Analysis of HIV-1 variants by cloning DNA generated with the ViroSeq™ HIV-1 Genotyping System

Susan H. Eshleman[1], Dana Jones[1], Tamara Flys[1], Olga Petrauskene[2], and J. Brooks Jackson[1]

[1]The Johns Hopkins Medical Institutions, Baltimore, MD and [2]Applied Biosystems, Foster City, CA, USA

BioTechniques 35:614-622 (September 2003)

There are three major classes of antiretroviral drugs approved in the United States for treatment of human immunodeficiency virus type 1 (HIV-1) infection: protease inhibitors, nucleoside and nucleotide reverse transcriptase inhibitors, and non-nucleoside reverse transcriptase inhibitors (NNRTIs). Unfortunately, the efficacy of these drugs is often limited by HIV-1 drug resistance, which is usually caused by mutations in the protease and reverse transcriptase enzymes. Sequence-based HIV-1 genotyping assays can be used to detect antiretroviral drug resistance mutations and help guide use of antiretroviral drugs for treatment and prevention of HIV-1 infection.

HIV-1 genotyping is complicated by the fact that individuals infected with HIV-1 typically harbor a large population of genetically diverse viral variants. In most genotyping assays, population sequencing is used to characterize the viruses in a plasma sample. Assays based on population sequencing may detect drug resistance mutations present in mixed viral populations (e.g., mixtures of viruses with and without a mutation) (1–3). However, those assays are not designed to detect minority variants. Furthermore, if multiple drug resistance mutations are detected as mixtures in a given sample, it is not possible to determine whether the individual drug resistance mutations are present together on a single virus or exist in different viral sub-populations. Genetic linkage of HIV-1 drug resistance mutations is important, since individual resistance mutations can enhance or reverse the effects of other resistance mutations in the same virus. Other methods are needed to examine further the genetic evolution of drug-resistant HIV-1 and to characterize the emergence and fading of drug resistance mutations following antiretroviral drug exposure.

Table 1. Nevirapine (NVP) Resistance Mutations Detected by Population Sequencing and Sequencing of Cloned Plasmids

	NVP Resistance Mutations
Sample 1	
Population sequencing	Y181C
Clone 1-1	G190S
Clone 1-2	None
Clone 1-3	Y181C
Clone 1-4	Y181C
Clone 1-5	Y181C
Clone 1-6	G190A
Clone 1-7	V108I, Y181C
Clone 1-8	Y181C
Clone 1-9	Y181C
Clone 1-10	None
Sample 2	
Population sequencing	K103N, Y181C, G190A
Clone 2-1	K103N
Clone 2-2	G190A
Clone 2-3	Y181C
Clone 2-4	G190A
Clone 2-5	V106A
Clone 2-6	Y181C
Clone 2-7	None
Clone 2-8	None
Clone 2-9	Y181C, G190A
Clone 2-10	Y181C

NVP, nevirapine. Mutations detected by population sequencing were detected as mixtures of mutant and wild-type amino acids (e.g., Y + C at position 181).

The Celera Diagnostics ViroSeq™ HIV-1 Genotyping System (previously marketed as the Applied Biosystems ViroSeq™ HIV-1 Genotyping System, Celera Diagnostics, Alameda, CA, USA) is a commercially available system for analysis of antiretroviral drug resistance mutations in HIV-1 (4). Analysis in the ViroSeq system involves HIV-1 RNA isolation and reverse transcription with murine Moloney virus reverse transcriptase, followed by a single 40-cycle PCR using Ampli*Taq* Gold™ polymerase (Applied Biosystems, Foster City, CA, USA). PCR products are purified using spin columns and analyzed by agarose gel electrophoresis prior to sequencing. Population sequencing of the PCR products is performed using premixed BigDye® (Applied Biosystems) sequencing reagents with 6–7 different primers. Analysis is performed using an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). Software provided with the ViroSeq system is used to assemble sequence data from the different primers into a contiguous consensus sequence that can be inspected for identification of drug resistance mutations. The ViroSeq system was approved in 2003 by the Food and Drug Administration for clinical use. This system has been used to analyze HIV-1 drug resistance in clinical trials, such as the AIDS Clinical Trials Group (ACTG)

368, 388, and 398 trials (unpublished data). The assay performs well for analysis of subtype (clade) B HIV-1 (5) as well as for non-B subtypes (6–8).

Analysis of HIV-1 using the ViroSeq system provides information about the predominant viral population in a plasma sample. Individual DNA molecules amplified in PCR generated in the ViroSeq system are derived from individual viruses. Therefore, cloning, isolation, and analysis of PCR products produced in this system would allow characterization of the *pol* regions of individual viruses. This would allow one to examine genetic linkage of drug resistance mutations, and may reveal the presence of minority variants with additional antiretroviral drug resistance mutations not detected by population sequencing. Well-characterized cloned viral variants could also be used to further characterize the effects of specific drug resistance mutations on viral fitness and phenotypic drug resistance.

Unfortunately, commercial cloning kits cannot be used to clone PCR products produced with the ViroSeq system. This is because the ViroSeq system uses a deoxyuridine triphosphate (dUTP)/uracil N-glycosylase (UNG) contamination control system. In this system, PCRs are performed with dUTP in place of dTTP. To prevent contamination of PCRs with PCR products generated in previous reactions, each PCR is proceeded by a brief incubation with the enzyme UNG, which destroys any contaminating dUTP-containing DNA (9). The UNG enzyme is then destroyed prior to amplification by heating the reaction to 93°C. Similar dUTP/UNG contamination systems have been incorporated into other PCR-based clinical assays, such as the AMPLICOR HIV-1 MONITOR™ Test viral load assay (Roche Diagnostics, Branchburg, NJ, USA). Cloning ViroSeq PCR products directly into most strains of *Escherichia coli* is problematic, since most strains naturally contain the UNG enzyme. Ligated plasmids with dUTP-containing inserts are degraded in the bacteria and cannot be isolated. In this report, we describe a method for cloning and isolating dUTP-containing HIV-1 *pol* region DNA amplified in the ViroSeq genotyping system.

Plasma samples for this study were obtained from two Ugandan women enrolled in the HIVNET 012 clinical trial, which compared the efficacy of two different antiretroviral regimens [nevirapine (NVP) or zidovudine] for prevention of HIV-1 mother-to-child transmission (10). These samples were collected 7 days after administration of a single dose of NVP given at the onset of labor. The women were antiretroviral drug naïve prior to NVP administration and did not receive any additional antiretroviral drugs, consistent with the standard of care in Uganda. These samples were genotyped using the ViroSeq system. PCR products generated in this system were diluted to 105 µL after column purification. Approximately 60 µL of PCR products from each reaction were used for population sequencing (HIV-1 genotyping with ViroSeq). While this system is designed for analysis of the major viral population, mixtures of nucleotides can also be detected, reflecting the genetic diversity of HIV-1 variants in clinical samples. Nucleotide mixtures are detected by the ViroSeq software (version 2.5) when a secondary peak is at least 30% of the primary peak (version 2.0 product insert). The electropherograms obtained from sense and antisense primers in the ViroSeq system can be compared to confirm the presence of the mixture. Analysis in the ViroSeq system revealed the Y181C NVP resistance mutation in sample 1, and revealed three NVP resistance mutations, K103N, Y181C, and G190A, in sample 2 (Table 1). All of these mutations were detected as mixtures [e.g., nucleotide sequences encoded a mixture of lysine (K) and asparagine (N) at position 103, a mixture of tyrosine (Y) and cytosine (C) at position 181, and a mixture of glycine (G) and alanine (A) at position 190 in HIV-1 reverse transcriptase].

To analyze sequences from individual HIV-1 variants in the samples, we first prepared electrocompetent *E. coli* using strains lacking the UNG enzyme (UNG-negative). Two UNG-negative *E. coli* strains, BD2314 [CGSC #6799, (11)] and BD1528 (CGSC #7221) were

available from the E. coli Genetic Stock Center (Yale University, New Haven, CT, USA). Bacteria were rendered competent for electroporation as described previously (12). Electroporation was performed using a MicroPulser™ Electroporator (Bio-Rad Laboratories, Hercules, CA, USA), according to the manufacturer's instructions, and bacteria containing plasmids were selected on LB/Amp plates. The transformation efficiencies of the two UNG-negative *E. coli* strains using a pUC19 control plasmid were similar ($1–5 \times 10^7$ colonies/µg). This was lower than the transformation efficiency seen with commercially prepared electrocompetent *E. coli* (approximately 1×10^9 colonies/µg with the TOP10 One Shot® Electrocomp™ cells provided with the TOPO TA Cloning® Kit, version N; Invitrogen, Carlsbad, CA, USA). Because the preparation of BD1528 *E. coli* often arced during the electroporation procedure, the BD2314 strain was used for subsequent experiments.

PCR products generated using the ViroSeq system were cloned using the TOPO TA Cloning Kit. One microliter of the dUTP-containing PCR products remaining after genotyping each of the samples was used for cloning. In the ViroSeq system, PCR products are amplified with a *Taq* DNA polymerase, which adds single deoxyadenosine (A) residues to the 3′ ends of PCR products (13). The pCRII-TOPO® vector (Invitrogen), which is synthesized with single 3′-thymidine (T) overhangs, is optimized for TA cloning of PCR products produced with *Taq* DNA polymerases. Storage of PCR products may be associated with loss of some terminal A residues. Therefore, to enhance the efficiency of the ligation reaction, additional nontemplated A residues were added to the ViroSeq PCR products prior to ligation, using procedures recommended by the manufacturer. PCR products were then ligated into the pCRII-TOPO vector. Ligation was performed according to the manufacturer's instructions. The resulting plasmids, which had dUTP-containing HIV-1 inserts, were isolated following electroporation into the UNG-negative *E. coli* and selected on LB/Amp plates. Plating of the entire electroporation reaction yielded 100–200 colonies for each sample. Plasmids were isolated from random colonies using a commercial kit (QIAwell® Ultra Plasmid Kit; Qiagen, Valencia, CA, USA) according to the manufacturer's instructions, except that the culture volume used for plasmid isolation was increased to 3 mL, and the cell lysis step was extended to 15 min. Approximately 50%–75% of the colonies for samples 1 and 2 contained inserts of the expected size (approximately 1.8 kb). For each sample, 18–20 plasmids were sequenced using the ViroSeq sequencing module. For plasmid sequencing, 8 µL of plasmid isolated by the miniprep procedure were used without further modification in each of the sequencing reactions.

DNA sequences of 10 representative plasmids cloned from each of the samples described above are shown in Figures

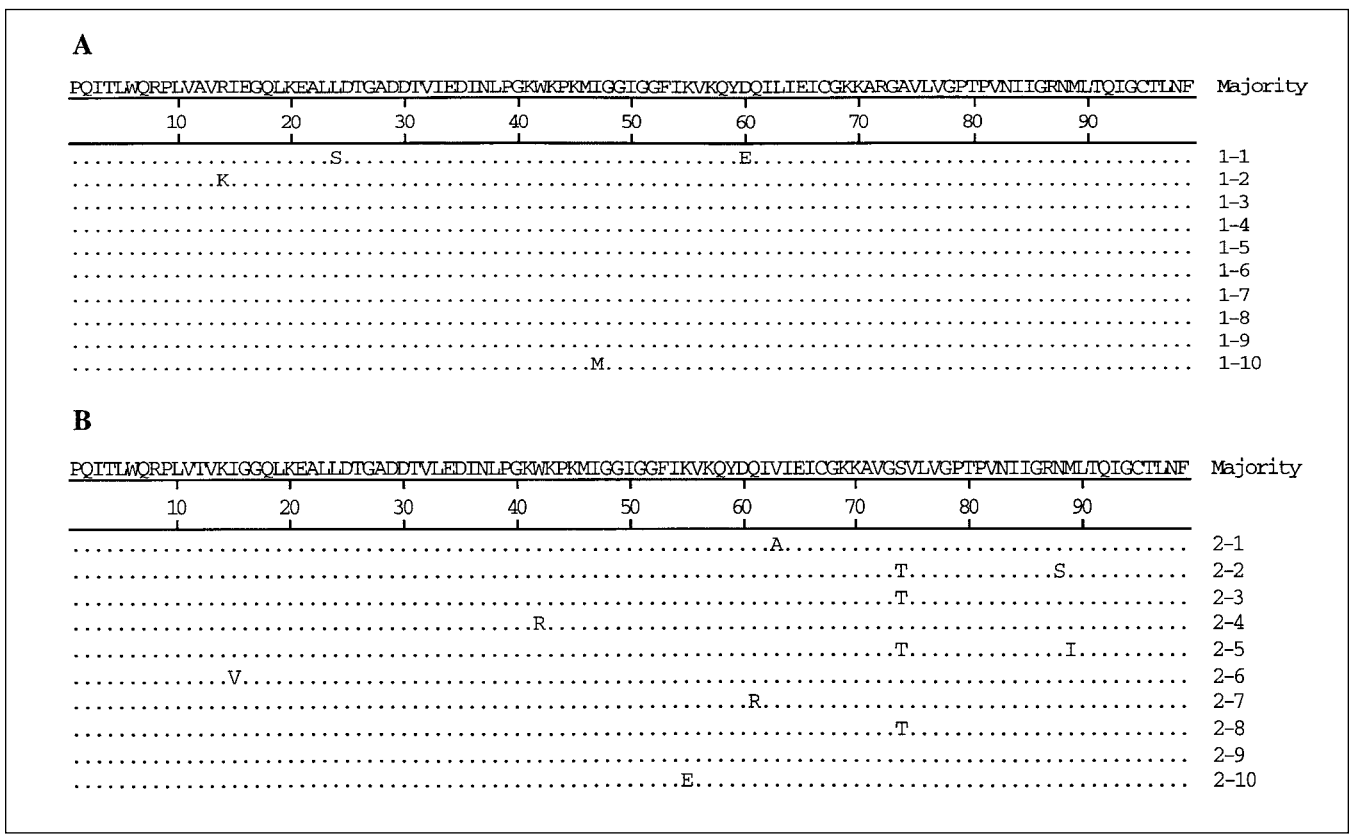

Figure 1. Protease sequences of cloned plasmids derived from two clinical samples. Protease sequences (amino acids 1–99) from cloned plasmids were aligned using the Clustal method (Lasergene; DNASTAR, Madison, WI, USA). (A) Clones derived from sample 1 (clones 1–1 to 1–10). (B) Clones derived from sample 2 (clones 2–1 to 2–10). A consensus sequence for the10 plasmid sequences from each sample is shown above each alignment (majority). Dots in the alignment indicate identity with the consensus sequence. Amino acids that differ from the consensus sequence are shown.

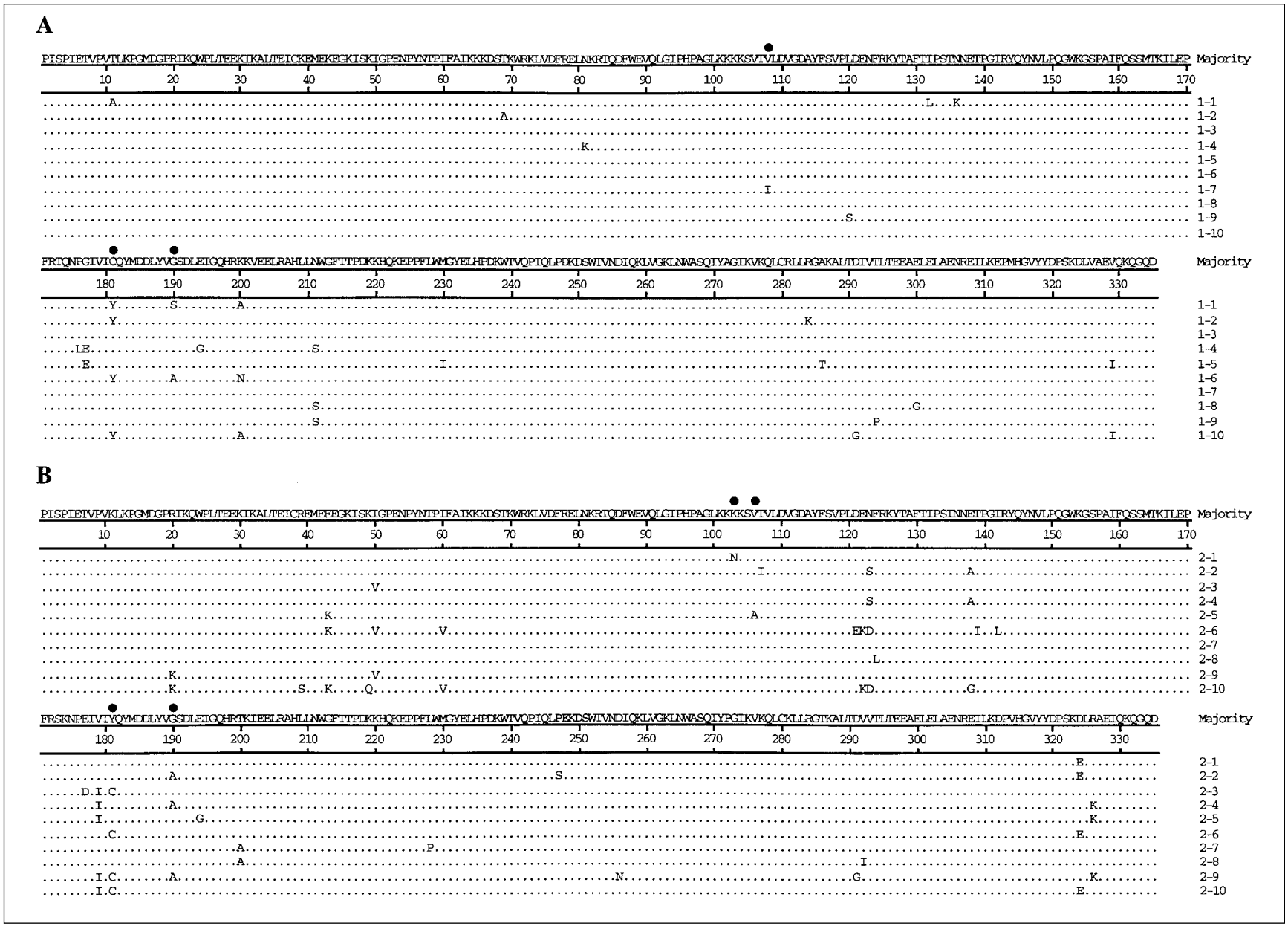

Figure 2. Reverse transcriptase sequences of cloned plasmids derived from two clinical samples. Reverse transcriptase sequences (amino acids 1–335) from cloned plasmids were aligned using the Clustal method (Lasergene). (A) Clones derived from sample 1 (clones 1–1 to 1–10). (B) Clones derived from sample 2 (clones 2–1 to 2–10). A consensus sequence for the 10 plasmid sequences from each sample is shown above each alignment (majority). Dots in the alignment indicate identity with the consensus sequence. Amino acids that differ from the consensus sequence are shown. Black circles over the consensus sequence indicate the position of nevirapine (NVP) resistance mutations detected in each sample.

1 and 2. The NVP resistance mutations detected in each sequence are shown in Table 1. For sample 1, the Y181C mutation, which was detected in population sequencing, was detected in only 6 of the 10 plasmids. For sample 2, all three mutations detected in population sequencing (K103N, Y181C, and G190A) were detected among the cloned plasmids. However, most plasmids contained only one of these mutations, suggesting independent selection of viruses with each of the mutations. For both samples, other NVP resistance mutations, not detected in population sequencing, were also seen in the cloned plasmids. For example, in sample 1, the NVP resistance mutations (V108I, G190A, and G190S) were each detected in 1 of the 10 plasmids. In sample 2, the NVP resistance mutation V106A was detected in 1 of 10 plasmids. Other minor sequence differences were noted among the protease and reverse transcriptase sequences of the plasmid clones (Figures 1 and 2), consistent with the natural genetic variation among HIV-1 viruses in infected individuals.

The polymerase used for DNA amplification in the ViroSeq system is Ampli*Taq* Gold. Like other *Taq* enzymes, Ampli*Taq* Gold lacks proofreading activity and is known to introduce errors in nucleotide sequences during amplification (14). These errors occur at approximately 1 in every 50,000 to 100,000 nucleotides incorporated (http://www.appliedbiosystems.com). This corresponds to approximately one error in every one or two cloned HIV-1 inserts in this system. Mutations could also be introduced during cloning by propagation in UNG-negative *E. coli*. The UNG enzyme normally catalyzes the first step of a pathway for repair of uracil-containing DNA. An elevated level of G:C → A:T transitions is observed in UNG-negative *E. coli* (15). Such errors are generally not apparent in population sequencing, but may be detected when sequences of cloned plasmids are analyzed. Some of these changes would be difficult to detect in clones from HIV-1 viruses, since the sequences of HIV-1 viruses are inherently variable. In contrast, termination (stop) codons and mutations that cause shifts in the reading frame (insertions or deletions) are easily identified.

Analysis of HIV-1 *pol* sequences in almost 40 cloned plasmids did not reveal any stop codons. However, nucleotide insertions and deletions were observed. In sample 1, such changes were detected in 3/18 cloned plasmids (data not shown). These included a single nucleotide insertion (insertion of an A near codon 101 in reverse transcriptase) and two single nucleotide deletions [protease codon 11 = (GT-) and reverse transcriptase codon 93 = (GG-)]. In sample 2, we detected such changes in 4/20 cloned plasmids (data not shown). Three plasmids had single nucleotide deletions [two with reverse transcriptase codon 249 = (AA-) and one with reverse transcriptase codon 87 = (T-C)], and two plasmids (one with the reverse transcriptase codon 87 deletion) had single nucleotide insertions, one with an A insertion near reverse transcriptase codon 104 and one with an A insertion near reverse transcriptase codon 30. Interestingly, all of the insertions involved an extra A residue in a homopolymeric run of six A residues. In contrast, single base deletions appeared to occur at random and did not occur in homopolymeric runs.

Single nucleotide frameshift mutations, including insertions of A nucleotides in homopolymeric runs of As, are commonly detected among sequences from HIV-1 viral variants (16). Some of these mutations may be artifacts introduced during amplification, cloning, or sequencing. However, because similar mutations have been observed in HIV-1 variants using different analytical methods, some of these mutations are likely to be naturally occurring inactivating mutations in defective viral genomes. The presence of these insertions and deletions did not interfere with the analysis of NVP resistance mutations.

The methods described in this report can be used for analysis of the genetic linkage of HIV-1 drug resistance mutations. Furthermore, characterization of as few as 10 cloned viral variants from each sample revealed additional drug resistance mutations not detected by routine genotyping. The cloning procedure described can be performed using as little as 1 µL of the PCR products that remain following routine HIV-1 genotyping, without the use of additional plasma. This may be advantageous in some settings (e.g., analysis of samples from clinical trials) where plasma samples may be limited in quantity or unavailable.

The method described is optimized using the commercial TOPO TA cloning vector, pCRII-TOPO. There are three advantages to using this vector for cloning: (*i*) TA cloning is more efficient than blunt-end cloning. Furthermore, the vector used is covalently bound to topoisomerase I, promoting rapid and efficient ligation (17). (*ii*) The TA cloning approach can be used to clone essentially any DNA insert, independent of its DNA sequence. This is an advantage, given the high genetic variability among HIV-1 sequences. (*iii*) The cloning site of the pCRII-TOPO is within the *lacZ*α gene, so blue/white color selection can be used in suitable bacterial strains. Although the UNG-negative *E. coli* strain described here cannot be used for color selection, it may be possible to use this strain for the initial recovery of plasmids with dUTP-containing inserts, and then transfer the plasmid mixture into a second strain of *E. coli* for plasmid isolation and color selection.

The methods described in this report use two commercial kits: the TOPO TA Cloning Kit is used for plasmid construction, and the ViroSeq HIV-1 Genotyping System is used for plasmid sequencing and mutation analysis. To reduce the costs involved, it should also be possible to use alternative methods for plasmid construction (e.g., alternative plasmid vectors) and alternative approaches for plasmid sequencing.

ACKNOWLEDGMENTS

The authors thank Manohar Furtado (Applied Biosystems) and Eric Shulse (Celera Diagnostics) for providing reagents used in this study. The authors thank Androulla Lambrianidou (Invitrogen) for helpful discussions and Mary Berlyn (The E. coli Genetic Stock Center, Yale University) for providing UNG-negative bacterial strains. The authors thank Bette Korber (Los Alamos National Laboratories) for providing information about inactivating mutations in defective HIV-1 viruses, and Christos Petropoulos, Ellen Paxinos, and Neil Parkin (ViroLogic, Inc.) for helpful discussions. This work was supported by the HIV Network for Prevention Trials (HIVNET) and sponsored by the U.S. Na-

tional Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Dept. of Health and Human Services (DHHS), through contract N01-AI-35173 with Family Health International, contract N01-AI-45200 with Fred Hutchinson Cancer Research Center, and subcontracts with Johns Hopkins University/Makerere University (NOI-AI-35173-417); the HIV Prevention Trials Network (HPTN) sponsored by the NIAID, National Institutes of Child Health and Human Development (NICH/HD), National Institute on Drug Abuse, National Institute of Mental Health, and Office of AIDS Research, of the NIH, DHHS (U01-AI-46745 and U01-AI-48054); the Adult AIDS Clinical Trials Groups (NIH, Division of AIDS, NIAID); and RO1- 42965-01 (NIH-CH/HD).

REFERENCES

1. **Leitner, T., E. Halapi, G. Scarlatti, P. Rossi, J. Albert, E.M. Fenyo, and M. Uhlen.** 1993. Analysis of heterogeneous viral populations by direct DNA sequencing. BioTechniques *15*:120-127.
2. **Chadwick, R.B., M.P. Conrad, M.D. McGinnis, L. Johnston-Dow, S.L. Spurgeon, and M.N. Kronick.** 1996. Heterozygote and mutation detection by direct automated fluorescent DNA sequencing using a mutant Taq DNA polymerase. BioTechniques *20*:676-683.
3. **Larder, B.A., A. Kohli, S.D. Kemp, M.N. Kronick, and R.D. Henfrey.** 1993. Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. Nature *365*: 671-673.
4. **Shafer, R.W., K. Dupnik, M.A. Winters, and S.H. Eshleman.** 2001. A guide to HIV-1 reverse transcriptase and protease sequencing for drug resistance studies. *In* Human Retroviruses and AIDS 2001: A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. Los Alamos National Laboratory, Los Alamos, NM.
5. **Cunningham, S., B. Ank, D. Lewis, L. Wei, J. Dileanis, J.B. Jackson, P. Palumbo, P. Krogstad, and S.H. Eshleman.** 2001. Performance of the Applied Biosystems ViroSeq HIV-1 Genotyping System for analysis of HIV-1 in pediatric plasma samples. J. Clin. Microbiol. *39*: 1254-1257.
6. **Mracna, M., G. Becker-Pergola, J. Dileanis, L.A. Guay, S. Cunningham, J.B. Jackson, and S.H. Eshleman.** 2001. Performance of the Applied Biosystems ViroSeq HIV-1 Genotyping System for sequence-based analysis of non-subtype B HIV-1 from Uganda. J. Clin. Microbiol. *39*:4323-4327.
7. **Fontaine, E., C. Riva, M. Peeters, J.C. Schmit, E. Delaporte, K. Van Laethem, K. Van Vaerenbergh, J. Snoeck, et al.** 2001. Evaluation of two commercial kits for the detection of genotypic drug resistance on a panel of HIV type 1 subtypes A through J. J. Acquir. Immune Defic. Syndr. *28*: 254-258.
8. **Jagodzinski, L.L., J.D. Cooley, M. Weber and N.L. Michael.** 2003. Performance characteristics of human immunodeficiency virus type 1 (HIV-1) genotyping systems in sequence-based analysis of subtypes other than HIV-1 subtype B. J. Clin. Microbiol. *41*:998-1003.
9. **Longo, M.C., M.S. Berninger and J.L. Hartley.** 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. Gene *93*:125-128.
10. **Guay, L.A., P. Musoke, T. Fleming, D. Bagenda, M. Allen, C. Nakabiito, J. Sherman, P. Bakaki, et al.** 1999. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-infant transmission of HIV-1 in Kampala, Uganda: HIVNET-012 randomised trial. Lancet *354*:795-802.
11. **Duncan, B.K.** 1985. Isolation of insertion, deletion, and nonsense mutations of the uracil- DNA glycosylase (ung) gene of Escherichia coli K-12. J. Bacteriol. *164*:689-695.
12. **Seidman, C.E., K. Struhl, and J. Sheen.** 1997. Introduction of plasmid DNA into cells: High efficiency transformation by electroporation, p. 1.22–21.23. *In* F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl (Eds.), Short Protocols in Molecular Biology. John Wiley & Sons, New York.
13. **Clark, J.M.** 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. *16*:9677-9686.
14. **Sambrook, S. and D.W. Russell.** 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
15. **Duncan, B.K. and B. Weiss.** 1982. Specific mutator effects of ung (uracil-DNA glycosylase) mutations in Escherichia coli. J. Bacteriol. *151*: 750-755.
16. **Kuiken, C.L., B. Foley, B. Hahn, B. Korber, P.A. Marx, F. McCutchan, J.W. Mellors, and S. Wolinksy.** 2001. HIV Sequence Compendium. Los Alamos National Laboratory, Los Alamos, NM.
17. **Shuman, S.** 1994. Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase. J. Biol. Chem. *269*: 32678-32684.

Received 16 May 2003; accepted 20 June 2003.

Address correspondence to Susan H. Eshleman, Department of Pathology, The Johns Hopkins Medical Institutions, Ross Bldg. 646, 720 Rutland Ave., Baltimore, MD 21205, USA. e-mail: seshlem@jhmi.edu