Long-Range HIV Genotyping Using Viral RNA and Proviral DNA for Analysis of HIV Drug Resistance and HIV Clustering

Vlad Novitsky, Melissa Zahralban-Steele, Mary Fran McLane, Sikhulle Moyo, Erik van Widenfelt, Simani Gaseitsiwe, Joseph Makhema, M. Essex
Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana

The goal of the study was to improve the methodology of HIV genotyping for analysis of HIV drug resistance and HIV clustering. Using the protocol of Gall et al. (A. Gall, B. Ferns, C. Morris, S. Watson, M. Cotten, M. Robinson, N. Berry, D. Pillay, and P. Kellam, J Clin Microbiol 50:3838–3844, 2012, doi:10.1128/JCM.01516-12), we developed a robust methodology for amplification of two large fragments of viral genome covering about 80% of the unique HIV-1 genome sequence. Importantly, this method can be applied to both viral RNA and proviral DNA amplification templates, allowing genotyping in HIV-infected subjects with suppressed viral loads (e.g., subjects on antiretroviral therapy [ART]). The two amplicons cover critical regions across the HIV-1 genome (including pol and env), allowing analysis of mutations associated with resistance to protease inhibitors, reverse transcriptase inhibitors (nucleoside reverse transcriptase inhibitors [NRTIs] and nonnucleoside reverse transcriptase inhibitors [NNRTIs]), integrase strand transfer inhibitors, and virus entry inhibitors. The two amplicons generated span 7,124 bp, providing substantial sequence length and numbers of informative sites for comprehensive phylogenetic analysis and greater refinement of viral linkage analyses in HIV prevention studies. The long-range HIV genotyping from proviral DNA was successful in about 90% of 212 targeted blood specimens collected in a cohort where the majority of patients had suppressed viral loads, including 65% of patients with undetectable levels of HIV-1 RNA loads. The generated amplicons could be sequenced by different methods, such as population Sanger sequencing, single-genome sequencing, or next-generation ultradeep sequencing. The developed method is cost-effective—the cost of the long-range HIV genotyping is under $140 per subject (by Sanger sequencing)—and has the potential to enable the scale up of public health HIV prevention interventions.

HIV genotyping is a critical tool for antiviral drug resistance testing that has revolutionized HIV care and advanced HIV-related research. Routine antiretroviral (ART) drug resistance testing is useful in choosing an optimal treatment regimen and monitoring its efficiency in clinical practice (1–12). HIV genotyping has been used successfully in research on HIV transmission clusters and HIV transmission dynamics (13–35).

Initial broadly used ART regimens included combinations of nucleoside reverse transcriptase (RT) inhibitors (NRTIs) and nonnucleoside reverse transcriptase inhibitors (NNRTIs). To monitor the emergence of drug resistance mutations associated with NRTIs and NNRTIs, HIV genotyping targeted viral sequences spanning an approximately 1,000- to 1,300-bp region of the HIV-1 genome encoding viral protease and partial RT, using viral RNA as a template for amplification. While the RNA-based approach works well in antiretroviral therapy (ART)-naive individuals, it is less successful if levels of viral replication are low, such as in individuals on ART. The sequence length of traditional RNA-based HIV genotyping for drug resistance is relatively short and does not cover the HIV-1 region encoding viral integrase or the viral envelope, hindering analysis of drug resistance mutations associated with integrase strand transfer inhibitors or entry inhibitors. The global scale up of ART treatment and successful introduction of integrase strand transfer inhibitors and entry inhibitors into clinical trials and clinical practice necessitate modification of traditional methods of HIV genotyping.

Two commercial genotyping assays, ViroSeq HIV-1 from Abbott Molecular and TruGene HIV-1 from Siemens Molecular Diagnostics, have been widely used for analysis of HIV-1-associated drug resistance. Both genotyping kits were extensively tested and validated (36–45). While the ViroSeq HIV-1 kit is still on the market, Siemens discontinued selling and supporting the TruGene HIV-1 kit in 2014. The ViroSeq HIV-1 kit covers the entire protease-coding region and the RT region encoding the first 320 amino acids. The TruGene HIV-1 sequences span the protease (amino acids 4 to 99)- and RT (amino acids 40 to 240)-coding regions. The CDC supplies WHO-designated and CDC-supported President’s Emergency Plan for AIDS Relief (PEPFAR) Genotyping Laboratories with the ATCC HIV-1 Drug Resistance Genotyping kit (46) for drug resistance testing. Many experienced genotyping laboratories have developed their own in-house amplification and sequencing protocols (11, 47–56), including identification of minor viral variants that are normally missed by commercial genotyping kits (57–61). All of these approaches generally include smaller and more restricted regions for testing HIV-1 drug resistance.

Recently, the protocol developed by Gall et al. (62) has enabled
high-throughput, nearly full-length HIV-1 genome genotyping in individuals infected with multiple HIV-1 subtypes. The method has become the cornerstone of the PANGEA (Phylogenetics and Networks for Generalized HIV Epidemics in Africa)-HIV Consortium (http://www.pangea-hiv.org/) aiming to establish worldwide scientific collaborations across phylogenetics, public health, and epidemiology. The Gall protocol (62) targets viral RNA as a template for cDNA synthesis and amplification and is very robust and reproducible when the HIV-1 RNA load is high (e.g., above 10,000 cps/ml). However, specimens with levels of HIV-1 RNA below 1,000 cps/ml, or lower thresholds, present a substantial challenge, and few of those samples could be genotyped. This is consistent with the commercially available assays for HIV drug resistance genotyping, ViroSeq and TrueGene, which are unable to genotype specimens with low or undetectable HIV-1 RNA loads.

In HIV infection, proviral DNA presents an alternative template for HIV genotyping. Drug resistance mutations detected in viral RNA from plasma and proviral DNA from peripheral blood mononuclear cells (PBMCs) or dried blood spots (DBS) show substantial correlation in treated patients, suggesting that either compartment is suitable for the detection of mutations as a virological guide for clinical care (63–65).

It is known that amplified HIV sequences and sequences from proviral DNA could have substantial numbers of guanine-to-adenine transitions. Such an inordinate number of identical G-to-A transitions in proviral DNA could have substantial numbers of guanine-to-ad-69). G-to-A hypermutations produce multiple stop codons and long-range HIV genotyping, the technique was applied to a set of extra round of PCR, (iii) selection of robust primers, and (iv) modifications in the pol region encoding HIV-1 protease and the first 335 amino acids of reverse transcriptase.

The outcome of the long-range genotyping using proviral DNA, as well as viral RNA, as a template for amplification and sequencing. The outcome of the long-range genotyping is two large fragments that span about 80% of the HIV-1 genome: (i) amplicon 1, spanning the region encoding gp120 V1C5 (34, 79, 80), nucleotide positions 2253 to 3554; and (ii) amplicon 2, spanning env, nef, and the TATA box in the U3 region of the 3' long terminal repeat (LTR) and corresponding to amplicon 4 in the study of Gall et al. (62), nucleotide positions 5967 to 9151; (iii) ViroSeq, a partial pol sequence spanning the region encoding HIV-1 protease and the first 335 amino acids of reverse transcriptase and corresponding to the sequence produced by ViroSeq (39, 44, 45, 78), nucleotide positions 2253 to 3554; and (iv) V1C5, a partial env sequence spanning the region encoding gp120 V1C5 (34, 79, 80), nucleotide positions 6570 to 7757. In addition, the following combinations of the subgenomic regions included concatenated amplicon 1 plus amplicon 2 and amplicon 1 plus V1C5. All multiple-sequence codon-based alignments were generated using MUSCLE (81) in MEGA6 (82).

To prevent sample contamination, basic laboratory rules were enforced, including controlled flow of specimens, use of dedicated areas and equipment, proper training, and routine implementation of a quality assurance/quality control (QA/QC) program.

Analysis of drug resistance. The WHO 2009 list of mutations for surveillance of transmitted drug-resistant HIV strains was used for analysis of protease inhibitor (PI)-, NRTI-, and NNRTI-associated mutations. The list of PI-associated mutations included 40 mutations at 18 positions across protease. The list of NNRTI mutations included 34 mutations at 15 positions in RT. The list of NNRTI mutations included 19 mutations at 10 positions across RT. The International AIDS Society (IAS)-USA list (2014 update) of drug resistance mutations in HIV-1 was used for analysis of integrase strand transfer inhibitors (20 mutations at 11 positions in integrase) and entry inhibitors (10 mutations at 7 positions in gp41) (3).

APOBEC-induced hypermutations. The APOBEC-induced hypermutations were assessed by Hypermut (83) at the Los Alamos National Laboratory (LANL) HIV Database (http://www.hiv.lanl.gov/). The HIV-1 subtype C (HIV-1C) consensus sequence was used as a reference. Two parameters related to APOBEC-induced hypermutations were analyzed: adjusted hypermutations and the hypermutation ratio. The adjusted hypermutations were expressed as a number of identified hypermutations adjusted by sequence length. The hypermutation ratio was computed as the ratio between weighted mutations (matched mutations out of potential mutations) and weighted controls (control mutations out of potential controls) and was derived as a statistical outcome of the Hypermut package (83).

Definition of the HIV cluster. An HIV cluster was defined as a viral lineage that gives rise to a monophyletic subtree of the overall phylogeny with strong statistical support. The bootstrapped maximum-likelihood
RESULTS
Long-range HIV genotyping. The original protocol for nearly full-length HIV-1 genome genotyping by amplification of four large overlapping amplicons in a single round of RT-PCR using viral RNA as a template was developed by Gall et al. (62). The protocol of Gall et al. (62) is robust and highly reproducible for samples with relatively high HIV-1 RNA levels. However, specimens with low or undetectable levels of HIV-1 RNA presented a substantial challenge for amplification from viral RNA. Attempts to apply the original protocol to proviral DNA produced large numbers of nonspecific products evident from smeared “ladders” on the electrophoretic gel (data not shown).

The modifications of the protocol of Gall et al. (62) included the following steps: (i) focus on 2 amplicons 2 and 4 in the original protocols) instead of 4 amplicons, (ii) an extra round of PCR where the amplified ~8.3-kb product was used as a template for the second round of PCR, (iii) highly specific primers for the first round of PCR and for cDNA synthesis (for viral RNA templates), and (iv) modified PCR running conditions.

The rationale for focusing on two instead of four amplicons was driven by a balance between sequencing data and cost. The two amplicons have lengths of 3,574 bp and 3,550 bp (HXB2 nucleotide length), which cumulatively covers about 80% of the unique full-length HIV-1 genome sequence (Fig. 1). The first amplicon (corresponding to amplicon 2 in the study by Gall et al. [62]) spans partial gag at the 3′ end and almost the entire pol (HXB2 nucleotide positions 1,486 to 5,058). The second amplicon (corresponding to amplicon 4 in the study by Gall et al. [62]) spans vpu, env, nef, and the 3′ LTR up to the TATA box in the U3 region (HXB2 nucleotide positions 5,967 to 9,517).

Amplification of a large fragment spanning almost the entire HIV-1 genome (Fig. 1, hatched bar) was introduced as the first round of PCR (RT-PCR for the RNA template). Primers OFM19 and SK145 (see Table S1 in the supplemental material) substantially increased the specificity of viral amplification. For the proviral DNA template, the 1st round of PCR was run with primers SK145 and OFM19. The PrimeStar GXL DNA polymerase (Takara; catalog number R050A) was used in 30 amplification cycles with the annealing temperature at 62°C (98°C for 10 s, 62°C for 15 s, and 68°C for 9 min cycling). For the RNA template, cDNA synthesis with primer OFM19 was followed by PCR with primers SK145 and OFM19 in a single-tube RT-PCR. The SuperScript III One-Step RT-PCR High Fidelity enzyme (Invitrogen; catalog number 12574035) was used with a cDNA synthesis step of incu-
and a single amplicon 2 sequence was considered a success. If a sequence per subject. Therefore, generation of a single amplicon 1 HIV quasispecies. The goal of this study was to obtain a single HIV most efficient sequencing strategy to overcome the complexity of amplicon 1 sequences and to 100% of amplicon 2 sequences is the variable regions of the HIV-1 genome, with multiple indels. Our study aimed to address the multiplicity of HIV infection or the diversity of viral quasispecies, multiple sequences (e.g., 20 per targeted region per subject) could be generated by appropriate amplification methods.

Cloning was performed with a PCR cloning kit (NEB; catalog number E1202S) using Fast-Media Amp XGal (Invivogen; catalog number fas-am-x). Ligation, transformation, and plating were performed according to the manufacturer’s instructions. Colonies were checked for inserts with EmeraldAmp GT PCR master mix (TaKaRa; catalog number RR310A) and submitted to GENEWIZ for colony sequencing. A list of sequencing primers used with clones is presented in Table S3 in the supplemental material.

All sequence contigs were assembled with SeqScape v.2.7.

Troubleshooting. Some amplification issues during long-range HIV genotyping, such as lack of or insufficient amplification, an overamplified product, or the presence of multiple bands, could be resolved by troubleshooting. The initial amplification results could guide troubleshooting. A lack of visible bands (or weak bands) on the gel after second-round PCR could be resolved by decreasing the annealing temperature in the first-round PCR to 58°C and/or increasing the number of cycles in the first-round PCR to 35 or increasing the amount of RNA template (e.g., up to 5 μl). The overamplified products could be overcome by reducing the number of cycles in the first-round PCR to 25 or in the second-round PCR to 20 to 25 or by decreasing the amount of input template. Multiple bands on the gel could be resolved by either extracting the band of the right size from the gel using the Wizard SV Gel and PCR Clean-Up System (Promega; catalog number A9281) or rerunning the first-round PCR in replicates and with serial dilutions.

HIV genotyping results. Amplicon 1 was amplified and sequenced in 649 HIV-infected subjects (a single sequence per subject), while amplicon 2 was amplified and sequenced in 90 subjects.

The long-range HIV genotyping from proviral DNA was applied to 212 specimens collected from subjects participating in the BCPP baseline household survey in the first four communities, Ranaka, Digawana, Molapowaboajang, and Otse, from November 2013 until June 2014. The distribution of amplified and sequenced samples from proviral DNA is presented in Table 1. Amplicon 1 was successfully amplified in 89.6% (95% CI, 84.5% to 93.2%) of the cases. Viral sequences were obtained for all amplified samples. The majority of amplified amplicon 1 sequences, 144 of 167 (86.2%; 95% CI, 79.8% to 90.9%), were obtained by direct Sanger

<p>| TABLE 1 Summary of HIV genotyping from proviral DNA, amplicon 1 (BCPP) |
|---------------------------------|------------------|------------------|</p>
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value for proviral DNA specimens</th>
<th>Subset with available HIV-1 RNA load data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>n</td>
<td>Proportion (95% CI)</td>
</tr>
<tr>
<td>Attempted cases</td>
<td>212</td>
<td>0.896 (0.845–0.932)</td>
</tr>
<tr>
<td>Amplified cases</td>
<td>190</td>
<td>0.104 (0.068–0.155)</td>
</tr>
<tr>
<td>Cases in which amplification failed</td>
<td>22</td>
<td>0.879 (0.822–0.920)</td>
</tr>
<tr>
<td>Cases sequenced by direct Sanger sequencing</td>
<td>167</td>
<td>0.121 (0.080–0.178)</td>
</tr>
<tr>
<td>Cases with partial sequences by direct Sanger sequencing</td>
<td>23</td>
<td>0.121 (0.080–0.178)</td>
</tr>
</tbody>
</table>

a The proportion of failed cases was calculated from the number of attempted cases.

b The proportion of sequenced cases was calculated from the number of amplified cases.

c Gaps at ≤10% of sequence length resolved by cloning.
specimens from proviral DNA (n/H11005 = 21) (Fig. 2, pie charts).

The proportion of sequenced cases was calculated from the number of amplified cases. The proportion of failed cases was calculated from the number of attempted cases.

Amplification from viral RNA. To assess the utility of long-range HIV genotyping for amplification and sequencing from a viral RNA template, we performed small-scale genotyping (n = 32) from viral RNA in plasma (Table 2). The HIV-1 RNA load was available for 31 of these samples and was above 1,000 cps/ml in 29 cases. A subset of 23 cases were successfully amplified and sequenced. Interestingly, two of nine specimens that failed amplification from proviral DNA (HIV-1 RNA loads, 1,576 cps/ml and 5,620 cps/ml), were successfully amplified from viral RNA.

The nine failed cases included one sample with unknown and eight specimens with available viral loads. Among the latter group, two samples had viral loads below 1,000 cps/ml (181 and 497 cps/ml), 5 samples had viral loads between 1,191 and 8,528 cps/ml, and 1 sample had a viral load of 156,821 cps/ml. The last failed sample, with a high viral load, also failed amplification from proviral DNA, apparently suggesting an intrinsic problem with mismatch of amplification primers.

Overall, the long-range HIV genotyping from proviral DNA (for amplicon 1) was successful in about 90% of targeted blood specimens collected in a cohort where a majority of the patients had suppressed viral loads, including 65% of patients with undetectable HIV-1 RNA loads.

Amplification from viral RNA. To assess the utility of long-range HIV genotyping for amplification and sequencing from a viral RNA template, we performed small-scale genotyping (n = 32) from viral RNA in plasma (Table 2). The HIV-1 RNA load was available for 31 of these samples and was above 1,000 cps/ml in 29 cases. A subset of 23 cases were successfully amplified and sequenced. Interestingly, two of nine specimens that failed amplification from proviral DNA (HIV-1 RNA loads, 1,576 cps/ml and 5,620 cps/ml), were successfully amplified from viral RNA.

The nine failed cases included one sample with unknown and eight specimens with available viral loads. Among the latter group, two samples had viral loads below 1,000 cps/ml (181 and 497 cps/ml), 5 samples had viral loads between 1,191 and 8,528 cps/ml, and 1 sample had a viral load of 156,821 cps/ml. The last failed sample, with a high viral load, also failed amplification from proviral DNA, apparently suggesting an intrinsic problem with mismatch of amplification primers.

Overall, the long-range HIV genotyping from proviral DNA (for amplicon 1) was successful in about 90% of targeted blood specimens collected in a cohort where a majority of the patients had suppressed viral loads, including 65% of patients with undetectable HIV-1 RNA loads.

Amplification from viral RNA. To assess the utility of long-range HIV genotyping for amplification and sequencing from a viral RNA template, we performed small-scale genotyping (n = 32) from viral RNA in plasma (Table 2). The HIV-1 RNA load was available for 31 of these samples and was above 1,000 cps/ml in 29 cases. A subset of 23 cases were successfully amplified and sequenced. Interestingly, two of nine specimens that failed amplification from proviral DNA (HIV-1 RNA loads, 1,576 cps/ml and 5,620 cps/ml), were successfully amplified from viral RNA.

The nine failed cases included one sample with unknown and eight specimens with available viral loads. Among the latter group, two samples had viral loads below 1,000 cps/ml (181 and 497 cps/ml), 5 samples had viral loads between 1,191 and 8,528 cps/ml, and 1 sample had a viral load of 156,821 cps/ml. The last failed sample, with a high viral load, also failed amplification from proviral DNA, apparently suggesting an intrinsic problem with mismatch of amplification primers.
Analysis of mutations associated with antiretroviral drug resistance. Amplicon 1 covers almost the entire HIV-1 pol gene and allows analysis of mutations associated with antiretroviral drug resistance to PIs, NRTIs, NNRTIs, and integrase strand transfer inhibitors. Amplicon 2 covers the entire HIV-1 env gene and allows analysis of mutations associated with drug resistance to virus entry inhibitors.

To illustrate the validity of long-range HIV genotyping for analysis of mutations associated with antiretroviral drug resistance, we estimated drug resistance profiles within two groups of specimens originating from the MPP and BCPP studies. Amplicon 1 sets included 192 MPP sequences and 186 BCPP sequences. Amplicon 2 sets included 35 MPP and 55 BCPP sequences.

Despite relatively rare use of protease inhibitors in Botswana, mutations associated with resistance to PIs were detected at five positions in protease: D30N (5% in MPP and 6% in BCPP), M46I (5% in MPP and 10% in BCPP), G73S (10% in MPP and 9% in BCPP), I85V (1% in MPP), and N88S (1% in BCPP). The encod-

The presence of G-to-A hypermutations. The number of matched amplicon 1 plus 2 sequences was limited to 83. The extents of HIV clustering within this small set were compared for three long loci—amplicon 1 (3,574 bp), amplicon 2 (3,550 bp), and the concatenated ampli-

G-to-A hypermutations. The presence of G-to-A hypermutations in the products amplified from proviral DNA is not surprising, as massive APOBEC-induced G-to-A transitions in retrovi-

HIV cluster analysis. To demonstrate the utility of the long-range HIV genotyping for analysis of HIV transmission dynamics and viral linkage, we compared the extents of clustering within viral sequences generated in this study. The concatenated ampli-

The majority of viral sequences with antiretroviral mutations had high rates of APOBEC-induced hypermutations, suggesting association between hypermutations and drug resistance mutations. The distribution of APOBEC-induced hypermutations among 36 MPP sequences with drug resistance mutations within protease, RT, and integrase is presented in Table S4 in the supplement-

HIV-1C sequences with identified drug resistance mutations demonstrated high rates of APOBEC-induced hypermutations. The horizontal box plots in Fig. 3 indicate the distributions of the adjusted numbers of hypermutations (Fig. 3A and C) and the hypermutation ratios (Fig. 3B and D) among viral sequences with drug resistance mutations in relation to the distribution of hypermutation parameters in the entire set of sequences. Comparison of these distributions indicates an association between APOBEC-induced hypermutations and drug resistance mutations.

To further address how G-to-A hypermutations can affect drug resistance mutations, we compared hypermutations between two groups with and without M184I mutations. Individuals with M184I mutations have higher adjusted numbers of hypermuta-

The proportions of clustered HIV sequences were compatible within long loci (Table 3). For example, at a bootstrap support of 0.80, the proportions of clustered HIV sequences were 0.265, 0.289, and 0.337 for amplicon 1, amplicon 2, and concatenated amplicons 1 plus 2, respectively. For the short loci ViroSeq and V1C5 at the same bootstrap support of 0.80, the proportions of HIV sequences in clusters were 0.157 and 0.145, respectively. The
The proportion of clustered sequences seemed to be higher for long regions than for short loci, although the difference reached significance at the 0.05 level in selected comparisons only.

A larger set of available HIV-1C sequences (n/H11005/547) included matched viral sequences for amplicon 1 and the V1C5 region of gp120 generated in our previous studies (34, 35, 80). Clustering patterns were compared for two long loci, amplicon 1 and concatenated amplicon 1 plus V1C5, and for two short regions across the HIV-1 genome, ViroSeq and V1C5. Similar to the small set of HIV sequences (n/H11005/83), the proportion of clustered sequences in the large set (n/H11005/547) was higher for long loci than for short regions (Table 4).

To address whether longer loci are associated with a greater extent of HIV clustering, we analyzed congruent (/H11001/H11001/ and /H11002/H11002/) and discordant (/H11001/H11002/ and /H11002/H11001/) clustering between different combinations of long and short HIV-1C sequences (Fig. 5). At all bootstrap thresholds from 0.70 to 1.0, amplicon 1 and concatenated amplicon 1 plus V1C5 demonstrated greater extents of HIV clustering than ViroSeq and V1C5 sequences (Fig. 5, gray background, indicating a significant difference).

**Estimated cost of long-range HIV genotyping.** The estimated cost for amplification and Sanger sequencing of both amplicons 1 and 2 in this study was $137.50 for proviral DNA and $139.75 for viral RNA. This includes the cost of reagents, materials, and disposables for nucleic acid isolation, amplification (RT-PCR and PCR), purification of amplicons, cloning up to 30% of amplicon 1 products and 100% of amplicon 2 products, and Sanger sequencing. The estimated cost does not include labor, training, supervision, or indirect costs.

**DISCUSSION**

A technique for long-range HIV genotyping from both viral RNA and proviral DNA has been presented. Using proviral DNA as a template, long-range HIV genotyping was successfully performed in one of the BCPP cohorts with a high proportion of virologically suppressed individuals, with a success rate of about 90%.

Both clinical trials and clinical care could benefit from routine use of long-range HIV genotyping. The proposed long-range HIV genotyping has the potential to improve the methodology of drug resistance testing, to broaden the spectrum of monitored ARVs, and to enable surveillance of transmitted drug resistance. Mapping of HIV transmission networks performed by long-range genotyping could help reveal transmitting viral variants in treatment-as-prevention studies. Implementation of long-range HIV genotyping could allow greater refinement of viral linkage analyses in HIV prevention studies and better coordination with evaluation of prevention strategies based on such interventions as behavior change, male circumcision, and treatment as prevention.

**FIG 3** Distributions of APOBEC-induced hypermutations in sequences amplified from proviral DNA (histograms). The horizontal box plots outline the distributions of APOBEC-induced hypermutations in subsets of sequences with identified drug resistance mutations. The box plots are drawn to the x-axis scale. The left and right box boundaries indicate lower and upper quartiles, the line within the box is the median, and the left and right whiskers indicate minimum and maximum values without outliers. (A and B) Amplicon 1 (n = 649), distribution of hypermutations adjusted by sequence length (A) and distribution of hypermutation ratio data (B) (see Materials and Methods). (C and D) Amplicon 2 (n = 90), distribution of hypermutations adjusted by sequence length (C) and distribution of hypermutation ratio data (D) (see Materials and Methods).
A comprehensive strategy of HIV genotyping could include two steps. First, a viral RNA template for amplification could be targeted, if the HIV-1 RNA load is relatively high (e.g., above 1,000 cps/ml). If amplification is successful, there is no need for proviral DNA. However, if amplification from viral RNA does not work, or the HIV-1 RNA load is below 1,000 cps/ml or undetectable, using proviral DNA is a logical step toward successful HIV genotyping. Complementary use of both viral RNA and proviral DNA templates could be an efficient and cost-effective approach for HIV genotyping.

Long-range HIV genotyping enables analysis of drug resistance (both transmitted and acquired) for all major groups of ARVs, including protease inhibitors, NRTIs, NNRTIs, integrase strand transfer inhibitors, and virus entry inhibitors. A comprehensive analysis of HIV drug resistance is feasible due to the long sequence length of generated amplicons that span the HIV-1 pol and env genes. While long-range HIV genotyping is able to identify drug-resistant mutations, its potential to enable the scale up of public health HIV prevention interventions across communities.

In this study, we demonstrated that long-range HIV genotyping using proviral DNA could be successfully applied to a population with a high level of ART-experienced individuals, which normally presents a challenge for HIV genotyping from viral RNA. In fact, more than 70% of individuals in the BCPP cohort participating in the baseline household survey in the first four communities had HIV-1 RNA levels below 1,000 cps/ml, including 65% with undetectable levels of HIV-1 RNA, below 40 cps/ml. The ongoing scale up of national ARV programs in Africa has led to a growing number of individuals with suppressed HIV-1 RNA loads across communities. The presented technique of long-range HIV genotyping from proviral DNA should alleviate challenges and enable analysis of HIV drug resistance and HIV transmission dynamics using samples collected from individuals on ART.
resistance mutations, it cannot distinguish transmitted and acquired drug resistance mutations without additional information on the sampling strategy and stage of HIV infection. For example, drug resistance mutations identified in individuals during early stages of HIV infection (e.g., in seroconverters) are likely to represent transmitted drug resistance mutations. In contrast, specimens collected in chronic HIV infection, or from individuals on ART, are more likely to be associated with acquired HIV drug resistance.

The G-to-A hypermutations observed in sequences amplified from proviral DNA and their relation to drug resistance mutations should be interpreted cautiously in the context of a specific study. Our data suggest that G-to-A hypermutations are likely to contribute to critical drug resistance mutations, such as M184I. We recommend controlling viral sequences generated from proviral DNA for the adjusted number of hypermutations and/or the hypermutation ratio using the online package Hypermut (83) at the LANL HIV Database (http://www.hiv.lanl.gov/) and the subtype consensus sequence as a reference. Based on interquartile range (IQR) boundaries in our data, the adjusted number of hypermutations above 2.8% (the 1st IQR in individuals with M184I) indicates a hypermutated sequence and that below 0.5% (the 3rd IQR in individuals without M184I) suggests a nonhypermutated sequence. Whether HIV-associated drug resistance mutations should be interpreted differently depending on the extent of G-to-A hypermutations still needs to be addressed in future studies.

The study showed the utility of long-range genotyping for analysis of HIV transmission dynamics and HIV clustering. A greater extent of clustering for longer HIV sequences in this study corroborates the results of our recent study (93), which used a set of nearly full-length HIV-1C sequences from the LANL HIV Database (http://www.hiv.lanl.gov/). Longer HIV sequences are more informative for HIV cluster analysis due to a larger number of informative sites (93). The technique of long-range HIV genotyping allows the use of amplicon 1 and amplicon 2 sequences either separately or in concatenation for a powerful cluster analysis. The concatenated amplicons 1 and 2 span about 80% of the unique HIV-1 genome sequence and could be considered a cheaper alternative to nearly full-length HIV-1 sequencing. A combination of conserved (amplicon 1) and variable (amplicon 2) regions could help to deal with different and/or unknown stages of HIV infection in an analyzed set of viral sequences. The choice of a particular bootstrap value and filtering by the threshold of pairwise distances and/or internode certainty (94, 95) could depend on the specific scientific question and take into account the specifics of an analyzed set of sequences, including sampling density (35).

In summary, the presented technique of long-range HIV genotyping using viral RNA and proviral DNA can help in analysis of HIV drug resistance and HIV clustering in cohorts and populations on ART when amplification from viral RNA is unsuccessful due to low HIV-1 RNA loads.

ACKNOWLEDGMENTS

We are very grateful to all participants in the BHP projects in Botswana. We thank the CDC and the Botswana Ministry of Health for their collaboration. We thank Lendsey Melton for excellent editorial assistance.

The Mochudi Prevention Project in Botswana was supported and funded by NIH grant R01 AI083036, an HIV Prevention Program for Mochudi, Botswana. The GWAS on Determinants of HIV-1 Subtype C Infection study was supported and funded by NIH grant R01 AI092715. The Botswana Combination Prevention Project (BCPP) has been supported by the President’s Emergency Plan for AIDS Relief (PEPFAR) through the U.S. Centers for Disease Control and Prevention under the terms of grant number U01 GH000447.

REFERENCES

Novitsky et al.


of APOBEC3 editing in the absence of drug exposure. AIDS 26:1619–1624.  
http://dx.doi.org/10.1097/QAD.0b013e3283560703.

http://dx.doi.org/10.7448/IAS.16.1.18472.

http://dx.doi.org/10.1089/aid.2008.0072.

http://dx.doi.org/10.1089/aid.2014.0211.

http://dx.doi.org/10.1177/1740774514523351.

http://dx.doi.org/10.1016/j.ijgev.2013.05.019.


http://dx.doi.org/10.1038/ncomms1325.

http://dx.doi.org/10.1093/bioinformatics/btl446.

http://dx.doi.org/10.1093/bioinformatics/btu033.

http://dx.doi.org/10.1093/molbev/mst120.

http://dx.doi.org/10.1093/bed/16.6.400.
Correction for Novitsky et al., Long-Range HIV Genotyping Using Viral RNA and Proviral DNA for Analysis of HIV Drug Resistance and HIV Clustering

Vlad Novitsky, Melissa Zahralban-Steele, Mary Fran McLane, Sikhulile Moyo, Erik van Widenfelt, Simani Gaseitsiwe, Joseph Makhema, M. Essex

Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA; Botswana Harvard AIDS Institute Partnership, Gaborone, Botswana

Volume 53, no. 8, p. 2581–2592, 2015. Table S1 in the supplemental material: The sequence for primer OFM19 was incorrect; it was missing the triplet ATT. Revised supplemental material is posted at http://jcm.asm.org/content/53/8/2581/suppl/DCSupplemental.