201I, T218I, L234I, A265V, R269K, D278A and S283G. Among the 17 mutations 11 were further increased in our cohort

In this study we analyzed the IN region of HIV-1 infected, cART naïve patients, for the presence of InSTI treatment compromising polymorphisms and mutations. We showed that no primary major resistance mutations against InSTIs were circulating in our study cohort at the time of collection. One accessory mutation (G140E) was observed, while the highly polymorphic mutations L74I/M and S230N were present in 5,5% (5/91) and 1,1% (1/91), respectively. Neither of these mutations is associated with reduced susceptibility to InSTIs. L74I/M, a polymorphism, that has been described in both, cART naïve and RAL or EVG experienced patients before, does not diminish the effect of InSTIs by itself, but can contribute to a high-level resistance, only if co-occurring with major resistance mutations^{29,30}. S230N has been reported as a natural variant with a polymorphism rate ranging between 0,5% to 2,0%³¹. It has also been selected by RAL and/or EVG before, *in-vivo* and *in-vitro*, but does not seem to confer resistance to any of the available drugs^{31,32}.

These findings are in line with previous studies on cART naïve patients confirming the variability of the genomic IN region as well as its lack of major resistance mutations. In 2013 Bessong and Nwobegahay reported the absence of major RAMs in a study conducted in the north-eastern part of South Africa, and a year before, in 2012, Oliveira *et al.* analyzed HIV-1 positive samples from Mozambique for genetic diversity of the IN gene^{33,34}. While resistance-associated mutations were not present in this study, the L74M polymorphism was found in 3,4% of the cases. Similar results were observed in Brazil and Europe, before the widespread use of InSTIs^{35,36}.

Among the 9/314 (2.86%) major InSTI RAMs, present in the database-derived sequences, only Q148H, detected in 1/314 (0.3%), may profoundly affect second-generation InSTIs susceptibility. If co-occurring with additional RAMs, mutations in the Q148 pathway can lead to higher fold resistances against all InSTIs. Despite both first-generation InSTIs, RAL, as well as EVG, selecting for these mutations, they have not yet been described to emerge under initial second-generation InSTI treatment³⁷. Y143H (2/314, 0.6%) is usually selected by RAL, and is considered to be a transitional mutation as part of the Y143R resistance pathway. Alone Y143H does not influence the effect of InSTIs, but by further mutating to Y143R it may confer moderate to high-level resistance to RAL, but minimal if any resistance to DTG^{38,39}.

T66S, T66A, E92G and S147G, found in 0.3%, 0.6%, 0.3%, and 0.3%, respectively, are non-polymorphic mutations, normally selected by EVG treatment. They are associated with moderate to high-level resistance against EVG, although T66 mutations also bear cross-resistance to and are selected by RAL^{40,41}.

The most frequent IN accessory RAM within the online, retrieved sequences was T97A, being present in 1.6% (5/314), followed by E157Q in 0.96% (3/314), G163R in 0.6% (2/314) and S230R in 0.3% (1/314) of the cases. All of these mutations are found to be within their natural prevalence rates, and although they can confer low-level resistance to both, RAL and EVG, none of these mutations are known to reduce DTG susceptibility, neither *in vitro* nor *in vivo*^{41–44}.

Interestingly, one case report found single E157Q to be associated with treatment failure of a DTG containing regimen⁴⁵. Therefore, Anstett *et al.* investigated this association in 2016, but could not confirm the result⁴⁶. On the other hand, however, a recent study has also shown that eight patients, who had E157Q mutation and initiated with DTG-based therapy, did not suppress the viremia below detection level after six months of therapy⁴⁷. Hence, causality between E157Q and a reduced DTG susceptibility is debatable and needs further long-term follow-up studies.

Despite higher fold RAMs against InSTIs being absent in most treatment naïve settings, they can emerge under treatment, particularly with first generation InSTIs, as Rossouw *et al.* presented in their case report from May 2016. In this report, they describe the first South African patient to fail EFV based first-line consecutively, ritonavir-boosted Lopinavir backbone second-line, and RAL containing third-line therapy⁴⁸. Poor adherence to the therapy was reported throughout the patient's history, and a final drug-resistance test, performed three years after the initiation of third-line treatment, for the first time included InSTI-resistance testing. This test ultimately confirmed a high-level resistance against RAL and high- or intermediate-level resistance against three of the other four drugs.

Furthermore, this test also revealed cross-resistance to EVG and a low-level resistance against DTG. This cross-resistance to DTG is seldom observed in the only RAL exposed patients and therefore is worrying, especially because InSTI resistances develop significantly less frequently if initially treated with DTG, instead of RAL^{49,50}.

Nevertheless, this case study raises the concern of emerging InSTI resistance patterns in the South African context. Hence, proper drug resistance surveillance within South Africa will be required, in particular also because, a recent study identified subtype-specific differences in DTG cross-resistance patterns in patients failing RAL³⁹. Further, sequence and structure-based analyses showed that the subtype-specific effects were caused by polymorphic residues across subtypes, which significantly affected native protein activity, structure and function of importance for drug-mediated inhibition of enzyme activity⁵¹. Although DTG showed a high genetic barrier to resistance, subtype-specific differences have been observed in the selection of DTG resistance mutations.

Therefore, we analyzed the position of naturally polymorphic mutations in the context of their ability to impact the stability of intasome. The polymorphisms noted in our analyses appear to be essential for the stability of tetramer and/or binding of DNA substrate in catalytically competent mode. The topological positions of polymorphisms also suggest that the intasome complex stability may differ in different subtypes, which may alter the architecture of the complex and thereby affect InSTI-based therapy outcome.

As the analysed cohort was recruited before the initiation of the HIV treatment program in South Africa, the possibility of RAMs being transmitted by treatment-experienced individuals is highly unlikely. Therefore, we consider the described findings to be a true baseline InSTI resistance rate.

Our data suggest that the introduction of this class of ART drugs, especially second-generation InSTIs, into the national treatment program could help in managing the HIV epidemic in South Africa. However, the possible emergence of formerly described, as well as a subtype and setting specific resistance pathways, requires proper drug resistance surveillance in the future, in order to track the evolution of the virus in a subtype C predominated setting under the pressure of the new treatment.

Conclusion

In the absence of a cure for HIV, long-term cART outcomes need to be monitored efficiently for maximum efficiency. RAMs lead to therapy escape mutants, which can ultimately cause cART failure. We have shown that in the South African context InSTIs is potentially a viable option for salvage therapy. However, there is still a need to keep assessing the RAMs to ensure patients receive the best treatment and care possible.

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 analyses on the samples, as they were obtained between 2000 and 2001, stored since 2001.

Study design. The samples used for analyses ($n = 91$) were part of a previously described Tygerberg Virology (TV) cohort³². The cohort contains treatment naïve patient samples from multiple ethnic groups, as well as different sexual orientations. Patients were sampled between 2000 and 2001, before the initiation of South Africa's national HIV treatment program and the introduction of InSTIs.

Collection of patient samples. Whole blood was collected from patients with ethylenediaminetetraacetic acid (EDTA) tubes. Plasma was separated after centrifugation at 2000 rpm for 10 minutes at 4 °C. The samples were stored at -80 °C for long-term storage. We selected randomly from the stored samples for IN amplification.

Nucleic acid extraction. HIV-1 RNA extraction was performed using the QIAamp Viral RNA Mini Extraction Kit's Spin protocol according to the manufacturer's instructions (Qiagen, Germany). Briefly, 140 µl of plasma was used as a starting volume. Larger starting volumes of 280 µl of plasma was used for some samples with very low viral titers. Viral RNA was stored at -80 °C until use.

cDNA synthesis and PCR amplification. The synthesis of complementary DNA (cDNA) and first-round PCR amplification was performed using the Invitrogen SuperScript[®] III Reverse Transcriptase (RT) reagents (Invitrogen, Germany). HIV-1 Integrase specific primers used for amplification were Poli5 (5'-CACACAAAGGRATTGGAGGAAATG-3') and Poli8 (5'-TAGTGGGATGTGTACTTCTGAAC-3'), position 4162–4185 and 5195–5217 on the HXB2 reference strain, respectively, with an expected fragment size of 1056 base pairs⁵³. The 25 µl reaction volume contained 8.5 µl nuclease-free water 12.5 µl of 2 × reaction mix buffer, 0.5 µl of both forward and reverse primers at a concentration of 5nmol, 0.5 µl of RT/Platinum Taq mix and 2.5 µl of extracted RNA. Reverse transcription was performed at 50 °C for 20 minutes, followed by an initial denaturation at 94 °C for 2 minutes. Forty cycles of amplification were carried out at 94 °C for 15 seconds for denaturation, 30 seconds for primer annealing at 94 °C and 90 seconds for elongation at 68 °C. The final elongation step was done at 68 °C for 10 minutes. For second round PCR amplification, primers Poli7 (5'-ACAAGTAGATAAATAGTCAGT-3') and Poli6 (5'-ATACATATGRTGTTTTACTAARCT-3'), with an expected fragment size of 945 base pairs relative to position 4186–4209 and 5107–5130, were used, in combination with the GoTaq[®] Flexi DNA Polymerase Kit (Promega, USA)⁵³. The 50 µl reaction mix consisted of 28.5 µl nuclease-free water, 10 µl of 5 × reaction mix buffer, 0.5 µl of both primers at 20 nmol, 3.0 µl of MgCl₂ at 75 nmol and 3 µl of amplified DNA. Second round of amplification started with a denaturation at 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 20 seconds, annealing at 55 °C for 30 seconds and elongation at 72 °C for 90 seconds. The final elongation step was done at 72 °C for 10 minutes and amplicons were stored at 4 °C until further use. Positive PCR amplicons were purified from agarose gel according to the manufacturer instructions using the QIAquick PCR Purification Kit (Qiagen, Germany).

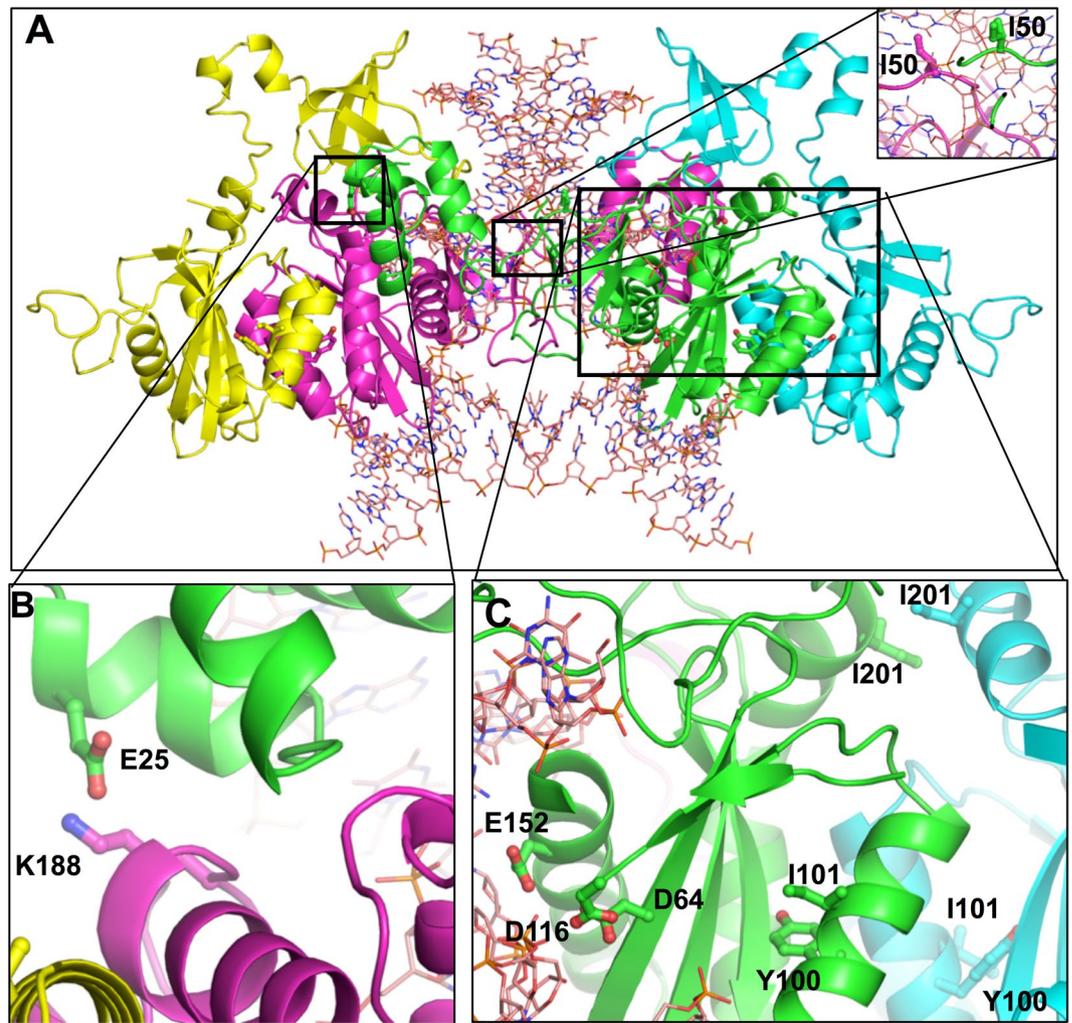


Figure 3. Homology derived molecular model of consensus HIV-1C_{ZA}. Homology derived a molecular model of Con_C_ZA. Figure A shows an intasome consisting of a tetramer of subtype C_ZA and substrate DNA. This structure was generated using the cryoEM structure of HIV-1B IN intasome (PDB file 5U1C) using ‘Prime’ of Schrodinger Suit using the protocol discussed in Neogi *et al.*, 2016. Inset in panel shows the proximity of I50 to DNA. Two I50 residues from two different subunit interact with DNA from two different sides. Figure B shows the position of E25 (in subunit colored green) that forms a ion-pair with K188 of subunit colored magenta. This is a symmetric interaction as E25 from magenta subunit interacts with K188 of green subunit. This interaction is important in maintaining the tetramer of IN. Figure C shows the active site residues D64, D116 and E152 of IN in one subunit (colored green) together with Y100 and I101 in the same and in the neighboring subunit. This figure also shows the position of I201 in two neighboring subunits. This interaction also appears critical for the maintenance tetramer organization of IN.

Sanger DNA Sequencing. All amplicons were sequenced on both strands with conventional Sanger DNA sequencing using the ABI Prism Big Dye[®] Terminator 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. Briefly, sequencing PCR cycling conditions were as follows: Initial denaturation of 95 °C for 60 seconds, followed by 25 cycles of 95 °C for 60 seconds, 55 °C for 7 seconds and 60 °C for 4 minutes. Sequencing primers consist of the above Poli6 and Poli7. Additional sequencing primers were used namely; Poli2 (TAAARACARYAGTACWAATGGCA), relative to position 4745–4766 and KLVO83 (GAATACTGCCATTTGTACTGCTG), corresponding to position 4750–4772⁵⁴. After that, sequences were assembled into contiguous fragments following (Phred quality score > 20) and edited manually using Sequencer version 5.0 (Gene Codes Corporation, USA). The bases were considered ambiguous if any nucleotide was present > 25% of the major peak.

HIV-1 Subtyping and phylogenetic analyses with online programs. HIV-1 subtyping based on integrase sequences was carried out using REGA v3 and COMET-HIV, followed by maximum likelihood phylogenetic analysis⁵⁵. The best fitted general time reverse (GTR) model of nucleic acid substitution with an estimated Gamma shape parameter and invariant sites model using Randomized Accelerated Maximum Likelihood (RAxML) as described previously^{56,57}.

Additional sequences. To compare our sequences with the rest of the IN sequences from South Africa, we performed a search on the LANL HIV database (<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.comp>). Our search inclusion criteria included all South African IN sequences and those identified from treatment naïve patients. We selected one sequence per patient and all problematic sequences were excluded from further analyses. Finally, 314 HIV-1 subtype C (HIV-1C) sequences were included in the analyses. Both cohort and database derived South African IN sequences were used to generate the consensus HIV-1C_{ZA} sequence using the Consensus Maker tool available in HIV-1 Los Alamos database using majority value 0.5 (<https://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html>). HIVseq Program, a literature prevalence of mutations in submitted sequences were used to identify the prevalence of naturally occurring polymorphisms in the HIV-1C_{ZA} sequences in HIV-1B⁵⁸.

Molecular Modelling. The homology model of HIV-1C_{ZA} IN tetramer was generated using the cryoEM structure of HIV-1B IN intasome (PDB file 5U1C) in the presence of DNA substrate, using Prime version 4.2 of the Schrodinger Suite (Schrodinger, New York, NY, USA), integrated into Maestro of Schrodinger Suite, (Schrodinger Inc., NY) as described previously^{59,60}. The homology model was subjected to energy minimization (5,000 steps) to reduce steric overlap between residues using the “Impact” utility of the Schrödinger Suite and the OPLS_2005 force field as described before⁶¹. The modeled structure was submitted to the Structure Analysis and Verification Server (SAVES) (<https://services.mbi.ucla.edu/SAVES/>) as well as Protein Structure Preparation tool of SYBYL-X (version 2.1). No bad contacts were noted in the structures. The backbone torsion angles were checked by Ramachandran plot for allowed conformations of ϕ and ψ angles. All angles were in the allowed range. The mutant modeling was conducted with ‘Prime’ utility of Schrodinger Suite Fig. 3.

References

- Johnson, L. F. *et al.* Life expectancies of South African adults starting antiretroviral treatment: collaborative analysis of cohort studies. *PLoS Med.* **10**, e1001418 (2013).
- Trickey, A. *et al.* Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* **4**, e349–e356 (2017).
- Gupta, R. K. *et al.* Global trends in antiretroviral resistance in treatment-naïve individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. *Lancet (London, England)* **380**, 1250–8 (2012).
- Kantor, R. *et al.* Pretreatment HIV Drug Resistance and HIV-1 Subtype C Are Independently Associated With Virologic Failure: Results From the Multinational PEARLS (ACTG A5175) Clinical Trial. *Clin. Infect. Dis.* **60**, 1541–1549 (2015).
- Steege, K. *et al.* Moderate Levels of Pre-Treatment HIV-1 Antiretroviral Drug Resistance Detected in the First South African National Survey. *PLoS One* **11**, e0166305 (2016).
- Clinton Health Access Initiative, ARV MARKET REPORT Issue 8, 2017. Available at: https://clintonhealthaccess.org/content/uploads/2017/09/2017-ARV-Market-Report_Final-2.pdf (Accessed: 17th January 2018)
- Paton, N. I. *et al.* Assessment of Second-Line Antiretroviral Regimens for HIV Therapy in Africa. *N. Engl. J. Med.* **371**, 234–247 (2014).
- Hazuda, D. J. *et al.* Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* **287**, 646–50 (2000).
- Rockstroh, J. K. *et al.* Durable efficacy and safety of raltegravir versus efavirenz when combined with tenofovir/emtricitabine in treatment-naïve HIV-1-infected patients: final 5-year results from STARTMRK. *J. Acquir. Immune Defic. Syndr.* <https://doi.org/10.1097/QAI.0b013e31828ace69> (2013).
- Taha, H., Das, A. & Das, S. Clinical effectiveness of dolutegravir in the treatment of HIV/AIDS. *Infect. Drug Resist.* **8**, 339–52 (2015).
- Heger, E. *et al.* Development of a phenotypic susceptibility assay for HIV-1 integrase inhibitors. *J. Virol. Methods* **238**, 29–37 (2016).
- Yoshinaga, T. *et al.* Antiviral Characteristics of GSK1265744, an HIV Integrase Inhibitor Dosed Orally or by Long-Acting Injection. *Antimicrob. Agents Chemother.* **59**, 397–406 (2015).
- Ndashimye, E. High Time to Start Human Immunodeficiency Virus Type 1–Infected Patients on Integrase Inhibitors in Sub-Saharan Africa. *J. Infect. Dis.* **216**, 283–284 (2017).
- Quashie, P. K., Mesplède, T. & Wainberg, M. A. Evolution of HIV integrase resistance mutations. *Curr. Opin. Infect. Dis.* **26**, (2013).
- Grobler, J. A. & Hazuda, D. J. Resistance to HIV integrase strand transfer inhibitors: *in vitro* findings and clinical consequences. *Curr. Opin. Virol.* **8**, 98–103 (2014).
- US Food&Drug Administration, Approved Drug Products, New Drug Application 204790 Tivicay (Dolutegravir). Available at: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/204790Orig1s000Approv.pdf. (Accessed: 17th January 2018) (2013).
- Llibre, J. M. *et al.* Genetic barrier to resistance for dolutegravir. *AIDS Reviews* **17** (2014).
- Kobayashi, M. *et al.* *In Vitro* antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob. Agents Chemother.* **55**, 813–21 (2011).
- Raffi, F. *et al.* Once-daily dolutegravir versus twice-daily raltegravir in antiretroviral-naïve adults with HIV-1 infection (SPRING-2 study): 96 week results from a randomised, double-blind, non-inferiority trial. *Lancet Infect. Dis.* **13**, 927–935 (2013).
- Walmsley, S. L. *et al.* Dolutegravir plus Abacavir–Lamivudine for the Treatment of HIV-1 Infection. *N. Engl. J. Med.* **369**, 1807–1818 (2013).
- World Health Organization (WHO). The use of antiretroviral drugs for treating and preventing hiv infection. 97 Available at: http://apps.who.int/iris/bitstream/10665/208825/1/9789241549684_eng.pdf?ua=1. (Accessed: 17th January 2018) (2016).
- UNAIDS. Country Report South Africa. Available at: <http://www.unaids.org/en/regionscountries/countries/southafrica>. (Accessed: 17th January 2018) (2016).
- Meintjes, G. *et al.* Adult antiretroviral therapy guidelines 2017. *South. Afr. J. HIV Med.* **18**, 24 (2017).
- Goethals, O. *et al.* Primary mutations selected *in vitro* with raltegravir confer large fold changes in susceptibility to first-generation integrase inhibitors, but minor fold changes to inhibitors with second-generation resistance profiles. *Virology* **402**, 338–346 (2010).
- Tsiang, M. *et al.* Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile. *Antimicrob. Agents Chemother.* **60**, 7086–7097 (2016).
- Van Zyl, G. U. *et al.* Trends in Genotypic HIV-1 Antiretroviral Resistance between 2006 and 2012 in South African Patients Receiving First- and Second-Line Antiretroviral Treatment Regimens. <https://doi.org/10.1371/journal.pone.0067188> (2013).
- 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**, e90845 (2014).
- NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES. Annual Overview 2016/17. 23–24 (2017). Available at: http://www.nicd.ac.za/wp-content/uploads/2017/03/NICD_AR_2016_17.pdf (Accessed: 17th January 2018)

29. Garrido, C. *et al.* Broad phenotypic cross-resistance to elvitegravir in HIV-infected patients failing on raltegravir-containing regimens. *Antimicrob. Agents Chemother.* **56**, 2873–8 (2012).
30. Temesgen, Z. & Siraj, D. S. Raltegravir: first in class HIV integrase inhibitor. *Ther. Clin. Risk Manag.* **4**, 493–500 (2008).
31. Rhee, S.-Y. *et al.* Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74 (2008).
32. Hombrouck, A. *et al.* Mutations in human immunodeficiency virus type 1 integrase confer resistance to the naphthyridine L-870,810 and cross-resistance to the clinical trial drug GS-9137. *Antimicrob. Agents Chemother.* **52**, 2069–78 (2008).
33. Bessong, P. O. & Nwobegahay, J. Genetic Analysis of HIV-1 Integrase Sequences from Treatment Naive Individuals in Northeastern South Africa. *Int. J. Mol. Sci.* **14**, 5013–24 (2013).
34. Oliveira, M. F. *et al.* Genetic diversity and naturally polymorphisms in HIV type 1 integrase isolates from Maputo, Mozambique: implications for integrase inhibitors. *AIDS Res. Hum. Retroviruses* **28**, 1788–92 (2012).
35. Casadellà, M. *et al.* Primary resistance to integrase strand-transfer inhibitors in Europe: Table 1. *J. Antimicrob. Chemother.* **70**, 2885–2888 (2015).
36. Passaes, C. B. *et al.* Lack of Primary Mutations Associated With Integrase Inhibitors Among HIV-1 Subtypes B, C, and F Circulating in Brazil. *JAIDS J. Acquir. Immune Defic. Syndr.* **51**, 7–12 (2009).
37. Anstett, K., Brenner, B., Mesplede, T. & Wainberg, M. A. HIV drug resistance against strand transfer integrase inhibitors. *Retrovirology* **14**, 36 (2017).
38. Huang, W., Frantzell, A., Fransen, S. & Petropoulos, C. J. Multiple genetic pathways involving amino acid position 143 of HIV-1 integrase are preferentially associated with specific secondary amino acid substitutions and confer resistance to raltegravir and cross-resistance to elvitegravir. *Antimicrob. Agents Chemother.* **57**, 4105–13 (2013).
39. Doyle, T. *et al.* Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades. *J. Antimicrob. Chemother.* **70**, 3080–3086 (2015).
40. Hardy, I. *et al.* Evolution of a novel pathway leading to dolutegravir resistance in a patient harbouring N155H and multiclass drug resistance. *J. Antimicrob. Chemother.* **70**, (2015).
41. Shimura, K. *et al.* Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J. Virol.* **82**, 764–74 (2008).
42. Abram, M. E. *et al.* Lack of impact of pre-existing T97A HIV-1 integrase mutation on integrase strand transfer inhibitor resistance and treatment outcome. *PLoS One* **12**, (2017).
43. Rhee, S.-Y. *et al.* Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology* **5**, 74 (2008).
44. Malet, I. *et al.* Mutations Associated with Failure of Raltegravir Treatment Affect Integrase Sensitivity to the Inhibitor *In Vitro*. *Antimicrob. Agents Chemother.* **52**, 1351–1358 (2008).
45. Danion, F. *et al.* Non-virological response to a dolutegravir-containing regimen in a patient harbouring a E157Q-mutated virus in the integrase region. *J. Antimicrob. Chemother.* **69**, 2118–22 (2015).
46. Anstett, K., Cutillas, V., Fusco, R., Mesplède, T. & Wainberg, M. A. Polymorphic substitution E157Q in HIV-1 integrase increases R263K-mediated dolutegravir resistance and decreases DNA binding activity. *J. Antimicrob. Chemother.* **71**, 2083–2088 (2016).
47. Neogi, U. *et al.* *Ex vivo* antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV-1 non-B subtypes. *AIDS* **1** <https://doi.org/10.1097/QAD.0000000000001726> (2017).
48. Rossouw, T. M., Hitchcock, S. & Botes, M. The end of the line? A case of drug resistance to third-line antiretroviral therapy. *South Afr. J. HIV Med.* **17**, 3 (2016).
49. Fourati, S. *et al.* Cross-resistance to elvitegravir and dolutegravir in 502 patients failing on raltegravir: a French national study of raltegravir-experienced HIV-1-infected patients. *J. Antimicrob. Chemother.* **70**, 1507–1512 (2015).
50. Cahn, P. *et al.* Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naïve adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study. *Lancet* **382**, 700–708 (2013).
51. Quashie, P. K., Han, Y.-S., Hassounah, S., Mesplède, T. & Wainberg, M. A. Structural Studies of the HIV-1 Integrase Protein: Compound Screening and Characterization of a DNA-Binding Inhibitor. *PLoS One* **10**, e0128310 (2015).
52. Jacobs, G. B. *et al.* Emergence and diversity of different HIV-1 subtypes in South Africa, 2000–2001. *J. Med. Virol.* **81**, 1852–1859 (2009).
53. Swanson, P., Devare, S. G. & Hackett, J. Molecular Characterization of 39 HIV-1 Isolates Representing Group M (Subtypes A-G) and Group O: Sequence Analysis of *gagp24*, *pol* Integrase, and *env gp41*. *AIDS Res. Hum. Retroviruses* **19**, 625–629 (2003).
54. Laethem, K. V. *et al.* A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J. Virol. Methods* **153**, 176–181 (2008).
55. Pineda-Peña, A. C. *et al.* V. A. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: Performance evaluation of the new REGA version 3 and seven other tools. *Infect. Genet. Evol.* **19**, 337–348 (2013).
56. Struck, D., Lawyer, G., Ternes, A.-M., Schmit, J.-C. & Bercoff, D. P. COMET: adaptive context-based modeling for ultrafast HIV-1 subtype identification. *Nucleic Acids Res.* **42**, e144–e144 (2014).
57. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
58. Rhee, S.-Y. *et al.* HIV-1 *pol* mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes.
59. Passos, D. O. *et al.* Cryo-EM structures and atomic model of the HIV-1 strand transfer complex intasome. *Science* **355**, 89–92 (2017).
60. Häggblom, A., Svedhem, V., Singh, K., Sönnnerborg, A. & Neogi, U. Virological failure in patients with HIV-1 subtype C receiving antiretroviral therapy: an analysis of a prospective national cohort in Sweden. *lancet. HIV* **3**, e166–74 (2016).
61. Singh, K. *et al.* Biochemical mechanism of HIV-1 resistance to rilpivirine. *J. Biol. Chem.* **287**, 38110–23 (2012).

Acknowledgements

This study was funded by the National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa. DB was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Sciences, University of Würzburg. KS acknowledges partial support by National Institute of Health, Clinical and Translational Science CTSA grant UL1 TR002345. UN acknowledges support received from Swedish Research Council Establishment grant (2017–01330). Grant support for this study was received from the National Research Foundation (NRF) of South Africa and the Poliomyelitis Research Foundation (PRF) of South Africa. DB was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Sciences, University of Würzburg. KS acknowledges partial support by National Institute of Health, Clinical and Translational Science CTSA grant UL1 TR002345. UN acknowledges support received from Swedish Research Council Establishment grant (2017–01330). We also thank Stellenbosch University and the Harry Crossley Foundation for additional financial support.

Author Contributions

G.B.J. and S.E. conceptualized the study SE obtained the T.V. samples. D.B. performed the PCR amplification and sequence analyses with assistance from A.E. D.B. and A.E. both wrote the first draft. K.S. and U.N. performed data analysis and prepared Figures 1–3. G.M.I. and R.C. read, and edited the final manuscript. All authors performed quality control check, data analysis, edited, read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-22914-5>.

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) 2018