Heteroduplex Mobility Analysis

From the Laboratory of James I. Mullins
Department of Microbiology
University of Washington School of Medicine
Seattle, WA, 98195-8070
Tel: 206-732-6166
Fax: 206-732-6055

jmullins@u.washington.edu

This manual and further information about the HMA is available on the World Wide Web at:
http://ubik.microbiol.washington.edu/HMA/index.html

Important contributions to these protocols and preparation of this manual were made by Eric L. Delwart, Edward W. Hanley, Belinda Herring, Gerald H. Learn Jr., Frederick Lyagoba, Allen G. Rodrigo, and Raj Shankarappa.

Live hyperlinks in this document are shown in red. Click on any red text to open your Internet or email program and connect to the linked document or email address, or to link to the cross-referenced text within this document.

Introduction

The following protocols describe PCR amplification and polyacrylamide gel electrophoresis conditions for subtype determination of HIV-1 envelope (*env*) sequences using the heteroduplex mobility assay (HMA). PCR primers and reference plasmids containing *env* genes from different subtypes are provided separately as part of the HMA kit. Representative polyacrylamide gel images are also presented and described to assist in the interpretation of the data. For a detailed discussion of the background of the HMA, please refer to *Genetic subtyping of Human Immunodeficiency Virus Using a Heteroduplex Mobility Assay* by Delwart et al.[4]

This HMA kit will forever be a work in progress since HIV continues to evolve, with certain regions of the genome becoming too diverse for study by this technique, and other regions only recently becoming sufficiently diverse to analyze. We solicit your help in making this endeavor more fruitful for everyone. The rapid communication of technical improvements and the distribution of subclones of newly identified HIV subtypes, circulating recombinants, and genetic outliers should be added to improve later versions of the kit.

This kit was made possible by the generous sharing of cloned HIV-1 *env* DNA reagents by a large number of researchers for use as subtype references. We are also indebted to the many individuals who tested a prototype of this kit and protocol. In particular, we acknowledge help from Drs. Francine E. McCutchan, Joost Louwagie, Andrew Artenstein, Beatrice Hahn, Nice Shindo, Cecilia Graziozi, Anthony Fauci, Elisabeth Menu, Françoise Barré-Sinoussi, Marc Girard, Wouter Janssens, Guido van der Groen, Alexsei Bobkov, Jonathan Weber, Leonardis Kostrikis, David Ho, Helga Rübsamen-Waigmann, Manual Grez, Jay Levy and other members of the UNAIDS Network for HIV Isolation and Characterization. In addition, we thank Drs. Saladin Osmanov and Jose Esparza from the UNAIDS Global Program on AIDS, Dr. James Bradac from the Division of AIDS, NIH-NIAID, and especially Drs. Opendra Sharma and Martha Matocha and Ms. Kate Traynor from the NIH AIDS Research and Reference Reagent Program for their support and interest in this project.

Strategy for Subtype Determination - an Overview

Subtype determination by HMA is based on evaluating the mobility of heteroduplexes formed with DNA fragments from the test sample and the subtype references provided in the kit. As described in detail below, nested PCR is used to generate either 1.2 Kb, 0.7, or 0.5 Kb *env* gene fragments from uncharacterized strains of HIV-1. The 1.2 Kb fragments encompass the V1 through V5 variable regions of the gp120 coding sequence, while the 0.7 Kb fragments encompass an internal C2–V5 fragment, and the 0.5 Kb fragment covers the V3 loop and the adjoining constant regions (C2–V3–C3). We currently recommend performing initial subtyping experiments with the 0.7 kb fragment, and if it is not possible to obtain this fragment by PCR or if a subtype cannot be unambiguously assigned using this fragment, then we suggest using the 0.5 kb fragment and then, if needed, the larger, 1.2 fragment. In previous versions of the kit we recommended the 1.2 kb fragment for initial subtyping efforts. We have changed our recommendation due to the fact that the 1.2 kb fragment is amplified with lower efficiency, causing the occasional specimen to be scored as PCR negative.

Depending on which fragment you plan to evaluate, subtype reference sequences can be amplified directly from the plasmids provided with the kit without the need for nested PCR amplification. PCR reactions are then checked for appropriate product length and yields by agarose gel electrophoresis. Heteroduplexes are then formed by mixing, denaturing and annealing amplified DNA from the unknowns with amplified DNA from the reference strains, and their mobilities analyzed on polyacrylamide gels. Heteroduplexes formed between the unknown sample and the most closely related sequences are expected to exhibit the fastest mobilities on these gels. If the closest relatives are all from a single genetic subtype, the likely subtype of that strain is thus determined. The goal should always be to positively identify the subtype of the unknown. Therefore, heteroduplexes formed with a set of references from the assigned subtype should have markedly faster mobilities than with other subtypes, although the degree of this distinction required for subtype assignment is subjective. The 0.5kb fragment is recommended for use because the extent of variation in this fragment, particularly with regard to length variation, is the lowest of the three fragments. Thus, more divergent strains can be typed using this fragment.

To familiarize yourself with the assay and to be sure to obtain reliable results, we recommend that the first few samples analyzed be compared to all of the reference strains provided in the kit, and that you compare your results with the gel photos reproduced in this document. When analyzing samples from a geographic region suspected of having only a single or a few subtypes (e.g. subtype B in North America, subtypes B and E in Thailand) it is possible to reduce the number of reference strains used for initial subtyping. In Thailand, for

example, where both the subtype B and E strains have to date not diversified as markedly as in other parts of the world, it is possible to positively identify the subtype of an unknown by comparison to references from only the B and E clades and one D subtype control. The low degree of diversity within these two subtypes in individuals infected in Thailand, and consequently the very fast mobility of their heteroduplexes, provides a high degree of confidence in assigning subtypes. Even though the B subtype predominates in the United States, subtype determination is more complex because this virus population, having been introduced earlier, has diversified for a longer time. Therefore, subtype B *env* genes are not as closely related to each other (thus resulting in slower heteroduplex mobilities) as the Thai B or E *env* genes are to each other. We recommend subtyping of North American-derived viral sequences by comparison to multiple B reference strains in addition to at least one strain from a few other subtypes, especially its closest relative, subtype D (corresponding to negative controls). Where there is the likely presence of multiple and highly diversified subtypes, such as in Sub-Saharan Africa, it is necessary to compare each unknown to a panel of multiple reference sequences from each of the available subtypes.

Over the years, HMA has proven quite reliable, with no known examples of erroneous subtype determinations from analysis of technically adequate polyacrylamide gels and multiple reference sequences from each of the relevant subtypes. In contrast, the failure to identify a subtype occurs with some frequency, depending again on the quality of the gel and the divergence of the sample from the reference samples employed. Comparison to only a single reference strain from a given subtype often provides an ambiguous result. Thus it is always best to use a panel of multiple reference sequences from each subtype expected to be related to your unknowns.

There is great flexibility in the choice of subtype reference strains that can be used. The use of references derived from your own cohorts is recommended, since within a given geographic region and cohort the resident HIV strains are more likely to be closely related to each other than to geographically more distant reference strains. This permits more rapid and confident assignments of subtype and intrasubtype relationships to be drawn. If *env* genes of known sequence from the region under study are available, they may also be used for generation of the standard curves required for inference of numerical estimates of sequence divergence.

As alluded to above, some samples may be difficult to subtype by HMA. Such results generally indicate detection of a genetic outlier within a known subtype, the detection of a new subtype[2], or a recombinant virus derived from parental viruses from different subtypes (e.g. Zam184[5] is now known to correspond to a subtype A–C recombinant in *env*). It is also possible, however, that the difficulty stems from a large deletion or insertion in the V1–V2 or V4–V5 region, each of which are highly prone to such length variations[3] or less likely, in another region of *env*. For this reason, the next step in characterizing an initially non-subtypable strain would be to try another fragment of *env* or determine a portion of the DNA sequence. Use of the 0.5 kb ED31–ED33 fragments covering the C2–V3–C3 region, which normally has little length variation[4], typically obviates this problem[3]. There is of course no barrier to performing HMA with other regions of the HIV-1 genome. Indeed, an HMA based on *gag* gene sequences has recently been developed[12]. However, it must be noted that with high prevalence of inter-subtype recombinants in certain regions of the world such as Africa, subtype assignments using one gene may not necessarily be identical to that using another gene. The degree of variation required for good discrimination of heteroduplexes in nondenaturing polyacrylamide gels is within a wide range of 3–20% (counting mismatches only; for detection of length variation, single base gaps are sufficient for their discrimination), hence the degree of mismatch expected to be encountered should also guide the choice of fragment to use for HMA.

Kit Components

Subtype References

The following set will grow as new subtypes are discovered and evaluated for use in the kit:

Reference Abbreviation*	HIV-1 Strain	Accession Number	Cloned Fragment	Plasmid Vector	Size bp	EcoRI Sites	HindIII Sites	Drug	Contributor**
A2	IC144	n/a	ED5/ED12	pCR1	4381	1	2	Kan	A
A3	SF170	M66533	gp160	pCR2	6550	2	1	Kan/Amp	B
B1	BR20	U08797	gp160	pCR2	6772	2	1	Kan/Amp	C
B2	TH14	U08801	gp160	pCR2	6750	2	2	Kan/Amp	C
B3	SF162	M65024	6.6kb 3'end	pUC19	3952	2	2	Amp	B
C1	MA959	U08453	gp160	pCR2	6783	2	1	Kan/Amp	D
C2	ZM18	L22954	ED5/ED12	pCR1	5466	2	2	Kan	A
C3	IN868	U07103	ED5/ED12	pCR1	4150	1	1	Kan	E
D1	UG21	U08804	gp120	pCR2	5621	2	2	Kan/Amp	C
D3	UG46	U08737	gp120	pGEM-T	273	0	0	Amp	C
E1	TH22	U09131	gp160	pCR2	6778	2	2	Kan/Amp	C
E2	TH06	U08810	gp160	pCR2	6294	2	2	Kan/Amp	C
E3	CAR7	n/a	ED5/ED12	pCR2	5197	2	1	Kan/Amp	F
F1	BZ162	L22084	gp120	pCR2	5664	2	1	Kan	A
F2	BZ163	L22085	gp120	pCR1	6226	1	1	Kan	A
G1	RU131	U08364	ED5/ED12	pUC18	1353	1	1	Amp	G
G2	LBV21-7	U09664	gp160	pUC19	2595	1	3	Amp	H
G3	VI525	U09665	gp160	pUC19	2604	1	2	Amp	H
H2	VI557	U09666	gp160	pCR2	4836	2	1	Kan/Amp	H
J1	SE9173	AF082395	ED5/ED12	pCR2	1240	Unknown	Unknown	Kan/Amp	H
J2	SE9280	AF082394	ED5/ED12	pCR2	1240	Unknown	Unknown	Kan/Amp	H

**Each subtype reference is supplied as plasmid DNA at a concentration of 10 ng/µl in 10 mM Tris HCl (pH 7.4) plus 0.1 mM EDTA (pH 8.0). Although transformation and growth of these plasmids in E. coli is possible, we strongly recommend that after any such procedure, the integrity of all reference fragments be verified. This can be accomplished by amplifying specific regions and testing them in an HMA as illustrated below. This will ensure that cross-contamination (>1 heteroduplex pattern present per lane) or mislabeling of plasmid preps (erroneous heteroduplex pattern for that fragment) are detected. Alignments of each of the available sequences are available at* http://ubik.microbiol.washington.edu/HMA/hma.htm*. Reference A1 (92RW20) was used for the HMA analysis of some of these standards (see figures) but was not available at the time of the release of the current kit.*

***Key to contributors: A: Drs. Francine E. McCutchan and Joost Louwagie from the Henry M. Jackson Foundation and Walter Reed Army Institute for Research; B: Dr. Jay Levy from the University of California at San Francisco and Chiron Corp.; C: Dr. Saladin Osmanov and the UNAIDS Network for HIV Isolation and Characterization; D: Dr. James Bradac from the NIH-NIAID program on Genetic Variation and Dr. Beatrice Hahn from the University of Alabama, Birmingham; E: Drs. Helga Rübsamen-Waigmann and Manuel Grez from the Georg-Speyer Haus, Frankfurt; F: Drs. Marc Girard and Françoise Barré-Sinoussi from the Pasteur Institute, Paris; G: Drs. Alexsei Bobkov and Jonathan Weber from St. Mary's Hospital, London; H: Drs. Wouter Janssens and Guido van der Groen from the Institute for Tropical Medicine, Antwerpen.*

Primers

The primers included in this kit have been chosen because of their success in amplifying HIV-1 *env* sequences from each of the known subtypes (A–J have been tested) within the Main (M) group. They are not expected to function, however, with sequences of the Outlier (O) and the Non M Non O (N) groups[11, 18, 20]. Please note that all reference standard plasmids may not support amplification from the outer primers ED3/ED14, due to a shorter fragment having been cloned into the plasmid. A schematic representation of the primer positions is provided for reference (page 18).

First Round (also an alternate Second Round when using ED3–ED14 in the first round)

ED5	5'-ATGGGATCAAAGCCTAAAGCCATGTG (6556–6581*)
ED12	5'-AGTGCTTCCTGCTGCTCCCAAGAACCCAAG (7822–7792*)

These amplify a fragment of ~1.25 Kb spanning the V1–V5 coding region of gp120.

Alternate First Round

ED3	5'-TTAGGCATCTCCTATGGCAGGAAGAAGCGG (5956–5985*)
ED14	5'-TCTTGCCTGGAGCTGTTTGATGCCCCAGAC (7960–7931*)

These amplify a ~2.0 Kb fragment spanning from the first exon of rev to the transmembrane protein gp41 coding region of *env*.

Second Round (0.7 kb V3–V5 region)

ES7	5'-tgtaaaacgacggccagtCTGTTAAATGGCAGTCTAGC (7001–7020*)
ES8	5'-caggaaacagctatgaccCACTTCTCCAATTGTCCCTCA (7667–7647*)

Lower case letters are complementary to the universal M13-primer (ES7) or the reverse M13-primer (ES8). These allow direct sequencing of the PCR product.

Alternate Second Round (0.5 kb C2–C3 region)

ED31	5'-CCTCAGCCATTACACAGGCCTGTCCAAAG (6816–6844*)
ED33	5'-TTACAGTAGAAAAATTCCCCTC (7359–7380*)

Additional Alternate Primers

BH2	5'-CCTTGGTGGGTGCTACTCCTAATGGTTCA (7697–7725*)
	Substituting BH2 for ED12 results in a 68 bp smaller fragment
DR7	5'-TCAACTCAACTGCTGTTAAATGGCAGTCTAGC (6990–7021*)
DR8	5'-GTGAAGAGGTTAACAGGGAGTATAGAGGAGG (7668–7638*)

DR7 and DR8 are identical to ES7 and ES8, respectively, except that they do not include universal or reverse primer sequences at their 5' ends. These primers are used when amplified fragments are to be cloned rather than directly sequenced.

* Coordinates from the HIV-1_{HXB2} genome (Genbank accession no. K03455).

Positive Control DNA

pNL4-3 DNA, 1 copy/µl, in 0.02 µg/µl human placental DNA. Use 10 µl per reaction.

Reagents Not Included in the Kit

Some reagents are only required for specific procedures. Read this manual carefully to determine which of these reagents you may need.

Lysis buffer

10 mM Tris HCl, pH 8.3
50 mM Potassium chloride
2.5 mM Magnesium chloride
0.45% NP-40
0.45% Tween 20
100 µg/ml Proteinase K

RBC Lysis Buffer

320 mM Sucrose
10 mM Tris HCl pH 7.5
5 mM $MgCl_2$
1% Triton X100

10X PCR Reaction Buffer (no MgCL2)

500 mM KCL
100 mM Tris HCl, pH 8.3
10% DMSO
10% Glycerol

10X dNTP Mixture

2000 µM of each dATP, dCTP, dGTP, TTP (Pharmacia)

10X Heteroduplex Annealing Buffer

1000 mM NaCl
100 mM Tris HCl, pH 7.8
20 mM EDTA

10X TBE Gel Electrophoresis Buffer

880 mM Tris-borate, pH 8.0
890 mM Boric Acid
20 mM EDTA

One liter of 10X stock is prepared by dissolving in deionized H_2O: 108 g of Tris base, 55 g of Boric acid and 40 ml of 0.5 M EDTA.

5% Non-denaturing Polyacrylamide Gel

5% Polyacrylamide (using a stock of 30% Acrylamide, 0.8% Bis-acrylamide) in 1X TBE Gel Electrophoresis Buffer (dilute from 10X TBE Gel Electrophoresis Buffer above)

A 50 ml mixture is prepared per gel using 8.3 ml acrylamide stock, 5 ml 10X TBE and 36.7 ml H_2O. Polymerization is initiated by mixing in 50 mg of Ammonium persulfate and 33 µl of TEMED.

10X TAE Gel Electrophoresis Buffer

400 mM Tris-acetate
10 mM EDTA

One liter of 10X stock is prepared by dissolving in deionized H_2O: 48.8 g of Tris base, 11.42 ml of glacial acetic acid, 20 ml of 0.5M EDTA (pH 8.0).

1% Agarose Gel

1% Agarose in TAE buffer

5X Ficoll/Loading Dye

25% Ficoll
1% Orange G

Miscellaneous Reagents

10 mM $MgCl_2$
Taq polymerase
Ethidium bromide (0.5 µg/ml in H_2O)
Mineral Oil
Ampliwax (Perkin-Elmer, http://www.pebio.com/pc/catalog2/ch6index.html)
Isopropanol
Glycogen
Ethanol
Phosphate buffered saline (PBS)

Electrophoresis equipment

We use the GIBCO/BRL V16 vertical gel apparatus, along with the glass plates, spacers and the 20 well comb provided with the apparatus. The front glass plates are 19 cm h x 19.5 cm w, the back plate is 16 cm h x 19.5 cm w (both plates are 3 mm thick). Teflon spacers are 1.5 mm thick. 15, 20 or 25 well combs can also be purchased from BioRad.

DNA or RNA Preparation

PCR amplification product carry-over contamination is an extremely serious source of false or misleading results[10, 13]. Therefore, manipulation of infected blood should be performed using proper biosafety precautions and in an environment free of contaminating HIV *env* DNA. Preferably, the rooms in which DNA is extracted, as well as reagents, pipettors and other supplies, should be separated from rooms where postamplification products are handled. A multitude of techniques are suitable for the preparation of HIV DNA or RNA, and the only requirement is that these procedures not leave impurities that will substantially inhibit PCR amplification. HIV DNA can even be amplified directly from whole or fractionated blood, PBMC, or cultured cells by the addition of a small amount of infected cells directly to a PCR reaction in which the Taq polymerase is initially omitted. This reaction mix is subjected to repetitive heating and cooling cycles to lyse the cells. Taq polymerase is then added and amplification is performed as for purified DNA[15]. This process eliminates the need for lengthy DNA extractions. However, inhibitors present in red blood cells may inhibit amplification of HIV DNA. Thus, specimens prepared in this way that are HIV-PCR negative upon subsequent testing should be extracted to purify genomic DNA and re-tested before being judged negative.

Protocols for several methods follow. The *Puregene* protocol is recommended in situations in which supply costs are available for kit purchase and a cold chain can be maintained between blood collection and DNA extraction. The *Cell Lysate Method* is less expensive and thus is recommended when a cold chain can be maintained but supply funds are more limited. *The Dried Blood Spot* (DBS) method is recommended when a cold chain cannot be maintained after sample collection and prior to extraction.

A. DNA Purification Using the Puregene DNA Isolation Kit

This is the method of choice, since it is fast, results in highly pure DNA, and does not lead to significant sample loss. However, it requires purchase of an extraction kit. We currently use the kit from Gentra Systems, Inc. (Minneapolis, USA, phone: 800-866-3039 from the USA, or 612-476-5858, fax: 612-476-5850; http://www.gentra.com/).

1. Add 300 µl of whole blood to a 1.5 ml microcentrifuge tube containing 900 µl *RBC Lysis Solution*. Invert to mix, incubate 10 minutes at room temperature. Invert once during incubation.

2. Centrifuge 30 seconds in a microcentrifuge at full speed.

3. Remove supernatant with a micropipet, leaving the pellet and 10–20 µl of liquid behind.

4. Vortex vigorously to resuspend cells.

5. Add 600 µl *Cell Lysis Solution* to the tube, pipet up and down to lyse the cells. This and subsequent volumes corresponds to twice the volumes called for in the *Puregene* protocol.

6. Add 1.5 µl *RNase A Solution* and incubate at 37°C for 15 minutes. Samples are then stable at room temperature (18–22°C) for at least 18 months.

7. Cool sample to room temperature and add 200 µl *Protein Precipitation Solution* to the lysate. Vortex vigorously for 20 seconds to mix.

8. Microcentrifuge at full speed for 3 minutes; precipitated proteins will form a dark brown pellet.

9. Pour the supernatant into a fresh 1.5 ml tube containing 600 µl of 100% isopropanol, being careful to leave the precipitate behind. If <300,000 white blood cells are expected, add 1 µl of 20 mg/ml glycogen to enhance DNA precipitation.

10. Mix by gently inverting the tube 50 times to completely form a white fibrous clump of DNA.

11. Spin in a microcentrifuge at full speed for 10 minutes.

12. Pour off supernatant and drain tube on clean absorbent paper, taking care not to pour off the small white DNA pellet.

13. Add 600 µl cold 70% ethanol, and mix by inverting the tube several times.

14. Centrifuge at full speed then pour off supernatant (carefully! the pellet will be looser than after the first precipitation) and drain on absorbent paper.

15. Air dry for 15 minutes (do not over-dry in a vacuum dessicator).

16. Add 100 µl *DNA Hydration Buffer* and allow rehydration overnight at room temperature.

17. Store at +4°C to –20°C.

B. Extraction of DNA from Cell Lysates

Dr. Marcia L. Kalish from the Centers for Disease Control and Prevention, Atlanta, GA, USA provided this method. It is very fast and inexpensive, but can leave behind some inhibitors of the PCR reaction, and thus while samples are typically PCR positive, the amount of virus detected may be low.

Prepare lymphocytes by Ficoll-Hypaque separation. Cells can then be processed immediately or stored frozen at –70°C or in liquid nitrogen as a dry pellet for subsequent DNA preparation. If cells are frozen in DMSO to retain cell viability, wash twice with cold PBS before proceeding to the next step.

1. Add Lysis Buffer (page 7) to washed cells or frozen pellet and mix thoroughly by shaking. Final concentration should be ~ 6 x 10^6 cells/ml.

2. Incubate at 56°C for one hour.

3. Heat to 95°C for 10 minutes to inactivate the Proteinase K.

4. Use 10 µl of the preparation (~60,000 cells) for a 100 µl PCR reaction.

5 Store remainder at –70°C or in liquid nitrogen for future use.

- If the PCR reaction is negative, change the amount of lysate in a follow-up experiment. A positive could result with less lysate if the problem is an inhibitor, or it could require more lysate if the problem is a scarcity of viral genome templates in the reaction. We recommend the routine use of *Gene-Releaser* to remove inhibitors, added prior to the first round of PCR according to the procedures recommended by the manufacturer (BioVentures, Murfreesboro, TN; http://www.bioventures.com).

C. Extraction of DNA from Dried Blood Spots

Blood collection

1. Collect blood from one patient onto all four tips of a “IsoCode Stix” collection device (Schleicher & Schuell, http://www.s-and-s.com). Each collection device has four triangular tips that absorb blood, either from a finger puncture, or by dipping the tips into blood. Each tip holds about 10–12 µl of blood.

2. Dry samples for 15–20 minutes at 80°C, or place in a clean dry container with desiccant for a minimum of 3 hours (it is critical that the sample be dried for storage and subsequent use).

3. Place each sample device in its own storage bag along with desiccant, at room temperature (18–22°C) and protected from light. Dried blood spots can be stored for months to years this way.

DNA Extraction

4. Detach one triangular tip by placing over a sterile microcentrifuge tube and closing the lid. Store remainder at room temperature.

4. Rinse tip with 500 µl of distilled water, vortex 3 times for 5 seconds each.

6. Pellet tip in a microcentrifuge, remove the water with a pipette.

7. Repeat steps 2 and 3 twice.
8. Add 100 µl of distilled water to the tube containing the tip then incubate 95°C for 30 minutes. Vortex twice during incubation.
9. Pellet paper tip in a microcentrifuge, and then transfer the supernatant to a clean microcentrifuge tube.
10. Use 10 µl for PCR and store the remainder at –20°C.

D. Direct Amplification from Cultured Cells

1. Microfuge sample 1 minute to pellet cells (~10^6).
2. Remove supernatant leaving ~30–50 µl in the tube.
3. Resuspend the cell pellet in the remaining supernatant (results may be enhanced by washing the cells in an isotonic buffer such as PBS).
4. Prepare a 50 µl PCR reaction for each sample as described in step 1 of First Round PCR (page 12), omitting Taq polymerase.
5. Overlay each reaction with 30 µl of mineral oil or add an Ampliwax pellet.
6. Add 1–2 µl of the resuspended cells to the PCR reaction and cycle at 94°C for 3 minutes, followed by 55°C for 3 minutes.
7. Repeat heating/cooling three times.
8. Add 0.5 µl of Taq polymerase (5 U/µl) to each reaction.
9. Continue amplification as described below from Step 4 of the First Round PCR (page 12).

E. Direct Amplification from Whole Blood

Same as for cultured cells above except that 1–2 µl whole blood is added to 100 µl of First Round PCR reaction without Taq polymerase before heating and cooling. 1 µl is recommended as 2 µl may introduce more inhibitors, such as hemoglobin.

F. Amplification from Blood After Removal of Red Blood Cells (RBC)

The method described below is for 5 ml of whole blood collected in EDTA tubes. If the blood sample is 1 ml or less a microcentrifuge tube may be used with volumes adjusted accordingly.

1. Transfer blood to a 10 or 15 ml centrifuge tube and centrifuge at 2000 rpm for 10 minutes. White blood cells (WBC) will form a buffy coat layer on top of the RBC.
2. Remove plasma with a sealed plastic Pasteur pipette leaving ~0.5 cm of plasma above the buffy coat (plasma can be stored and processed for PCR by a variety of techniques not detailed here).
3. Carefully remove the buffy coat cells to a clean tube, taking as little of the RBC as possible.
4. Wash the WBC with 5 ml of PBS and centrifuge at 2000 rpm for 15 minutes.
5. Remove supernatant leaving a small volume (500 µl) with the cell pellet.
6. Lyse the remaining RBC by the addition of 4X volume (2 ml) of RBC Lysis Buffer (page 7) and mix gently.
7. Centrifuge at 2000 rpm for 10 minutes to pellet WBC.
8. Remove supernatant leaving ~50 µl with the cell pellet.
9. Resuspend cells in the residual volume (wash optional as in above whole blood procedure).

10. Add 2 µl of cell suspension directly to 100 µl First Round PCR reaction without Taq polymerase and proceed as described above for whole blood. Freeze remainder at –20°C or lower.

- If amplification is not initially successful, DNA can be extracted from stored cell pellets using the *Puregene Kit* or other method and the amplification repeated.
- If blood has been frozen the RBC will preferentially lyse when the sample is thawed. Following centrifugation only a few RBC will be present in the cell pellet, and thus, the lysis step may potentially be omitted. In this case, following a PBS rinse and centrifugation, resuspend in 50 µl PBS and proceed as for whole blood.

PCR

Nested PCR reactions are required for the preparation of sufficient quantities of HIV *env* DNA fragments from infected human cells to perform 10–20 HMA analyses. Reference fragments are amplified from 10 ng of plasmid DNA using the Second Round primers only (NOTE: not all reference plasmids contain annealing sites for the ED3–ED14 primers). Positive (10 copies of pNL4-3) and multiple negative control PCR reactions (using reaction mix alone, without template DNA) should be carried through both rounds of amplification.

First Round PCR

1. For a single 50 µl reaction only, mix, in this order:

5 µl	10X PCR Reaction Buffer (page 7)
6.25 µl	10 mM $MgCl_2$
28.25 µl	H_2O
5 µl	10X dNTP Mixture (page 7)
2 µl	ED5 (5 pmoles/µl)
2 µl	ED12 (5 pmoles/µl)
0.5 µl	Taq polymerase (2.5 U)

2. Overlay reaction mixture with an Ampliwax pellet (then heat and cool to re-solidify—see comment below) or 30 µl mineral oil (if recommended by the thermocycler manufacturer).

3. Add 0.1–2.0 µg infected cell DNA.

4. Amplification conditions:

   3 cycles of 94°C for 1 minute, 55°C for 1 minute and 72°C for 1 minute

   32 cycles of 94°C for 15 seconds, 55°C for 45 seconds, 72°C for 1 minute

   Final extension at 72°C for 5 minutes

5. When completed, store samples at –20°C.

- When setting up multiple reactions, prepare a “master mix” of reagents minus template DNA. The volume should be a little larger than the calculated amount (e.g., n + 1) to allow for losses during pipeting. The order of addition of reagents should also be considered and adhered to in each experiment to avoid contamination of reagents. Once the master mix is prepared and aliquoted into reaction tubes add 30 µl of mineral oil. Subsequently, add the DNA template to the tube below the mineral oil, cap the tube and begin thermal cycling.
- Alternatively a “hot start” procedure may be employed. Briefly, two master mixes are made called “lower” and “upper”. The lower buffer, sufficient for 30 µl of the final reaction volume, contains 10X PCR Reaction Buffer (page 7), $MgCl_2$, H_2O, dNTPs, and primers. The upper buffer, sufficient for the remaining 20 µl of each

reaction, contains 10X PCR Reaction Buffer (page 7) and Taq polymerase. The lower buffer is aliquoted to each tube, a 30 µl Ampliwax bead is added and the tube is capped, heated to 90°C for several minutes to melt the wax, and cooled to room temperature. The upper buffer is then distributed to the tubes (the upper buffer must be of sufficient volume to ensure that the phases will mix as the wax melts during the first heating step). Finally, the template is added and the thermocycling is begun. This "hot start" procedure reduces amplification of spurious products from complex templates. Please note that the $MgCl_2$ concentration will possibly need to be increased because some of it is absorbed in the wax. Using control templates, the amount of $MgCl_2$ which is needed to ensure robust product formation will need to be re-determined empirically, but will be more than that shown for an entirely aqueous PCR cocktail.

Second Round PCR

1. Prepare a 100 µl reaction per sample by mixing, in this order:

10 µl	10X PCR Reaction Buffer (page 7)
12.5 µl	10 mM $MgCl_2$
57 µl	H_2O
10 µl	10X dNTP Mixture (page 7)
4 µl	ES7 (5 pmoles/µl)
4 µl	ES8 (5 pmoles/µl)
0.5 µl	Taq polymerase (5 U/µl)

2. Overlay reaction mixture with 30 µl mineral oil or add an Ampliwax pellet as needed.

3. Add 2 µl of the First Round reaction, or 1 µl reference plasmid.

4. Amplification conditions:

   3 cycles of 94°C for 1 minute, 55°C for 1 minute, 72°C for 1 minute
   32 cycles of 94°C for 15 seconds, 55°C for 45 , 72°C for 1 minute
   Final extension at 72°C for 5 minutes

5. When completed, store samples at –20°C.

6. Check Second Round reactions for amplification product by loading 5 µl mixed with 1 µl 5X Ficoll/Loading Dye (page 8) on a 1% Agarose Gel (page 8) in 1X TBE Gel Electrophoresis Buffer (see 10X stock, page 7) or 1X TAE Gel Electrophoresis Buffer (see 10X stock, page 8). Electrophorese at 100 V for 1 hour. Stain the gel with ethidium bromide for 30–60 minutes and detect the DNA by UV transillumination.

- Agarose gel electrophoresis is conducted to ensure that the correct size DNA fragment was amplified in sufficient amounts (i.e. a bright band should be seen). Detection of multiple bands, especially ones with less than the expected size in agarose gels, is indicative of either the presence of a substantial population of molecules with large internal deletions or products derived from internal priming with Second Round primers. The presence of these molecules occasionally interferes with distance measurements due to the formation of heteroduplexes between DNA strands of substantially different size. Even though such molecules have greatly reduced mobilities, they have not, to date, prevented subtype determination when present as a minority component of the Second Round PCR. However, knowledge of the presence of such deletions is helpful in preventing erroneous conclusions. The majority of the amplified DNA seen on agarose gels may in some cases consist of the smaller fragments, since because of their size, they are preferentially amplified. To remove the smaller products, nested PCR is repeated using serially decreasing amounts of input genomic DNA until only the correct sized DNA fragment is amplified[6]. This dilution procedure can be considered successful if, as in all of the cases we have encountered to date, the full-length fragments represent the majority of the provirus population.

- In some instances, standard amplification conditions (55°C annealing) do not yield a product visible following agarose gel electrophoresis and ethidium bromide staining. In these instances, a *step-up amplification* is carried out with 5 initial cycles at lower annealing temperatures (from 52°C to 37°C) followed by 30 cycles under standard conditions.
- If samples are still negative after step-up amplification, another amplification can be attempted using the above conditions, this time using an alternative set of First and/or Second Round primers.
- Optimum $MgCl_2$ concentration for primer pair ES7/ES8 = 1.8 mM; ED31/ED33 = 1.4 mM; ED3/ED12 = 1.4 mM; ED3/ED14 = 1.25 mM.

Heteroduplex Mobility Assay in Polyacrylamide Gels

1. Assess the genetic diversity of each PCR-amplified sample. Mix, in a 500 µl Eppendorf PCR tube:

1.1 µl	10X Heteroduplex Annealing Buffer (page 7)
5 µl	H_2O
5 µl	Second Round PCR reaction (~100–250 ng of DNA)

2. Heat to 94–96°C for 2 minutes in thermocycler (or in boiling water bath).
3. Cool tubes rapidly by transferring to wet ice. Heteroduplexes can be kept at room temperature before loading or stored at 4°C for several minutes before loading the gel.

- This analysis provides a baseline heteroduplex pattern (resulting from quasispecies diversity in vivo) with which to compare deliberately formed heteroduplexes. It is CRITICAL and MOST useful if this sample is run on the same gel as the reactions between references and the unknown, so that bands present due to within-quasispecies heteroduplexes can be easily identified.

4. Form heteroduplexes using approximately equal amounts of amplified fragments from the unknown and each reference. Mix, in a 500 µl Eppendorf PCR tube:

1.1 µl	1.1 µl 10X Heteroduplex Annealing Buffer (page 7)
5 µl	Second Round PCR reaction from the analogous fragment of *env* derived from a reference strain
5 µl	Second Round PCR reaction from the unknown (~100–250 ng of DNA)

- Rapid cooling facilitates stable formation of heteroduplexes between highly divergent sequences and thus increases the yield of heteroduplexed molecules relative to the more stable homoduplexes. This seems to be less important for creating heteroduplexes between closely related sequences. You will note that the apparent heteroduplex yield will, in any case, be greater between more closely related sequences (see enclosed figures).

5. Mix heteroduplex reaction with 3 µl 5X Ficoll/Loading Dye (page 8) and load onto a 5% Non-denaturing Polyacrylamide Gel (page 7).

Gel Electrophoresis

General Considerations

The mobility of heteroduplexes is noticeably affected by changes in temperature during electrophoresis, particularly when the reannealed DNA strands are from highly divergent viral strains. The higher the temperature in the gel (i.e. the faster the gel is run) the slower the mobility of the heteroduplexes. Temperature increases result in local increases in duplex melting, thus slowing heteroduplex mobility. In order to compare data acquired across

experiments, it is therefore important to consistently reproduce the electrophoresis conditions as closely as possible. For this purpose the gel units, plates, acrylamide concentration, buffer, and standard voltage/current conditions should be carefully adhered to in each experiment. Given a particular apparatus and conditions, an equation can be derived to fairly reliably estimate the genetic distance between two DNA fragments from their heteroduplex mobility. Estimates of genetic distances can then be used to infer phylogenetic relationships between multiple sequences without analyzing all of the pairwise heteroduplexes possible[5]. To date we have typically evaluated 25–33% of the [N x (N–1)]/2 possible comparisons (where N = number of sequences being compared)[1].

- If the electrophoresis apparatus or conditions are changed, the previously determined standard curve is no longer valid and therefore must be reestablished using newly determined heteroduplex mobilities using DNA fragments from known sequences.

Electrophoresis Conditions

For a 16 cm x 19.5 cm x 1.5 mm gel we recommend the following electrophoresis conditions:

0.5 kb fragments, 250 V conditions	To compare 500 bp fragments (C2–V3–C3) produced by Second Round amplification with ED31 and ED33, we use a constant voltage of 250V for 2.5 hours. This region has demonstrated little length variation and thus heteroduplexes tend to migrate faster than ones formed with the fragments described below.
0.7 kb fragments, 250 V conditions	To compare 700 bp fragments (V3 to V5) produced by Second Round amplification with primers ES7 and ES8 or DR7 and DR8, we use constant 250 V for 3 hours.
1.2 kb fragments, 200 V conditions	To compare 1.2 kb fragments (generated by Second Round amplification with ED5–ED12) we use a constant 200 Volts for 6 hours. Overall, these conditions were initially found to be the most useful for deriving phylogenetic information although the PCR efficiency is generally lower than with smaller fragments.
1.2 kb fragments, 70 mA conditions	We initially used 70 mA constant current for 1000 Volts x Hours (usually about 3 hours)[5]. However, we have found that the relationship between heteroduplex mobility and DNA distance to be less predictable using these conditions when compared to the 200 V conditions described above.

Analysis of Heteroduplexes

The Problem of Inherent Quasispecies Complexity

When amplifying viral sequences from a plasmid or single provirus or RNA template, only homoduplexes are seen on the polyacrylamide gel. When amplifying sequences from a viral quasispecies containing multiple templates, heteroduplexes can form between different, simultaneously amplified variants within the mixtures and multiple heteroduplexes are seen on polyacrylamide gels. Heteroduplexes can take the form of sharp bands or of a smear-like pattern. The complexity of the heteroduplex pattern in a single sample can vary widely, with complex quasispecies usually seen as both homoduplexes and as heteroduplexes with reduced mobility. Thus, if a single variant or a collection of highly related variants is amplified from the unknown sample, a single homoduplex band is seen in the gel. When such products are reannealed with a reference sequence the two fast migrating homoduplexes (with indistinguishable or similar mobilities) and only two sharp heteroduplex bands are seen (occasionally migrating at the same position). When a complex quasispecies is reannealed with a reference sequence the heteroduplexes formed between the reference and the multiple variants can take the form of a series of bands or a diffuse smear. Occasionally then, difficulties in identifying the inter-strain heteroduplexes may be encountered. When this happens, we have found it useful to serially dilute the DNA preparation prior to nested PCR in order to generate products derived from a less complex mixture or a single variant. Assigning subtypes with less complex quasispecies is easier because of the simpler pattern of heteroduplexes formed with reference sequences. Quasispecies complexity is typically lower in in vitro cultured isolates than in their PBMC of origin[4]. Thus, if available, the use of co-cultured PBMC as a source of proviral DNA can decrease the problem of complex quasispecies.

Discreet single or multiple bands migrating with a mobility of about 40% of that of homoduplexes are sometime seen when examining PCR products. The position of these signals varies only slightly in location between different samples, as shown on the enclosed figures (page 19). These bands correspond to collapsed single stranded DNA fragments that failed to reanneal with a complementary strand due to unequal strand amplification during PCR arising from unequal primer concentrations. The uniform positioning of these bands makes them useful for visual comparisons of heteroduplex mobilities in different lanes; however, they can be prevented by carefully titrating the ratio of primers in control PCR reactions.

Subtype Determination

To assign an uncharacterized strain to a known subtype, the use of an equation which converts heteroduplex mobility into estimated DNA distance and subsequent phylogenetic analysis is not essential. Rather, the PCR fragment derived from the unknown strain is reannealed with the corresponding fragment from multiple representatives from each of the previously identified subtypes (or from any other sequence suspected to be related to the unknown strain). The heteroduplexes exhibiting the fastest mobilities between the unknown and the most related subtype indicate the likely subtype of that strain. In general, the closer the relationship between an unknown and a given reference, the fewer references that need to be compared to comfortably assign a subtype. See the Overview section of this document (page 3) for more discussion of subtyping strategy.

Phylogenetic Inferences

Some users of this kit with access to and familiarity with computational tools may wish to estimate phylogenetic relationships from HMA mobility data. The following section is relevant only to these efforts.

Measurement of Heteroduplex Mobilities

The relative mobility of heteroduplexes is typically determined from photographs or video-captured images of ethidium bromide stained polyacrylamide gels. The distance between the bottom of the loading well and the mid-point between the two heteroduplexes is measured and divided by a value corresponding to the distance between the bottom of the loading well and the midpoint between the two homoduplexes (the latter often migrate with the same mobility) as given below. Often, when the complexity of the unknown quasispecies is high, more than two heteroduplexes are formed with the reference sequence. In such instances the approximate midpoint between the most prominent heteroduplexes is used.

$$\text{Heteroduplex Mobility} = \frac{\text{distance from the well bottom to the midpoint of heteroduplex bands}}{\text{distance from the well bottom to the midpoint of the homoduplex bands}}$$

Comparison to Phylogenetic Measurements

Readers are referred to Delwart et al.[5] for a detailed description of the method used to relate evolutionary distance to heteroduplex mobility. Briefly, to calculate DNA distances between HIV variants, the sequences of each clone, bounded by the Second Round primers, were first aligned using the programs GENALIGN (Intelligenetics Inc., Mountain View, CA) and MASE[7]. DNADIST, from the PHYLIP software package[8] was then used to calculate distances by counting mismatches and weighting to adjust for base substitution frequencies, after removal of sites with insertions/deletions introduced to maintain alignment because there is no generally accepted means of weighting these gaps for phylogenetic measurements.This method was chosen to provide a comparison of HMA analysis with a commonly used method for sequence-based investigation of phylogenetic origins of HIV[16]. Despite the fact that gaps in annealed sequences have a substantial effect on heteroduplex mobility, a generally reliable relationship between these two measures is possible[5].

To calculate DNA distances from heteroduplex data, standard curves are generated for the relevant electrophoresis conditions by annealing samples of DNA with a range of sequence differences between them. Relative mobilities are measured as described above. Relative mobility values are then plotted against genetic distances, and the curve approximated by an exponential function.

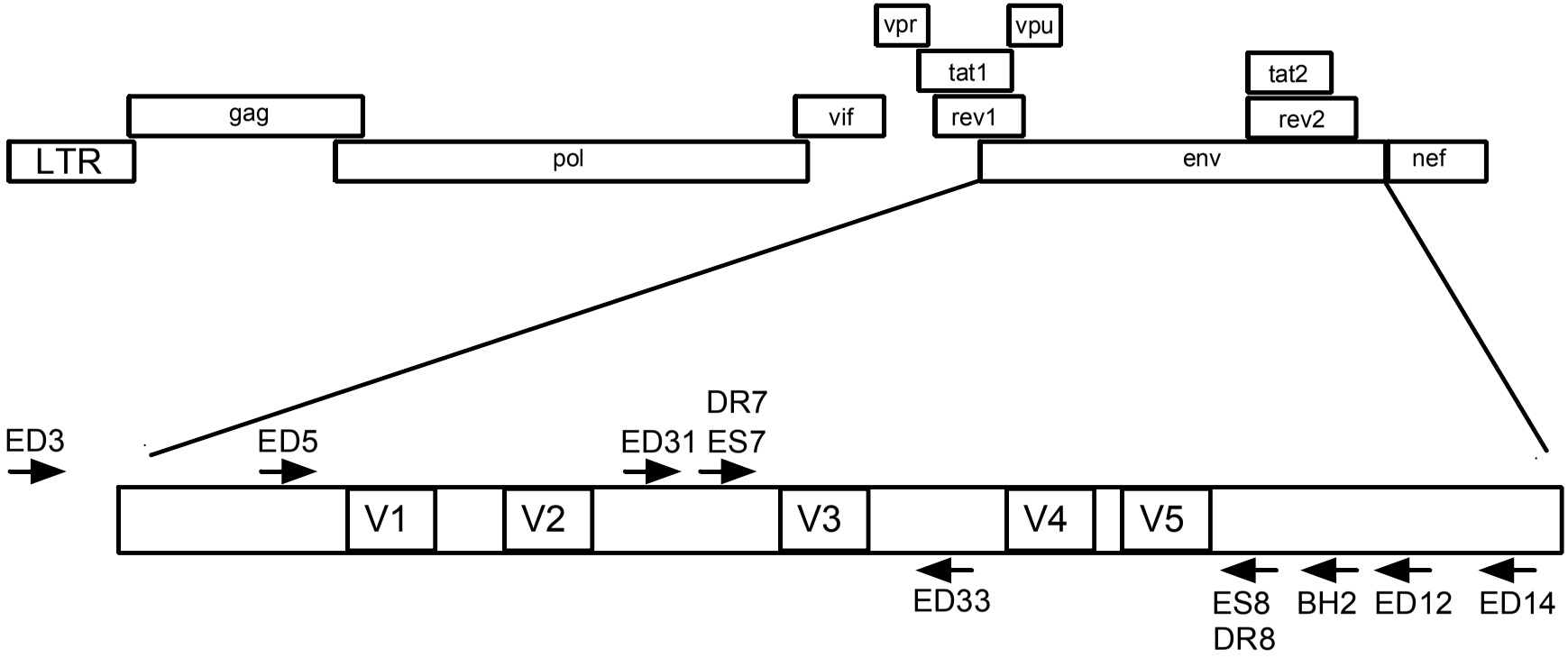
As noted above, various experimental conditions affect the mobility of the hetero- and homoduplexes. As a consequence, it is necessary to derive equations relating DNA distance to heteroduplex mobility that are unique for your own electrophoretic conditions. To facilitate these analyses, we include at our website (http://

ubik.microbiol.washington.edu/HMA/html/references.html) all reference sequences issued with this kit (for which sequences have been determined) bounded by three sets of Second Round primers.

Once a calibration curve has been obtained, a matrix of evolutionary distance between pairs of samples may be constructed. If comparisons between every pair of samples (including references) have been made, the matrix of genetic distances will be complete and can be used in any distance-based phylogenetic method, e.g., the neighbor-joining[17] and Fitch-Margoliash[9] methods. If, however, only certain pairwise comparisons are made, then the matrix of genetic distances will be incomplete and one is limited, at present, to the use of the Fitch-Margoliash least-squares method. This method has been implemented in the computer program FITCH, part of J. Felsenstein's PHYLIP software (J. Felsenstein, University of Washington, http://evolution.genetics.washington.edu/phylip.html). With FITCH, the user has the option of including missing values in the distance matrix. This is done by using the sub-replicate option. The documentation file that can be obtained from the same location describes how this may be done. FITCH writes the phylogenetic tree in parenthetical notation and places it in a file called *Treefile*. This tree description can be used in a number of different phylogenetic programs including, for example, PAUP 4.0[19] and MacClade 4.0[14] Tree view.

PHYLIP may be obtained at http://evolution.genetics.washington.edu/phylip.html. Users are required to register with J. Felsenstein (e-mail: joe@genetics.washington.edu). PHYLIP is available for Macintosh, IBM-compatible and UNIX computers.

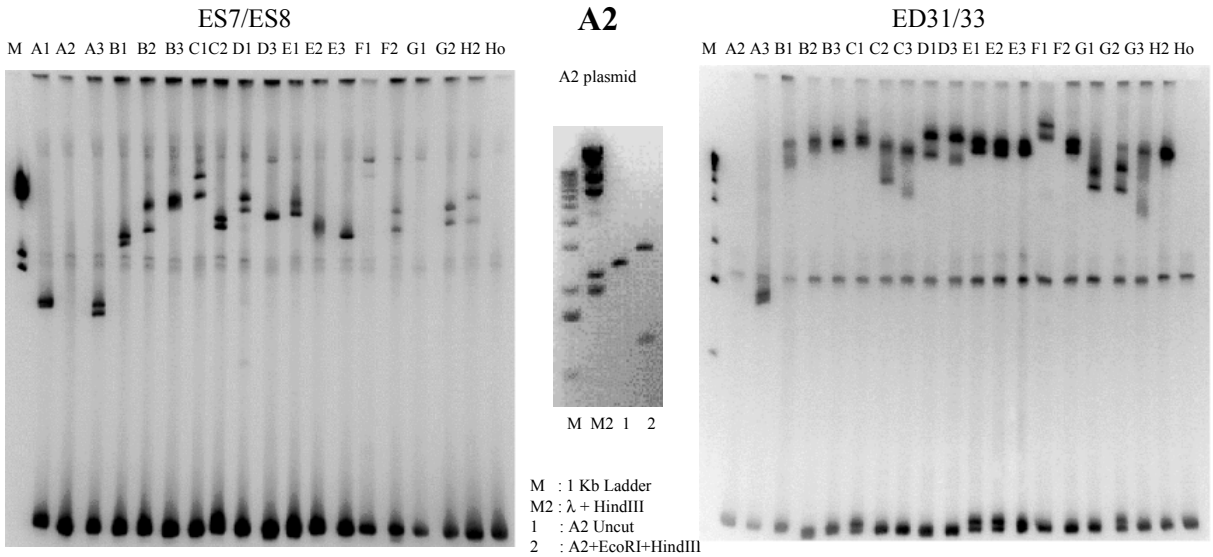
Schematic Representation of Primer Positions

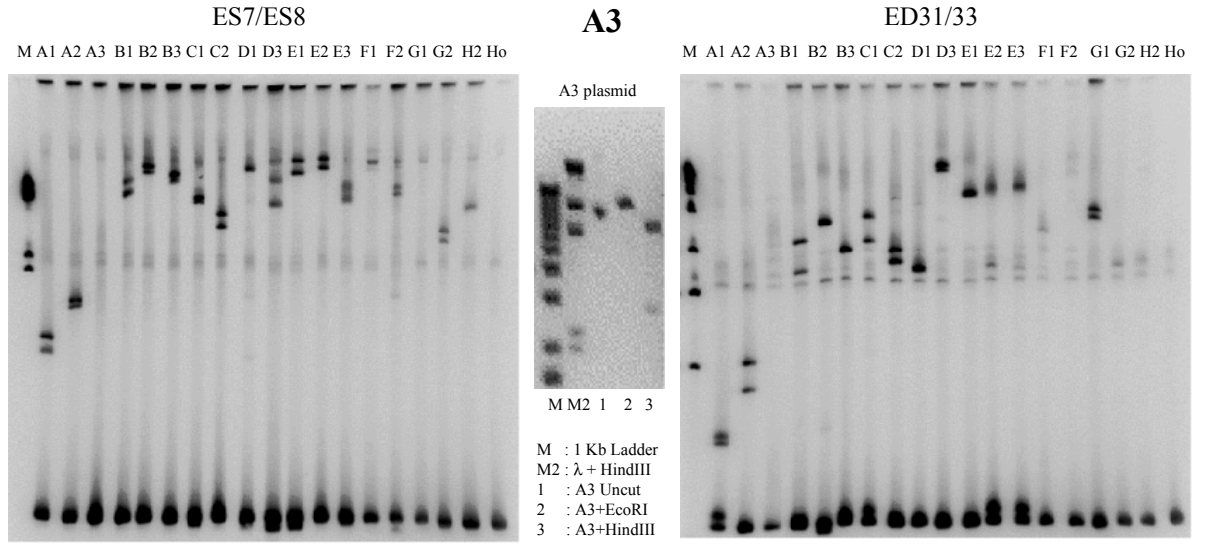


Genome organization of HIV-1 illustrating the position of protein coding regions. The envelope gene is magnified to show the relative position of the V1 through the V5 variable regions and the location of the PCR primers described in this manual.

Figures

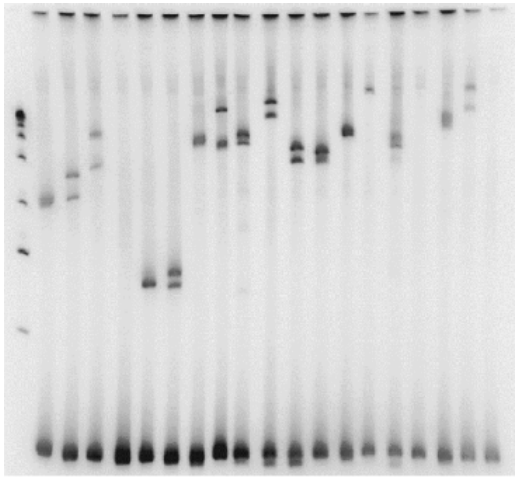
The following figures illustrate the mobility of heteroduplexes formed between one reference strain and each of others provided in the kit. The electrophoresis conditions used in these experiments are 5% polyacrylamide gel electrophoresed at 250 volts constant voltage for 3 hours (ES7–ES8) or 2.5 hours (ED31–ED33). Within each figure corresponding to different subtypes, the left panel illustrates the heteroduplexes formed between ES7–ES8 fragments and the right panel shows heteroduplexes formed between ED31–ED33 fragments. The letter/number at the top of each lane corresponds to the HIV-1 subtype for each of the fragments. This may be used for comparing with electrophoresis runs in your laboratory. The middle panel shows the undigested or restriction endonuclease-digested plasmid containing the envelope gene corresponding to the reference subtype used. This may be used to check the integrity or contamination of plasmids, if necessary. The molecular weight markers shown in these gels are kilobase molecular weight markers and lambda-DNA digested with *Hin*dIII (Gibco-BRL).





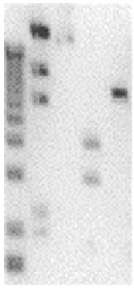
ES7/ES8

M A1 A2 A3 B1 B2 B3 C1 C2 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



B1

B1 plasmid

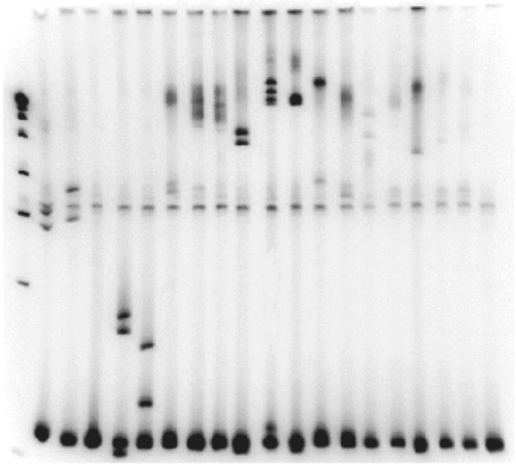


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : B1 Uncut
2 : B1+EcoRI
3 : B1+HindIII

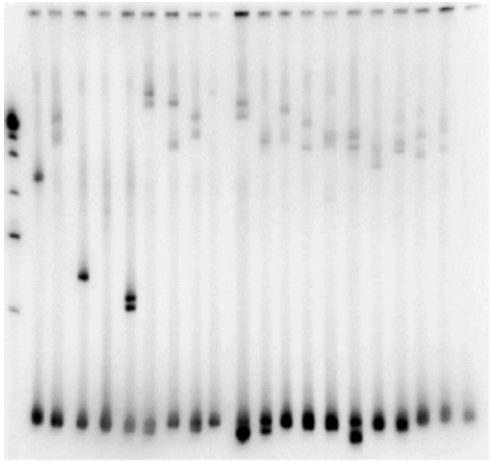
ED31/33

M A2 A3 B1 B2 B3 C1 C2 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



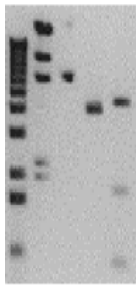
ES7/ES8

M A2A3 B1 B2 B3C1 C2 D1 D3 E1 E2 E3 F1 F2 G1G2 G3 H2 Ho



B2

B2 plasmid

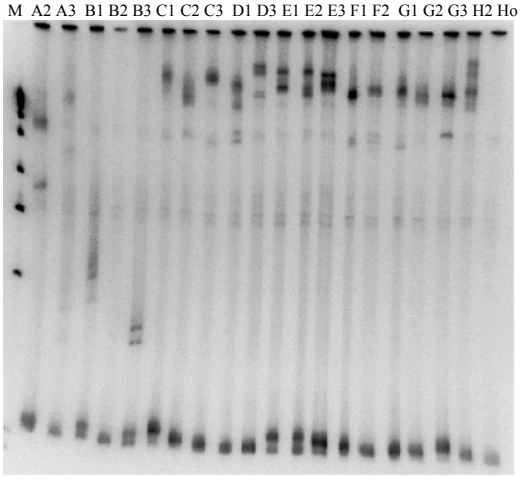


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : B2 Uncut
2 : B2+EcoRI
3 : B2+HindIII

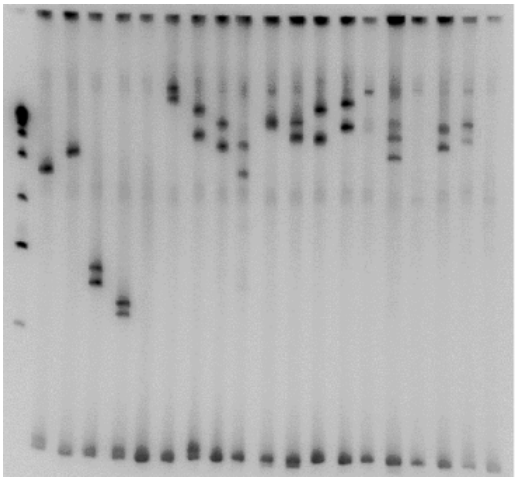
ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



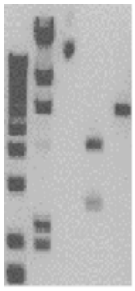
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



B3

B3 plasmid

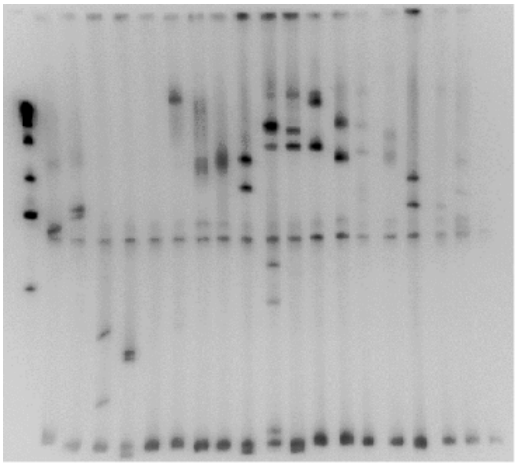


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : B3 Uncut
2 : B3+EcoRI
3 : B3+HindIII

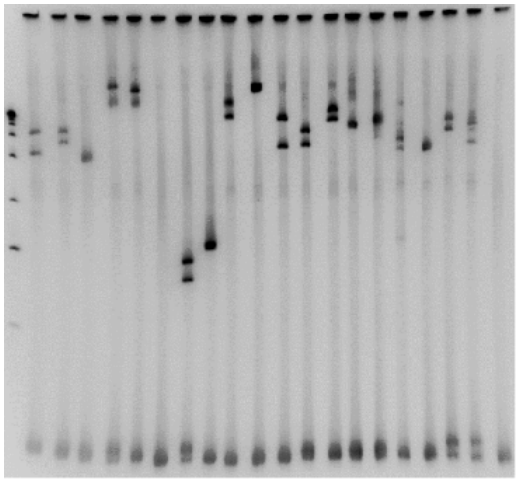
ED31/33

M A2 A3 B1 B2 B3 C1 C2 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



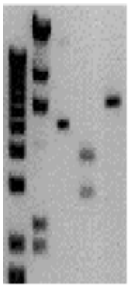
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



C1

C1 plasmid

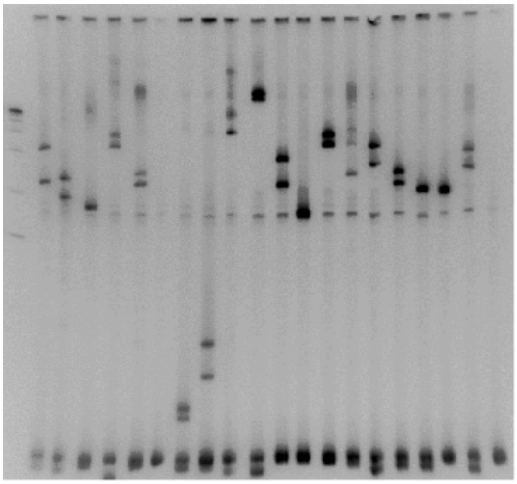


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : C1 Uncut
2 : C1+EcoRI
3 : C1+HindIII

ED31/33

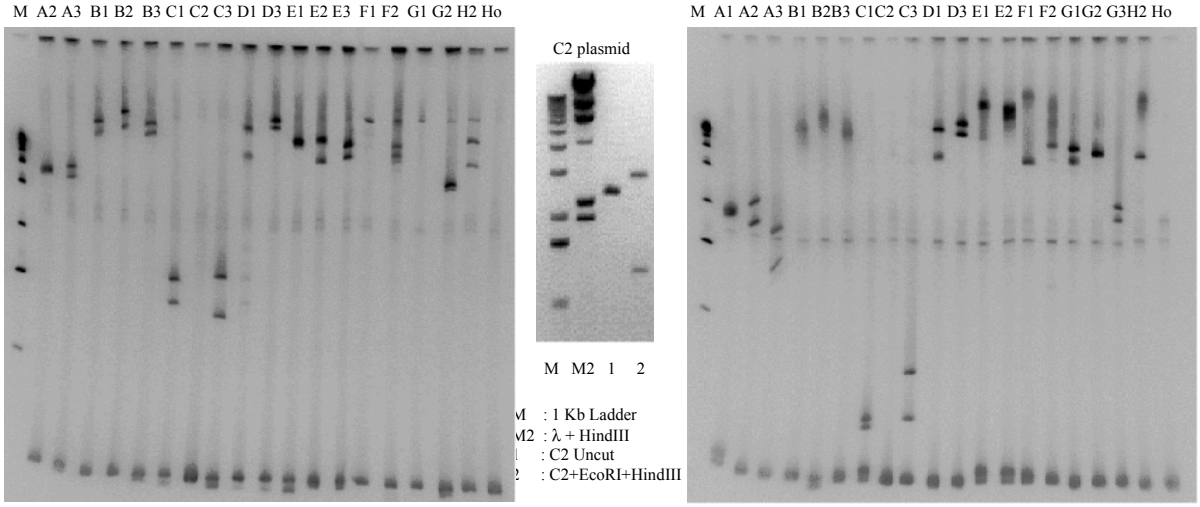
M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1G2 G3 H2 Ho



ES7/ES8

C2

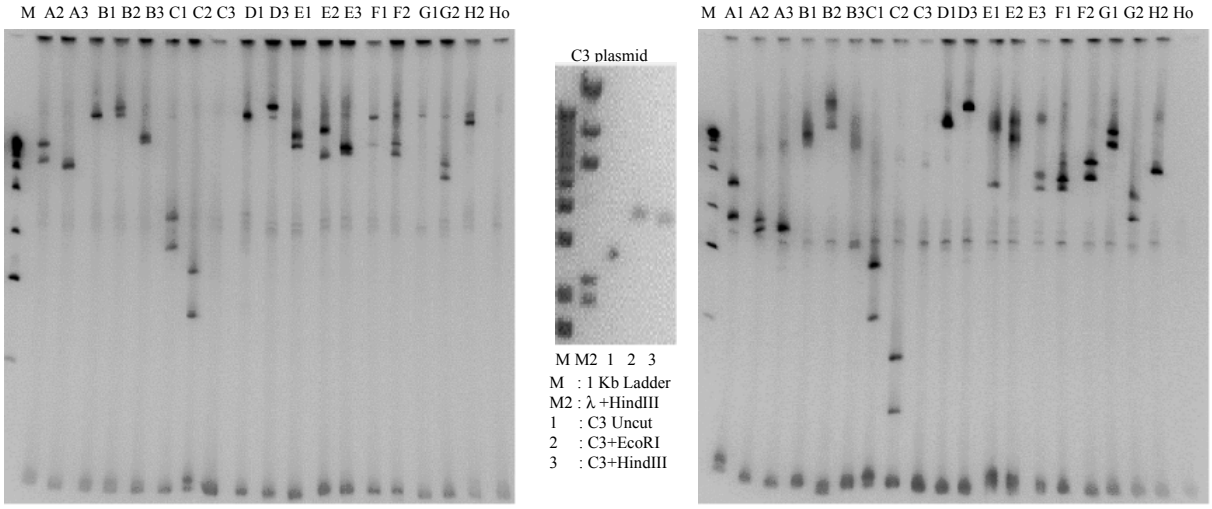
ED31/33



ES7/ES8

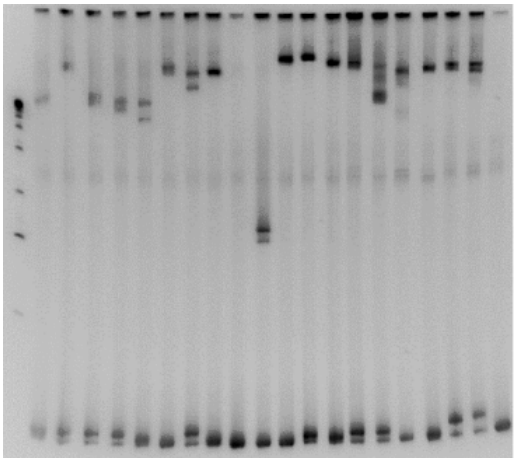
C3

ED31/33



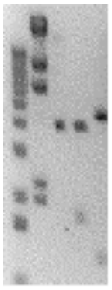
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



D1

D1 plasmid

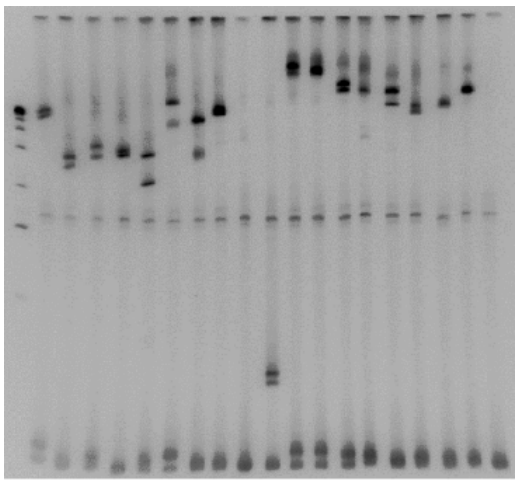


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ +HindIII
1 : D1 Uncut
2 : D1+EcoRI
3 : D1+HindIII

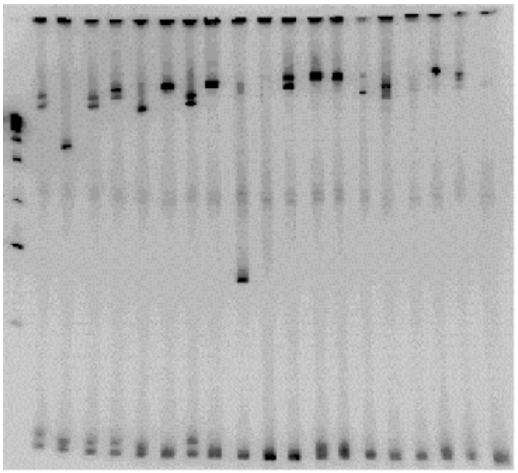
ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 F1 F2 G1 G2 G3 H2 Ho



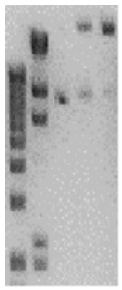
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



D3

D3 plasmid

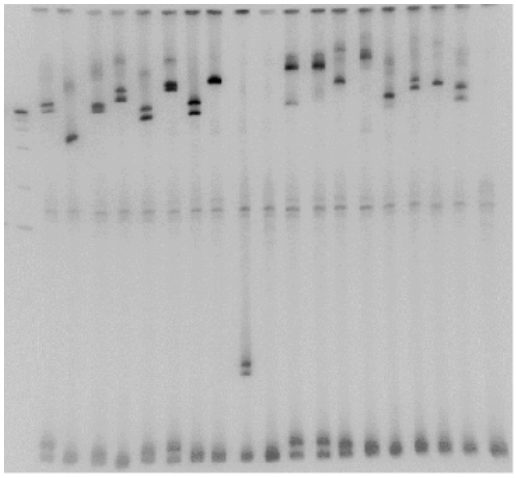


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : D3 Uncut
2 : D3+EcoRI
3 : D3+HindIII

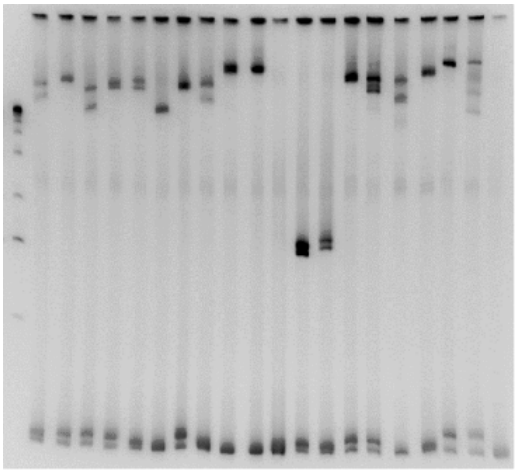
ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



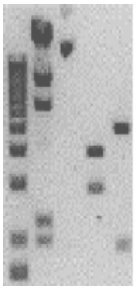
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 Ho



E1

E1 plasmid

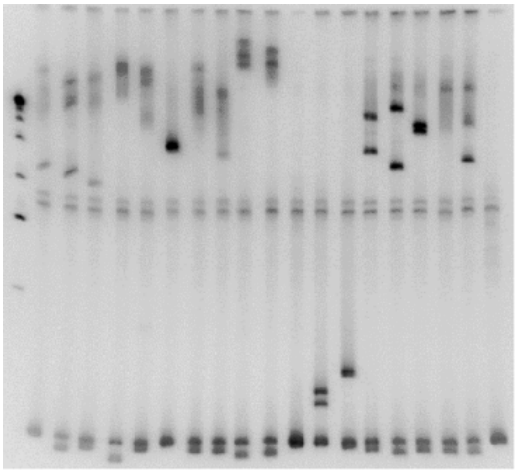


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : E1 Uncut
2 : E1+EcoRI
3 : E1+HindIII

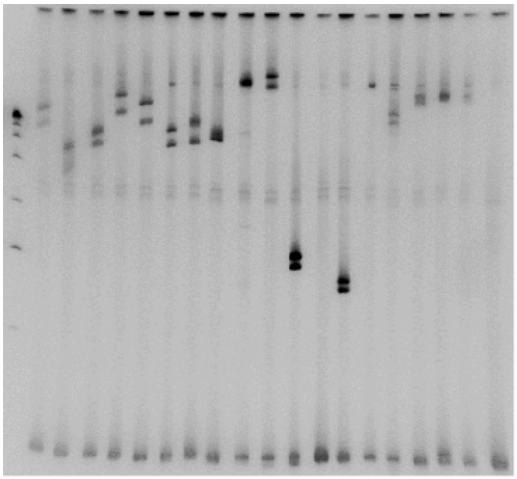
ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



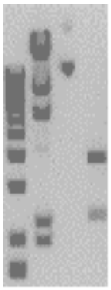
ES7/ES8

M A2 A3 B1 B2 B3 C1C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



E2

E2 plasmid

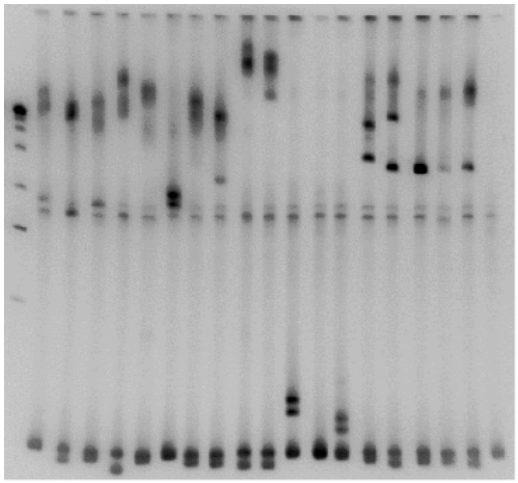


M M2 1 2

M : 1 Kb Ladder
M2 : λ + HindIII
1 : E2 Uncut
2 : E2+EcoRI

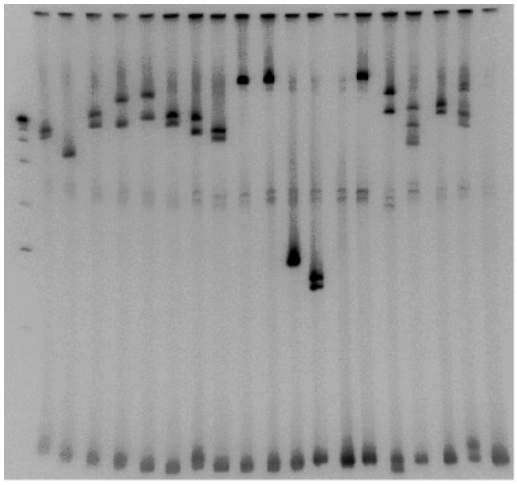
ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



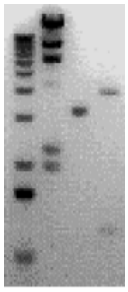
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



E3

E2 plasmid

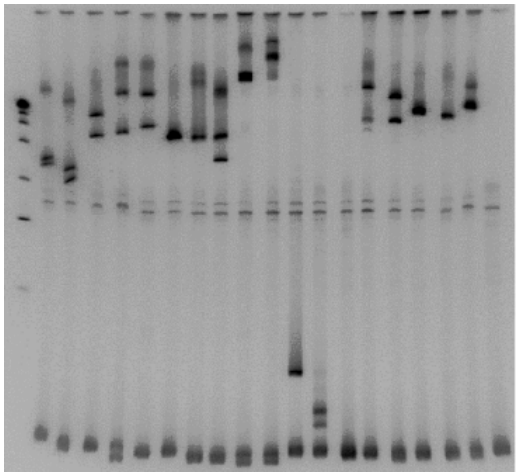


M M2 1 2

M : 1 Kb Ladder
M2 : λ + HindIII
1 : E2 Uncut
2 : E2+EcoRI+HindIII

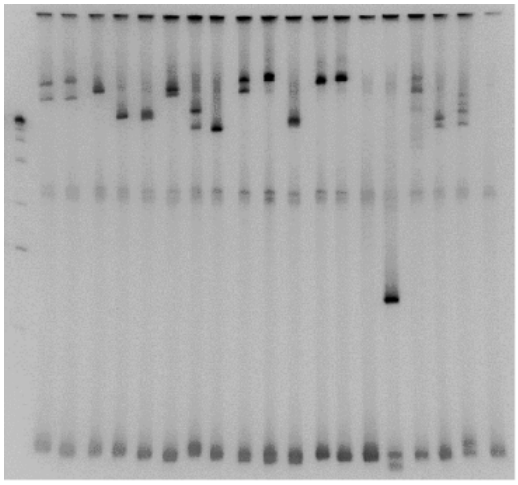
ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho



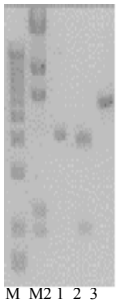
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1G2 H2 Ho



F1

F1 plasmid

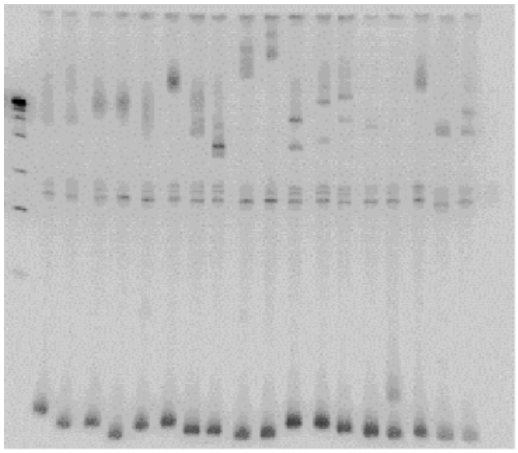


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ + HindIII
1 : F1 Uncut
2 : F1+EcoRI
3 : F1+HindIII

ED31/33

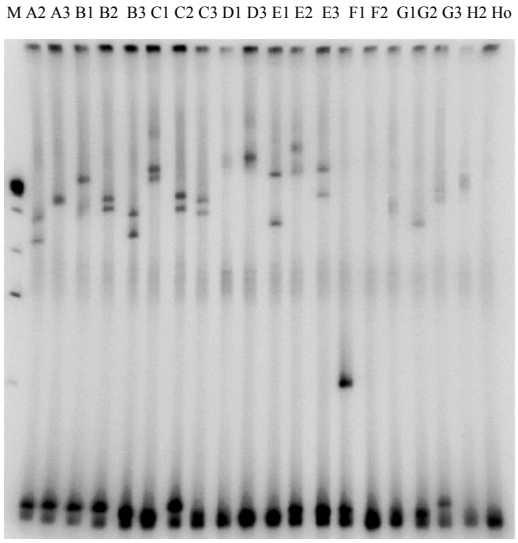
M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 H2 Ho

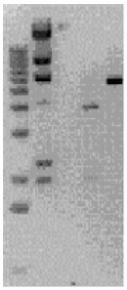


ES7/ES8

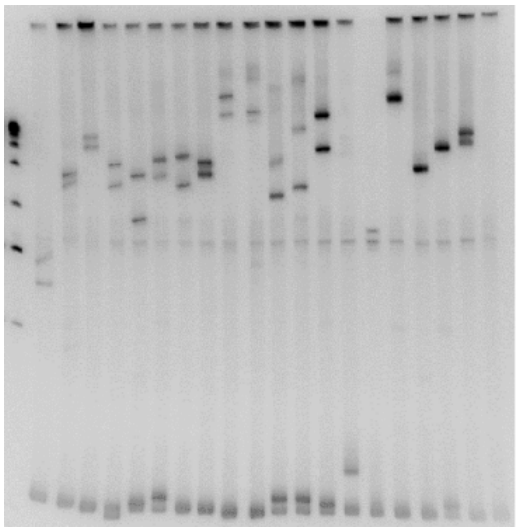
F2

ED31/33





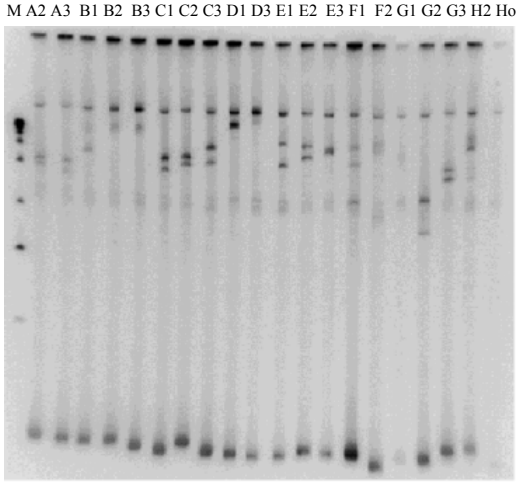
M : 1 Kb Ladder
M2 : λ + HindIII
1 : F2 Uncut
2 : F2+EcoRI
3 : F2+HindIII

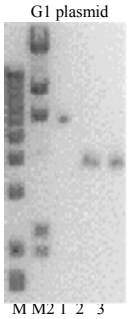


ES7/ES8

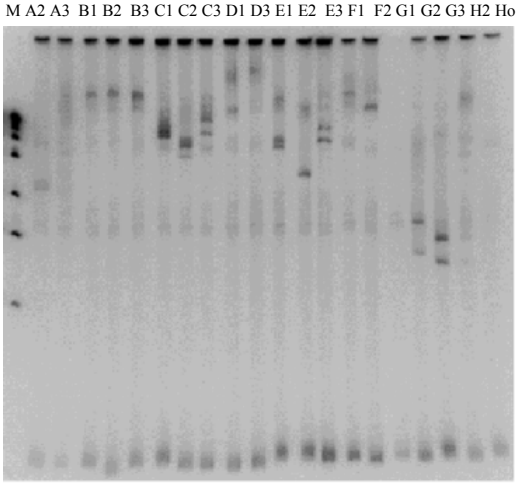
G1

ED31/33





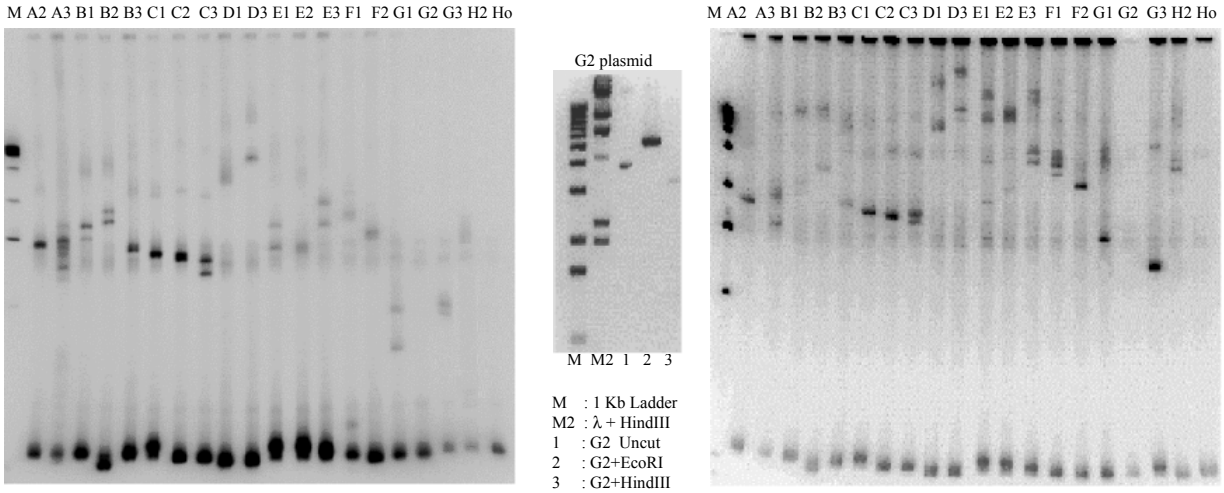
M : 1 Kb Ladder
M2 : λ + HindIII
1 : G1 Uncut
2 : G1+EcoRI
3 : G1+HindIII



ES7/ES8

G2

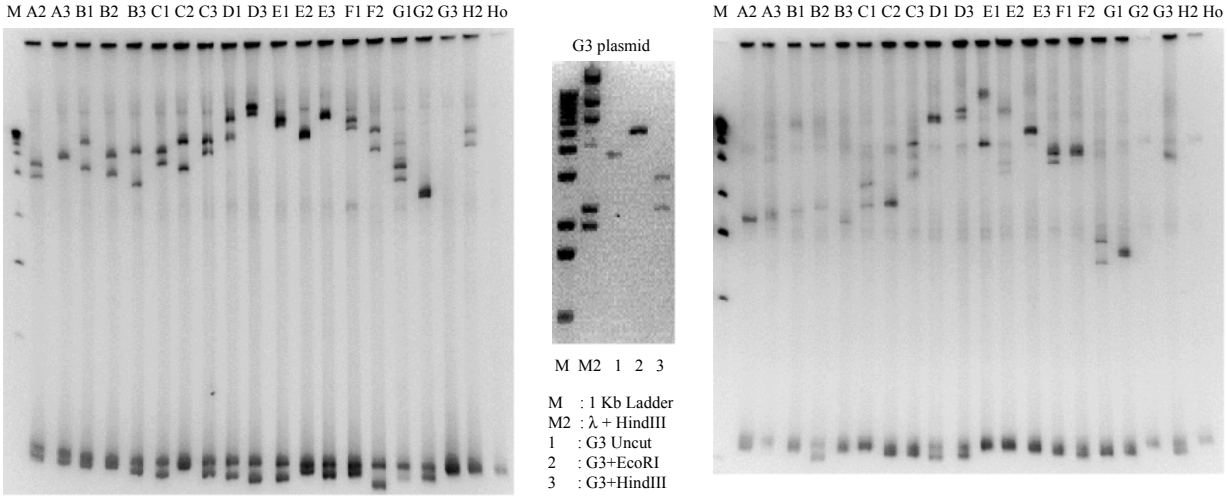
ED31/33



ES7/ES8

G3

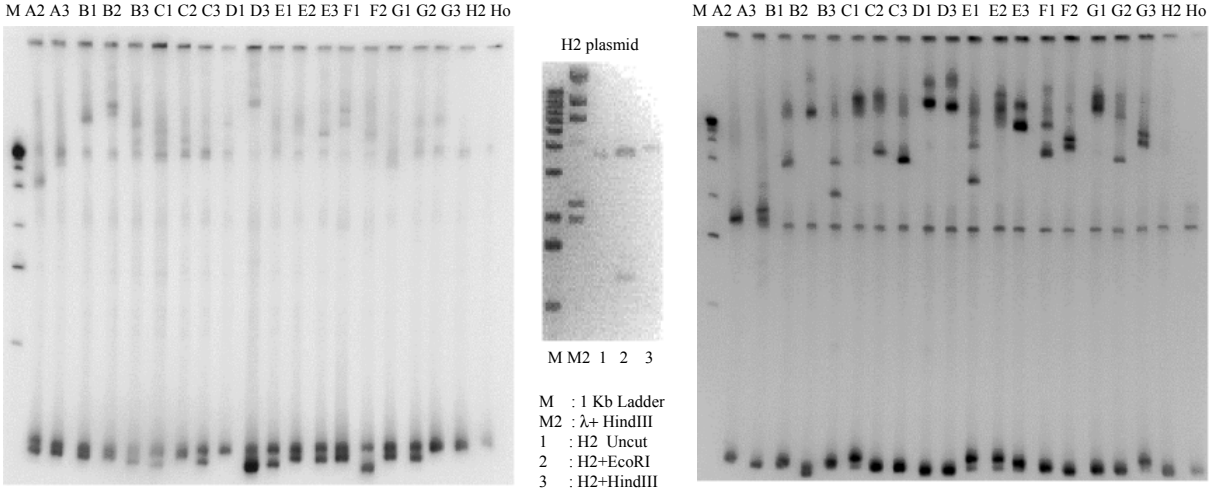
ED31/33



ES7/ES8

H2

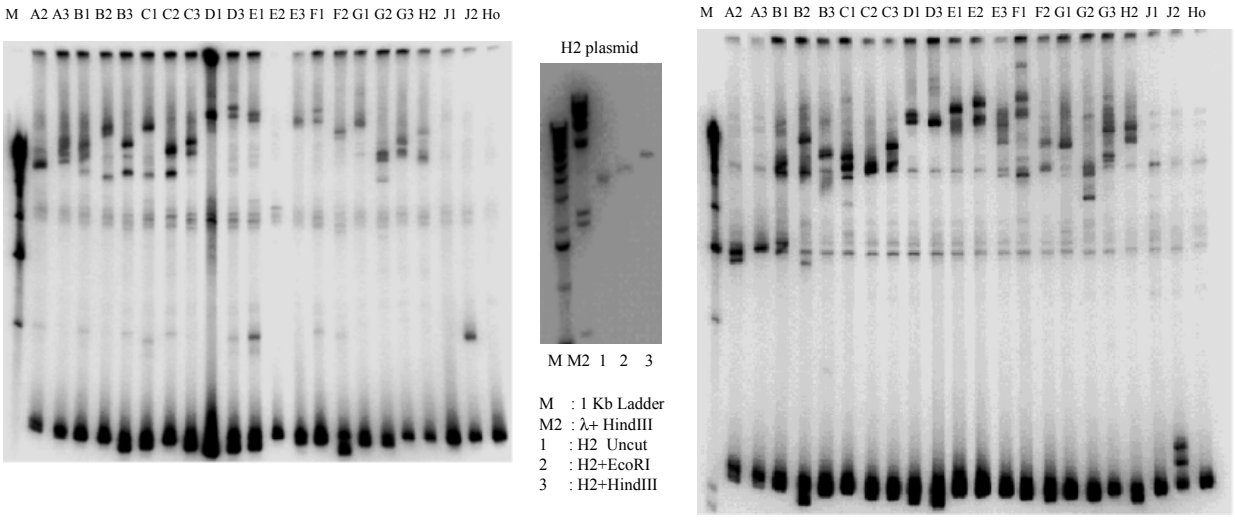
ED31/33



ES7/ES8

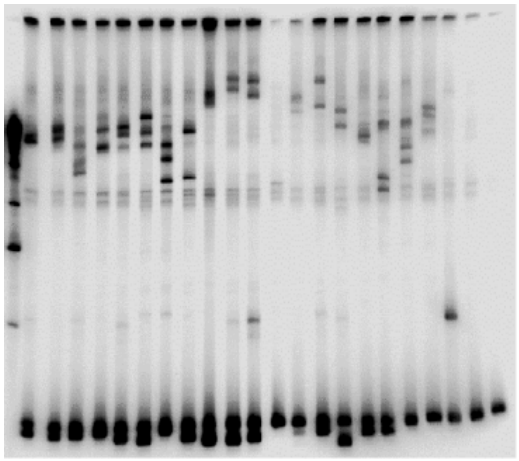
J1

ED31/33



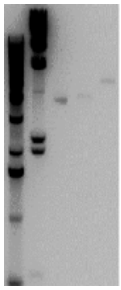
ES7/ES8

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 J1 J2 Ho



J2

H2 plasmid

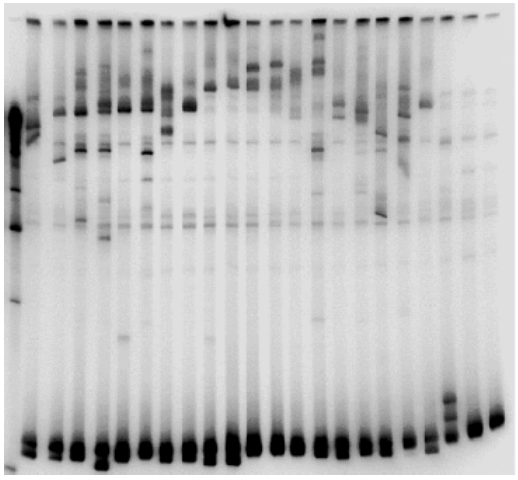


M M2 1 2 3

M : 1 Kb Ladder
M2 : λ+ HindIII
1 : H2 Uncut
2 : H2+EcoRI
3 : H2+HindIII

ED31/33

M A2 A3 B1 B2 B3 C1 C2 C3 D1 D3 E1 E2 E3 F1 F2 G1 G2 G3 H2 J1 J2 Ho



References

1. Bachmann MH, Delwart EL, Shpaer EG, Lingenfelter P, Singal R, Mullins JI, WHO Network on HIV Isolation and Characterization. Rapid genetic characterization of HIV-1 from four WHO-sponsored vaccine evaluation sites using a heteroduplex mobility assay. AIDS Res Hum Retroviruses 10:1343–1351, 1994.

2. Bobkov A, Cheingsong-Popov R, Garaev M, Rzhaninova A, Kaleebu P, Beddows S, Bachmann MH, Mullins JI, McCutchan F, Weber J. Identification of a new *env* G subtype and heterogeneity of HIV-1 strains in the former Soviet Union. AIDS 8:1649–1655. 1994.

3. Delwart EL, Busch MP, Kalish ML, Mosley JW, Mullins JI. Rapid molecular epidemiology of HIV transmission. AIDS Res Hum Retroviruses 11:1181–1193. 1995.

4. Delwart EL, Herring B, Rodrigo AG, Mullins JI. Genetic Subtyping of Human Immunodeficiency Virus Using a Heteroduplex Mobility Assay. PCR Methods Appl 4:S202–S216, 1995.

5. Delwart EL, Shpaer EG, McCutchan FE, Louwagie J, Grez M, Rübsamen-Waigmann H, Mullins JI.. Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 *env* genes. Science 262:1257–1261, 1993.

6. Edmonson PL, Mullins JI. Efficient amplification of half-genome sized fragments of human immunodeficiency virus from infected tissue samples. Nucl Acids Res 20:4933, 1992.

7. Faulkner DV, Jurka J. Multiple aligned sequence editor (MASE). TIBS 13:321–322, 1988.

8. Felsenstein J. PHYLIP–Phylogeny inference package. Cladistics 5:164–166, 1989.

9. Fitch WM, Margoliash E. Construction of phylogenetic trees. Science 155:279–284, 1967.

10. Frenkel L M, Mullins JI, Learn GH, Manns-Arcuino L, Herring BL, Kalish ML, Steketee RW, Thea DM, Nichols JE, Liu S-L, Harmache A, He X, Muthui D, Madan A, Hood L, Haase AT, Zupancic M, Staskus K, Wolinsky SM, Krogstad P, Zhao J, Chen I, Koup R, Ho DD, Korber BT, Apple R J, Coombs RW, Pahwa S, Roberts NJJ. Genetic evaluation of suspected cases of transient HIV-1 infection of infants. Science 280, 1073–1077, 1998.

11. Gürtler LG, Hauser PH, Eberle J, von Brunn A, Knapp S, Zekeng L, Tsague JM, Kaptue L. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J Virol 68:1581–1585, 1994.

12. Heyndrickx L, Janssens W, Zekeng L, Musonda R, Anagonou S, Van der Auwera G, Coppens S, Vereecken K, De Witte K, Van Rampelbergh R, Kahindo M, Morison L, McCutchan F E, Carr JK, Albert J, Essex M, Goudsmit J, Asjö B, Salminen M, Buvé A, Cities SG, Study Group on Heterogeneity of HIV Epidemics in African Cities, van der Groen G. Simplified Strategy for Detection of Recombinant Human Immunodeficiency Virus Type 1 Group M Isolates by gag/env Heteroduplex Mobility Assay. J Virol 74:363–70, 2000.

13. Learn GH, Korber BTM, Foley B, Hahn BH, Wolinsky SM, Mullins JI. Maintaining the integrity of HIV sequence databases. J Virol 70, 5720–5730, 1996.

14. Maddison WP, Maddison DR. MacClade 4.0. Sunderland, MA: Sinauer and Associates, 2000.

15. Mercier B, Gaucher C, Feugaes O, Mazurier C. Direct PCR from whole blood, without DNA extraction. Nucl Acids Res 18:5908, 1990.

16. Korber B, Kuiken C, Foley B, Hahn B, McCutchan F, Mellors JW, Sodroski J. Human Retroviruses and AIDS 1998. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences. In Human Retroviruses and AIDS (Los Alamos, NM 87545: Theoretical Biology and Biophysics, Los Alamos National Laboratory). 1998.

17. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425, 1987.

18. Simon F, Mauclere P, Roques P, Loussert-Ajaka I, Muller-Trutwin MC, Saragosti S, Georges-Courbot MC, Barre-Sinoussi F, Brun-Vezinet F. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med 4:1032–1037, 1998.

19. Swofford, DL PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, 2000.

20. Vanden Haeselvelde M, Decourt J-L, De Leys RJ, Vanderborght B, van der Groen G, van Heuverswijn H, Saman E. Genomic cloning and complete sequence analysis of a highly divergent African Human Immunodeficiency Virus Isolate. J Virol 68:1586–1596, 1994.

Alignments

```
          1         11        21        31        41        51        61        71        81        91       100
          >--------primer ED5-------->|         |         |         |         {V1       |         |        |
Consensus ATGGGATCAA AGCCTAAAGC CATGTGTAAA GTTAACCCCA CTCTGTGTTA CTTTAAATTG TACTAATGT- AAC------- ---------- ----------    72
SF170-A3  ......C..G ..T..... .......... .........T .....C.... .......... ..GCC..--- ...ATCACCA CC-------- ----------    79
BR20-B1   .......... .......... .T........ .......... .......... .......... C.T.G..T.G ..T------- ---------- ----------    73
TH14-B2   .......... .......... .......... A......... .......... .......C.. C...G..T.G .GG------- ---------- ----------    73
SF162-B3  .......... ..T..... .......... .......... .......... ..C..C.... C......T.G ..G------- ---------- ----------    73
MA959-C1  .......... .......... .......... ...G...... ........C. .......... ..GC....CT ...------- ---------- ----------    73
ZM18-C2   .......... ...T..... .......... ...G...... ........C. .....G.A.. .GGA.....T ..T------- ---------- ----------    73
IN868-C3  ........C. .....G.... .......... ...G...... ........C. .....G.A.. ..GAG.G..T ..T------- ---------- ----------    73
UG21-D1   .......... ........A. .......... A...C..... ........C. .......C.. C...G.ATGG ..G------- ---------- ----------    73
UG46-D3                                                                                                                    0
TH22-E1   .......... ..T..... .......... ......T..T .......... .......... ...C....CT ..TTTGATCA AT-------- ----------    82
TH06-E2   .......... ..T..... .......... ......T..T .....C.... .......... ...C....CT ..AGTGACCA AT-------- ----------    82
BZ162-F1  .......... .......... .......G.. .......... .......... .......... ....G...CC .T.------- ---------- ----------    73
BZ163-F2  .......... .......... .......G.. .......... .......... .......... ........CC .TT------- ---------- ----------    73
RU131-G1  .......G.. .......... .......... .........T .......... .......C.. ..GA.....A .G.TGTACTA ATTGTACTAA GGGAAAC---    97
LBV21-G2  .......G.. .......... .......... .........T .......... .C.....C.. .......--- ---------- -----GCAAT CGCTAATGTA    82
VI525-G3  .......G.. ...T...... ........G. .C.......T .......... .......C.. ......G..A .C.TACAATA GCACTGCAGC CAACAATAGA   100
VI557-H2                                                                                                                   0
SE9173-J1 .......G.. ...T...... .......... .A.......T ........C. .......... ..GCG.C--- ---------- ---------- -------GTC    70
SE9280-J2 .......G.. ...T...... .......... .A.......T ..A.....C. .......... ..G...C--- ---------- ---------- -------ATC    70

          101       111       121       131       141       151       161       171       181       191      200
          |         |         |         |         |         |         |      V1}          |         |        |
Consensus ---------A -T---ACTAA TA-TAATACT AGTAACGG-A A-A--ACTGA GAATGACACA GAAGAAATAA AAAACTGCTC TTTCAATGTA ACCACAGAAA   154
SF170-A3  ---------- -----CA... C.TC.CC..C .AC...CCC. .T.TC..CA. AT.CATGGA. .G........ .G........ ..A....A.G .........C   164
BR20-B1   ---------. A.AGT..... C.AC....A. ....GT..GG TA.AA..A.G ..TA....AG .G...G.... .......... .......ACC ......AGTG   164
TH14-B2   ---------. A.GCT..... ..CC.G..G. .CA------- ---------- -.TG..GGG. .G........ .......... .......A.C .....GAGC.   146
SF162-B3  ---------. A.GCT..... ..CC..G.G. .....T---- --TGG.AA.. ..TG....G. .G........ ....T..... ......G..C ......AGC.   158
MA959-C1  ---------- --------.. C.C.GC...C .A...T.TA. CTGCT..CA. T....TA..C AGT..C..G. ....T..T.. .........C. .........T   155
ZM18-C2   ---------- ---------- ---------- ------.TT. CCCATGAGA. T.GCACG.AG .GG.....G. ....T..... .........C. .........C   137
IN868-C3  ---------- --------G. .------... ..A..T..GG TT.ATGT.AC C...A.TGGG .......... ....T..... .........C. ..........   149
UG21-D1   ---------. A.GCC..C.C ..A.GCC... .A.G.G..C. TA-------- ---------- ---.G...G. .......... ....------ ---......G   134
UG46-D3                                                                                                                    0
TH22-E1   ---------- -----GTC.. ..AC.TA..C .A.GT.TCT. .C.TA.TA.G A...ATA... ..T...G... G......TA. ...T...A.G ..........   167
TH06-E2   ---------- -----GTC.G ..A..CA.TC .A.GC.CCTG .C.TA---.G A...ATA... ..T...G... G......T.. ...T...A.G .........C   164
BZ162-F1  ---------. A.AGC...-- -----..A.AC ..C.C.---- -----CTGAT .G.A..GT.. .GG.C....C .......T.. .......A.G .........G   149
BZ163-F2  ---------. A.GCC...GC C.T.C.AGAC G.GG..---- -----CTGA. .G.A.GAC.. .GG.C....C .......T.. .........G ....TG...G   155
RU131-G1  ---------- ---------- ---------- ------AAC. .T.GA....T .G.A.G..AG ..C.....G. .......... .......... ..T.......   161
LBV21-G2  ACCAACAGC. A.AAG....G ..AC..C.GC ..CCTT..G. .T.GC....T ..G.AG..T. ........G. .......... .......A.. .........G   182
VI525-G3  TCATCCACTG TCACC..CCC ..C..G.... GTA.C.AAC. .T.GC....T ..G..G..A. ........G. C......... .......... .........C   200
VI557-H2                                                                                                                   0
SE9173-J1 AATAGTAAC. A.AGT..AG. ..G.....G. ...GCAA.C. .C.AT.G.CC T--------- ...ATC..G. .......... .......... ..T.......   161
SE9280-J2 ACTAGTAAC. G.AAT..A.C ..G...C.G. ...GTAA.T- -----.G.CC T--------- ..TATC..G. C......... .......A.. ..T.......   155
```

```
          201        211        221        231        241        251        261        271        281        291      300
          |          |          |          |          |          |  {V2    |          |          |          |    V2} |
Consensus TAAGAGATAA GAAGCAGAAA GTATATGCAC TTTTTTATAA ACTTGATGTA GTACCAAT-- -------TGA TAATAATAA- -----TACTA GT-A-TATAG   237
SF170-A3  ....G..... ...A...... ..G...T... .......... .......... ........TG ATAAAAA.A. .GG......C AATATA.GG. C.C.G.....   264
BR20-B1   ...A...... .G..A....G .A......T. .C..C..... T......... ....A...-- -------A.G ....G.....- -----..... .C---.....   246
TH14-B2   ...A.AC... .GTA---... .AC....... .......... .......... ........-- -------A.. ....G....- -----...C. .C---.....   225
SF162-B3  .....A.... ..T....... .A........ .......... .......... ........-- -------A.. ....G....- -----...A. .C---....A   240
MA959-C1  .......... ..GA...... ..G....... .......... .......A.. ......C.TA ATGAGAAG.. ...C.GC.G- -----CT..G .AG.A.....   249
ZM18-C2   ...A...... A..A....G. ..G....... .......... .......A.. ......C.TA ATGAGAA.A. C..CTC..GT GAGGACT... ..G.G.....   237
IN868-C3  .........G ........C. ..G....... .........G .......A.. ......C.TG ATAATAAGA. G..G..C.-- -ACTC..G.G AGT.T.....   246
UG21-D1   .......... ....A..C.. .C........ .......... G........G ....A...GA ATG---A... .....G..C- ------..A.. CCAGT.....   225
UG46-D3                                                                                                              0
TH22-E1   .......... AC.......G ..CC...... .........G G......A.. ....A...-- -------GC. ....G.A..- -----..G.. ..G.G.....   252
TH06-E2   .......... .........G ..CC...... .......... G......A.. ....A.---- ---------- ----GT...- -----..G.. ..G.G..C..   240
BZ162-F1  .......... .C..A...G. ...C...... .......... .......A.. ........-- -------CA. .GGC.G..G- -----A..AT CA---..C..   231
BZ163-F2  .......... .C........ ...C...... .G...-...G .......A.. ........TA GCAATGGAAG A...GGC.G- ------.GGAG ACT.C..C..   248
RU131-G1  .......... ...AA...C. .A...C..G. ....C..... .......... ........TA ------A... .G.A.G..G- -----..A.. ..ACT.....   249
LBV21-G2  .......... ..GTA..... .A...C..T. ....C....G .........G ........TA ATGGTAA... A....G..G- -----...A. C.G.T.....   276
VI525-G3  .......... .C..A..GC. .A...C..G. ....C..... .......... ........-- -------... ........G- -----A.A.G A.AGT.....   285
VI557-H2                                                                                                             0
SE9173-J1 .....A.... A.G.A..C.. .A...C..G. ....C....G ..AA...... ........-- -------.A. ..G.G.C..- -----..AA. ..---....T   243
SE9280-J2 .....A.... A.G.A..C.. .AG..C..G. ....C....G ..AA...... ........-- -------... ..G...C..- -----..AA. A.---....T   237
```

```
          301        311        321        331        341        351        361        371        381        391      400
          |          |      >---------primer ED31---------->     |          |          |          |          |        |
Consensus GTTAATAAAT TGTAATACCT CAGCCATTAC ACAGGCTTGT CCAAAGGTAT CTTTTGATCC AATTCCCATA CATTATTGTG CTCCAGCTGG TTTTGCGATT   337
SF170-A3  A......... .......... .......... .......... .....A.... .C.....G.. .......... .......... .......... ..........   364
BR20-B1   A..G.C..G. .......... ...T...... C.....C... .........A .C.....G.. .......... .........A .C..G..... ..A.......   346
TH14-B2   ...G...... .....C.... ...T...... ...A...... .......... .C.....G.. .......... .........A .C..G..... ..........   325
SF162-B3  A..G...... .....C.... ...T...... ......C... .......... .C.....A.. .......... .......... .C..G..... ..........   340
MA959-C1  A......... ..C.G..... ..A.TG.A.. ...A..C... ........C. .......C.. ......T... .........A .......... ..A.......   349
ZM18-C2   A......... .......... .......A.. ...A..C... ........CA ....G..... ......T... .......... .......... A.A.......   337
IN868-C3  A......... .......... .......A.. ...A..C... .......CCA .......... ......T... ..C....... .......... ..........   346
UG21-D1   A......... ......G... ..A....... ......G... ......A... .C.....G.. .......... .......... .C........ A.....A...   325
UG46-D3                                                                                                              0
TH22-E1   .......... ........T. ...T.....A G......... ......A... .C........ ......T... .........A .......... ..A.......   352
TH06-E2   .......... ........T. ...T.....A G......... ......A... .C........ ......T... .......... .......... ..A.......   340
BZ162-F1  .C........ .......... ..A....... .......... .......... ...GG..... T......... .......... .......... ..A.A.....   331
BZ163-F2  .C........ ..C....... ..A....... .......... .....A.... ...GG.G... .......... .......... .......... ..A.......   348
RU131-G1  .C........ ......GT.. ..A......A .......... ........GA A......C.. .......... .......... .......... ..........   349
LBV21-G2  .C........ ......G... ..A......A ...A...... ........GA .......C.. .......... ........C. .......... G.........   376
VI525-G3  .C.....C.. ......GT.. ..A......A .......... .C........ .......A.. .......... .......... .......... ......A...   385
VI557-H2                                                                                                             0
SE9173-J1 TC.......C .......... ...TT....A .......... ........G. .....C.A.. .......... .......... .......... ..........   343
SE9280-J2 TC.......C .......... ...T.....A ...A...... ........G. .....C.A.. .......... .......... .C........ ..........   337
```

```
          401        411        421        431        441        451        461        471        481        491      500
          |          |          |          |          |          |          |          |          |    >---primer ES7—con't
Consensus CTAAAGTGTA ATGATAAGAA GTTCAATGGA ACAGGACCAT GTAAAAATGT CAGCACAGTA CAATGTACAC ATGGAATTAA GCCAGTGGTA TCAACTCAAC   437
SF170-A3  .......... .....GCAG. .......... .....G.... .C..G..... .......... .....C.... .......C.G ......AA.. .....C....   464
BR20-B1   .......... ...G...... .......... ........C. ...C...... .......... .......... .......... ......A..G ..C.......   446
TH14-B2   ...C...... .CA....A.. .......... .......... ...C...... ...T...... .......... .........G ......A...  ..........   425
SF162-B3  .......... .......... .......... T......... ...C...... .......... .......... .........G ......A..G .........T   440
MA959-C1  .......... ..A......C A........G .......... .CC.T..... .......... ..G....... .......... .......... ..........   449
ZM18-C2   .......... ..A......C A........G .......... .CC.T..... .......... .......... .C.....C.. .......... ..........   437
IN868-C3  A......... ..A......C A........G .......... .C.GT..... .......... .......... .......... .......... ..........   446
UG21-D1   .....A.... .......A.. .........G .....T.... .......... .......... ..G....... ....G..... ......A..G ..........   425
UG46-D3                                                                                                                    0
TH22-E1   T......... ....C..... T..T.C...G .....G.... .......... ....T..... .....C.... .......... .......... .........T   452
TH06-E2   T......... .......A.. C........G .....G.... .......... ....T..... .....C.... .......... .......... .........T   440
BZ162-F1  .......... .G.....A.. .........G .....G.... .CG.G..... .......... .......... .......... A......... .........T   431
BZ163-F2  T......... .......A.. .........G .....G.... .C..G..... .GC-...... ..G....... .......... A......... .........T   447
RU131-G1  T......... GG........ T....C.... .......... .......... ...T...... .......... .......... .......... .........T   449
LBV21-G2  ....CA.... GG........ .......... ........G. .......... ...T.....G .......... .......... .......... .........T   476
VI525-G3  T......... GG..G..... T....C.... .......T.. .......... ...T...... .......... .......... .......... ..........   485
VI557-H2                                                                                                                   0
SE9173-J1 .....A.... ..A......C T......... .......... .C........ ...T...... .......... .C........ .......... ..........   443
SE9280-J2 .....A.... .......... T......... ......T... .C........ ...T...... .......... .C........ .......... ..........   437

          501        511        521        531        541        551        561        571        581        591      600
          ---primer ES7 cont --->          |          |          |          |          |          |          |        |
Consensus TGCTGTTAAA TGGCAGTCTA GCAGAAGAAG AGATAATAAT TAGATCTGAA AATCTCACAA ACAATGCCAA AACCATAATA GTACAGCTTA ATGAATCTGT   537
SF170-A3  .......... .......... .......G.A G.G...AG.. .......... ...A...... .......... .......... .....A.... .CA.GA....   564
BR20-B1   .......... .......... .......... .T...G.... .......... .......... .......T.. .......... ..G.....A. .G..CC....   546
TH14-B2   .......... .......T.. .......... ..G..G.... ........AGC ...T....GG .......T.G .GT....... ........G. ..........   525
SF162-B3  .......... .......... .......... G.G..G.... .......... ...T.....G .......T.. ...T...... ........G. .G........   540
MA959-C1  .A........ ...T..C... .......G.. .A........ .......... .....G.... ......T... ...A...... ..G..C.... ..........   549
ZM18-C2   .A........ ...T..C... .......... .......... ...G...... .....A.... .......... ...A...... .....T.... ..........   537
IN868-C3  .A......G. ...T..C... .......... G......... .......... .....G.... .......... ..TA...... .....T...G ..C..C....   546
UG21-D1   ..T....G.. .......... .......... .......... .......... .......... .T.....T.. ..TT...... .....T.... ....G.....   525
UG46-D3                                     ........ .......... .......... .T.....T.. .......... .......... ....G.....    68
TH22-E1   .......... .......... .......... .......... C......... .......... .......... .......... ..G..C.... ..A.......   552
TH06-E2   .......... .......... .......... .......... C......... .......... .......... .......... ..GT.C.... ..A.......   540
BZ162-F1  ..T....... ......C... .........A GT...G.... C......C.. ...A..T..G .T.....A.. ..T....... .....C.... ..........   531
BZ163-F2  ..T....... ......C... .......... .T........ C......C.. ...A..T..G .T.....A.. .......... .....CT... ..........   547
RU131-G1  .AT..C.G.. ......CT.. .......... .A......G. .......... ..CT.....G .......... .GT....... ..G....... ..A.......   549
LBV21-G2  .A...C.G.. .......T.. .......... .A.....G.. .......... ..CT...... .......... ..A....... ..G..AT... ..A.....A.   576
VI525-G3  .A...C.G.. .......T.. ......A... .A..T..... .......... ..CT...... .......... .GT....... ..G....... .....A....   585
VI557-H2                              A.CC .TG....... .......A.. ...A.....G .....A.... ..A....... .......... .AAGTC....    74
SE9173-J1 .A...C.... .......G.. .......G.. .C........ .......... ...A..T..G .......T.. ..A....... .....A.... ....CA....   543
SE9280-J2 .A...C.... .......A.. .......G.. .C........ .......... ..CA..T..G .......... ..A....... .....A.... ..A..A....   537
```

```
          601        611        621        631        641        651        661        671        681        691      700
          |          |          | {V3      |          |          |          |          |          |     V3}  |        |
Consensus AGAAATTAAT TGTACAAGAC CCAACAACAA TACAAGAAAA AGTATACATA TAGGACCAGG ACAAGCATTC TATGCAACAG GAGACATAAT AGGAGATATA   637
SF170-A3  .......... .....C.... .T........ .......... ...G...G.. .......... .......... .......... .T........ ...G......   664
BR20-B1   ...C...... .......... .......... .........G .......... .......... GAG......T .......... .......... ..........   646
TH14-B2   .......... .......... .......... .......... .........C .......... G.G.....GG ...A...... ..C.A..... ..........   625
SF162-B3  .......... .......... .T.....T.. .......... ......AC.. .......G.. GAG......T .......... .......... ..........   640
MA959-C1  .......GTG .......... .......T.. .........G. ......AGA. .......... .....T.... .......ATA AT........ ......C...   649
ZM18-C2   .......GTG .......... ...G...T.. .......... ......AGG. .......... .......... .......... .T.G...... ....A.C...   637
IN868-C3  .......GTG C......... .A.....T.. .......... ......AGG. .......... ....A..... .......... .......... ......C...   646
UG21-D1   .CC....... ..C.....G. ..T..G.T.. GGT...CT.C ..G.C..C.. .....GT... G.G.....C. ...A.....A .---A....A ..........   622
UG46-D3   .CC....... ..CT....G. ..T..G.A.. ..A....CG. C...C..C.. ......T... G.....G.A. ...A.....A A---AT...A ....T.....   165
TH22-E1   ......C... .....C.... ..TC...... ........C. ......AC.. .......... .....T.... ...AG..... .......... ..........   652
TH06-E2   ......C... .....C.... ..TC...... ........C. ......AC.. .......... .....T.... ...AG..... .......... ..........   640
BZ162-F1  .C........ .......... .......... .......... .......... .......... ..GG...C.T .......... .......... ......C..C   631
BZ163-F2  .C.G...... .......... .......... .......... G......... .......... ..G......T .......... .......... ......C..C   647
RU131-G1  ........CA .....C.... .T..T..... .......... ......AC.T .T.C...... ......GC.. .......... .T..A..... ..........   649
LBV21-G2  ...T...GTG .....C.... .......... .......... ........C. .......T.. .........A .......... .C.C...... ...C......   676
VI525-G3  .......... .....C.... .......... .......... ......A.AT .C...A.... ..G..TG..A .......... .T.C...... ....A.....   685
VI557-H2  .CC......C .....C.... .......T.. ...G...... ......AG.. ....G..... .......... .......... .T.....C.. ..........   174
SE9173-J1 .......GTG .....C.... .T..T..... .......... G.......C. .G........ .....TGC.. ..C....... .G..A..... ..........   643
SE9280-J2 .......GTG ...TAC.... .T..T..... ......G... G.......C. .G........ .....TGC.. ..C....... ....A..... ....A.....   637

          701        711        721        731        741        751        761        771        781        791      800
          |          |          |          |          |          |          |          |          |          |        |
Consensus AGACAAGCAC ATTGTAACAT TAGTAGAACA AAATGGAATA AAACTTTACA AAAGGTAGCT AAAAAATTAA AAGAACACTT TAATAATAAA ACAATAATCT   737
SF170-A3  .........T .......TG. C......G.. G.C...... .......... .GG......C ..CC...... ..AGTT.... ..G...C... ..........   764
BR20-B1   .......... ........C. ...C...G.. C........G .C......AG C...A...T. .C........ G......A.. .G.G...... .......AG.   746
TH14-B2   .......... ........C. ......C... .......... .C......AG .C..A..A.. G......... G......A.. .GGG..C... ......G...   725
SF162-B3  .......... .......... ....G..GA. .......... .C......A. .C..A...T. .C.......C ...C...A.. .GGG...... ......G...   740
MA959-C1  .......... ....C..... .....AGGAT GTC.....C. GT........ ........G. .......... .......... CCC....... .......CA.   749
ZM18-C2   .......... .......... .....A.GAG ..C....... .......... ........GA .........G C...G..... CCC....... .......AA.   737
IN868-C3  ....G....T .......... ....GA..AC .........G .......... ..G...GAG. .........G C....T.... CCC....... ......CAA.   746
UG21-D1   .......... .......... ....G..GA. .......... .......... .C........ GCG....... G...C.TTC. ...CC.A.C. G.......T.   722
UG46-D3   ....C..... .......T.. ....G..G.. G......... .......... .C........ .........G G...C.TT.. C..CC.G.C. ........T.   265
TH22-E1   ...AG....T .....G.G.. ..A.G..... .......... ..G.....A. .C.....A.. G......... ....G..... ....---..G ..........   749
TH06-E2   .........T .....G.G.. ..A.GA.... .........G ........A. .C.....A.. .......... ....G..... .........G ..........   740
BZ162-F1  ...A.G.... ........G. .......... C......... .G.AG...G. CC.....AAG GC.G.G.... .GTCT..T.. CCC....GC. ......CAA.   731
BZ163-F2  ...A.G..C. .......TG. ....G..... C......... ....A...G. .C.A...AGG GC...G.... .GTCT..T.. .CC....GC. .......AA.   747
RU131-G1  .......... .......TG. .......A. G.T......G .G.TG..G.. G..T..GA.. .C....C... ...G.AT... ...C..C.C. .AC....CA.   749
LBV21-G2  .......... .......TG. ....GA.... G.C....GAG .C.TG..... G......AAA GC.C..C..C ...G.AT..A C..C..A.GC ---.....C..   773
VI525-G3  .......... .......TG. .....C.GA. ......GG.. .G.TG..... G..T...... GC....C..C ..A..G.A.. ......A--- .AC....C..   782
VI557-H2  ...A.....T .......T.. ..C...GGA. G.T.....A. GG........ TG......T. C.GC...... G......... .......C.. ........T.   274
SE9173-J1 ..GA.....T .......... ........A. G.T....... .C.......G T.GA...... ......C... G......... ......A.C. ---...GA..   740
SE9280-J2 ...G..A... .......... ....GA..G. G.T....G.. .C.......G T.GA...... .C....C... G......... ......A.C. ---.....A..   734
```

```
          801        811        821        831        841        851        861        871        881        891      900
          |          |          |          |          |          |   <-----primer ED33------<     |          | {V4    |
Consensus TTAAACCATC CTCAGGAGGG GACCTAGAAA TTACAACACA TAGTTTTAAT TGTAGAGGAG AATTTTTCTA TTGCAATACA TCAGAACTGT TTAATAGTAC   837
SF170-A3  ..GCTAGC.. .......... ..TT...... .......... .......... ...G...... .......T.. ...TG..... ....GC.... .......C..   864
BR20-B1   ..C.G...C. .......... ..T.C..... ..GT.TTT.. C......... ...G....G. .......... C..T...... A..C...... ..........   846
TH14-B2   ....T.A... .......... ....C...G. ..GT..TG.. C......... ...G....G. .......... C..T...... A..C...... ..........   825
SF162-B3  ....G.A... .......... ....C..... ..GT..TG.. C......... ...G....G. .......... C..T...T.. A..C.G..T. ..........   840
MA959-C1  ..G.....CA .........A ..T....... .......... .......... .......... .......... ....GG.... .........-   848
ZM18-C2   ..G.C.A.CA .......... .......... .......... ...C...... .......... .......... ...A.C.... ..........   837
IN868-C3  ..GC.T.... .......... .......... .......... ...C...... .......... .......... ...A.C.... .....G...-   845
UG21-D1   .......... ...G...... ....C..... .......... C..C...... ...G....G. .......... C......... ....G..... ....C.A..G   822
UG46-D3   ..C.....-- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   273
TH22-E1   ..C.....C. .........A ..T....... ......TG.. .CA....... ........G. .......... .......... A..CG..... ...G..A...   849
TH06-E2   ..C.C...C. .........A ..T....... ...T...G.. .CA....... ........G. .......... .......... C..A..T... .....G.A..   840
BZ162-F1  ....CT.... .......... .........G ......TG.. .CA.....T. ....A..... .......... C...G..... ...AG..... .C.G..AC..   831
BZ163-F2  ....TT.... .......... .......... ......TG.. .......... .......... .......... C......... ....G..... .---------   838
RU131-G1  A...CT..G. TG.......A ..T....... .......... .......... .......... .......... .......... ....G..... ......A..G   849
LBV21-G2  ....CT.... .......... .......... .....G.... .......... .......... .......... ...T...... ....G..... .........G   873
VI525-G3  ....CT.... .......... .......... .......... .......... .......... .......T.. ...T..G... .....G.... ..........   882
VI557-H2  .CG....... .......... ...A.G.... ......TG.. ..C....... .......... .......... .......... ...A..T... ....C.....   374
SE9173-J1 ...C.T..C. .......... ...A...... .......... .......... ...G...... .......... ...T...... ...AC..... .C.......G   840
SE9280-J2 ...C.T..C. .......... ..TA...... ..GTG..... .......... ...G...... .......A.. ...T...... ...A...... .C.......G   834

          901        911        921        931        941        951        961        971        981        991      1000
          |          |          |          |          |          |          |    V4}     |          |          |        |
Consensus TTGG----AT AATAA---TA ATACGACAAA TAATACA--- ---AATGGCA C--------- TATCACACTC CCATGCAGAA TAAAACAAAT TGTAAACATG   915
SF170-A3  A...GGGGGA ....GCAC.G .C.GC.T.C. GG.GT..AAT AGC.CA.AGT .AAATGACAC ...A.TT... .......... ....G..... .A....T...   964
BR20-B1   C...ACAA.. ..C.C---.G .AGG.....G ...C..TACA GGG....A.. .---------- A......... .....T.... .......... ..........   934
TH14-B2   ....AATG.. .C..G---.. C.TG.---.. .....ATACA GGA....... .---------- ......G... ..T....... .......... ..........   910
SF162-B3  ....A---.. ....C---.. TAGG.C.... ...C..T--- ---.....A. .---------- .......... .......... .......... .A......G.   919
MA959-C1  --------.C TT..----.G ....AGA.GG ....T..--- ---.C.TTA. G--------- C..A..C... .......... .......... .A........   922
ZM18-C2   A.ATA---.G CC..---.G ....A.AT.G ..CATATAAT CCT....AT. .---------- A......... .......... .......... .A........   922
IN868-C3  -----CAT.C .TG..---.G G...A.A..G ....T..--- ---.GCTCA. .---------- C......A.. .T........ ........G. ..........   922
UG21-D1   .GTATGGACA TCA..---.. G.----..G.T AGGGG..--- ---.....G. .---------- A......... .......... .......... .A....T...   901
UG46-D3   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   273
TH22-E1   ...CATAA.A ...G.---A. CC.T.G.GGG G---TGT--- ----.C.... .---------- .....T...T .......AG. ....G..... .A........   928
TH06-E2   CAT.GAGGGG CC..---.. .C...GAGGG GCC..AT--- ----.CACT. .---------- .....T...T .......AG. ....G..... ..........   922
BZ162-F1  ---------- ---------- -AGAAT.C.. .GGC...GAA TCC....... .---------- ......T... .....TC... .......... G.........   901
BZ163-F2  ---------- ---------- -------T.. CG.C...GGA TCC....... .---------- ......T... .T...TC... .......... G.........   902
RU131-G1  ----------. G.G..---.. .CT.C..T.T CG.G.AT--- ----..CTC.. .---------- .......... .....T..G. .......... .A...TA...   922
LBV21-G2  .ATATTATC. ..C..---.. ..G.AC..TC .....GC--- ---.G.AAG. A--------- .......... .....T.... .......... .....GA...   955
VI525-G3  ....TTGTT. ....GTAC.G ---G.CTGGT .....GTACT TGG....AT. .ACGAAAAGA ...T.....G .A.....A.. .......... ...G.GA...   979
VI557-H2  ....G---T. ..C.G---.. CAT.A.ATG. ..TCT..AAT GGC.CAACA. T--------- A..TGA...G .......... .......... .A........   459
SE9173-J1 ....GATG.G .....CAT.. .GGAC..... ..G....AAT GAC..CACA. .---------- .......A.A .....T.A.. .......... ...G.GA...   931
SE9280-J2 ....GATA.G ....GCAT.G .AG.T..... .G.C...AGT ---N.C.CA. .---------- .......A.A .....T.A.. .......... ...G.GA...   922
```

```
          1001       1011       1021       1031       1041       1051       1061       1071       1081       1091     1100
          |          |          |          |          |          |          |          |          |          |        {V5
Consensus TGGCAGGGAG TAGGACAAGC AATGTATGCC CCTCCCATC- --GCAGGAAA CATTACATGT AAATCAAATA TTACAGGACT ACTATTGACA AGAGATGGTG  1012
SF170-A3  ......A... .......... .......... .........- --CA....GT A..A.AG... .T........ .......... T.....A... ..........  1061
BR20-B1   .......A.. .....A.... .......... .........- --AAG..... A....A.... TC......C. .......G.. .......... ..G.......  1031
TH14-B2   .......A.. .....A.... .......... .........- --AG....C. A....G.... TC........ .......GT. G.....A... ..........  1007
SF162-B3  .......A.. .....A.... .......... .........- --AG....C. A....G...C TC........ .......... G.....A... ..........  1016
MA959-C1  .......AG. ......G... .......... ........T- --.A...... ...A...... .......... .C........ .......... C.T.G...A.  1019
ZM18-C2   ........G. .......... .......... ........T- --........ ...A...... .......... .C........ .......... C.G.....A.  1019
IN868-C3  .......AG. ......G... .......... ........T- --.A...... ...A...... .......... .C........ .......G.. C.T.....A.  1019
UG21-D1   .......... .....A.... .......A.. ........T- --.A....C. A..C.AC... TC.....C.. .......... .T.G...... ..........   998
UG46-D3   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   273
TH22-E1   .......... C......... .........T ........T- --AGG....G A....AT... GT........ ........A. .......... ..G.......  1025
TH06-E2   .......... C......... .........T .........- --AGT....G A....AT... GT........ ........A. ......A... ..........  1019
BZ162-F1  .......A.. ......G.A. .......... G........- --........ ......C... ..C..... .......T.. .......... .....C....   998
BZ163-F2  .......A.. .G....G... .......... G........- --T...A... ......C... ..C..G.... .......T.. .......... ..........   999
RU131-G1  ......A... .G..G..... .......... .........- --........ ...C...... .TT.....C. .......... ......A... ..........  1019
LBV21-G2  ......A.G. .G........ .......... ........T- --........ .......... ........C. .......... .A....A... ..........  1052
VI525-G3  ......A... .G........ .......... .........G CA........ ....GA.... .G.....GC. .......... .A....A... ..........  1079
VI557-H2  ......A... .......... .......... .........- --AA...... A......... G......... .......G.. .A........ ..T...CA..   556
SE9173-J1 .....AA..A C......... ...A...... .........- --........ .........C .......... ........T. .T........ C.T.....A.  1028
SE9280-J2 .....AA.GA C......... ...C...... ........T- --........ ......G..C .C........ ........T. .......... C.T.....A.  1019

          1101       1111       1121       1131       1141       1151       1161       1171       1181       1191     1200
          |          |          |          |          |    V5}   |          |        <----------primer ES8---------- cont'd
Consensus G--------- TAA------T AATAATAATA AA------AC TGAGACCTTC AGACCTGGAG GAGGAGATAT GAGGGACAAT TGGAGAAGTG AATTATATAA  1091
SF170-A3  .--------- GG.------. .....C.G.G C.-------A ...A...... ......AC.. .....A.... .......... .......... ..........  1140
BR20-B1   .--------- ...------. ...G.G.TG. .C---ACG.. C....T.... .......... .......... .......... .....G.... ..........  1113
TH14-B2   .--------- ...------. .G.G.G..C. .G---ACC.. C......... .......... .......... .......... .......... ..........  1089
SF162-B3  .--------- ...------A G.G.TC.G.. .C---ACC.. C....T.... .......... .T........ .......... .......... ..........  1098
MA959-C1  .--------- .CC------A ---.C...C. C.---AAG.. A.....A... ..G....... .......... .......... .......... ..........  1098
ZM18-C2   .--------- GTC------A ...G...CC. C.---AAC.. A.....A... .......... .......... .......... .......... ..C.......  1101
IN868-C3  .--------- A.C------A ---G.GTCA. .T---CAT.. A....TA... .......... .......... ....A..... .......... ..........  1098
UG21-D1   .--------- .GT------G ..A....... GT---CAG.A C......... .......... .G........ ...A.....C .......A.. .......C..  1080
UG46-D3   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   273
TH22-E1   .--------- .GC------. ......GCG- ---------A C......... .......... .....A.... A.A....... .......... ..........  1101
TH06-E2   .--------- .GC------. .......CG. C------T.A C......... .......... .....A.... A.A....... .......... ..........  1098
BZ162-F1  .GTTAGATGA .GGTAGGAA. .C..G.C.G. CTAATCGG.. .....T.... ........G. .....A.... ..AA...... .......... ..........  1098
BZ163-F2  .--------- .C.------G .........C .G-------. .......... ........G. .T..T..... ..AA...T.. .......... T.........  1078
RU131-G1  .--------- G..------C ........C. GCACAAAT.A .......... .......... .......... .......... .......... ..........  1104
LBV21-G2  .--------- G..------. .G..CC...G G.-------A .......... .......C.. .......C.. .......... .......... ..........  1131
VI525-G3  .--------- G..------. .......... CC-------A .......... .......... .......C.. .....G.... .......... ..........  1158
VI557-H2  ---------- ---------- -T....TC.G C.GAAAGT.A ..TT...... .......... .......... .......... .......... ..........   635
SE9173-J1 .--------- A..CA---GG ...GGC.G.G .GAATGGC.. .......... ......AC.. .....A.... ..AA..T... .......... ..........  1116
SE9280-J2 .--------- A..CAGGGGG ...GGC.G.G .GAATGGC.. .......... ......AC.. .....A.... ..A....... .......... ..........  1110
```

```
          1201       1211       1221       1231       1241       1251       1261       1271       1281       1291     1300
 ...primer ES8---<   |          |          |          |          |          |          |          |          |        |
Consensus ATATAAAGTA GTAAAAATTG AACCACTAGG AGTAGCACCC ACCAAGGCAA AGAGAAGAGT GGTGGAG--- AGAGAAAAAA GAGCAGTGGG A---ATAGGA  1185
SF170-A3  G......... .......... .......... .......... ...CC..... .......... ....C..--- .......... .......T.. .---......  1234
BR20-B1   .......... ....G..... .....T.... .A........ ....G..... .......... ....C.A--- .......... .......... .ACAT.....  1210
TH14-B2   .......... ..G....... .....T.... .......... ......C... .......... ....C..--- .......... .......... .ACG......  1186
SF162-B3  .......... .......... .G...T.... .......... .......... .......... ....C..--- .......... .......... -ACGC.....  1192
MA959-C1  G........G ...G.....A .G...T.G.. .......... ..T....... .A..G..... .......--- .......... .......... .---......  1192
ZM18-C2   .........G ...G.....A .....T.G.. CA........ ..TGC..... .A..G..... .......ACG .......... .......... .---......  1198
IN868-C3  ...C.....G ...G.....A .G...T.G.. .....T.... ..T.CA.... .A..G..... .......--- .......... .......... .---..G...  1192
UG21-D1   .......... ....G..... .......... TC........ .......... G......... ......A--- .......... .....A.A.. .---C.G...  1174
UG46-D3   ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------   273
TH22-E1   .......... ...C...... .......... .A........ ....G..... .......... .......--- .......... .......... .---......  1195
TH06-E2   .......... ...C...... .......... .A........ ....G..... .......... .......--- .......... .......... .---......  1192
BZ162-F1  .......... ...G...... .......... .......... .......... .A...CA... ....A..--- .......... .......... .---......  1192
BZ163-F2  G......... ...G...... .......... .......... ....G..... .A...CA... ....AG.--- .......... .......... .---..G...  1172
RU131-G1  G......... ....G....A .......... .......... ....G.---- ---------- ---------- ---------- ---------- ----------  1150
LBV21-G2  G......... A........A .......... .......... .......... G......... .....G.--- .......... .....A.TAC T-------...  1222
VI525-G3  G......... ........CA ..T....... .......... ....G..... G......... .......--- .......... .....A.T.. .---C.G...  1252
VI557-H2  .......... ....G.G.C. .......... .T........ ....G..... G......... ......A--- .......... ....T..... .---C.G.A.   729
SE9173-J1 .......... ...G.GC... .G........ .......... .......... .A........ .......--- .......... .......... .---......  1210
SE9280-J2 .......... ...G.G.... .G........ .......... .......... .A........ .......--- .......... .......... .---......  1204
```

```
          1301       1311       1321       1331
          |        <---------primer ED12----------<
Consensus GCTGTGTTCC TTGGGTTCTT GGGAGCAGCA GGAAGCACT 1224
SF170-A3  .....A...A .......... A........C ......... 1273
BR20-B1   ...A...... .......... .......... ......... 1249
TH14-B2   ...A...... .......... ....A..... ......... 1225
SF162-B3  ...A...... .......... .......... ......... 1231
MA959-C1  .......... .......... .......... ......... 1231
ZM18-C2   .......... .......... .......... ......... 1237
IN868-C3  .......... .......... .......... ....C.... 1231
UG21-D1   ...T...... .......... .......... ........G 1213
UG46-D3   ---------- ---------- ---------- ---------  273
TH22-E1   ...A..A..T .......... A......... ......... 1234
TH06-E2   ...A..A..T .......... A......... ......... 1231
BZ162-F1  ...A...... .......... .......... ......T.. 1231
BZ163-F2  ...T...... .......... .A.C..G... ......... 1211
RU131-G1  ---------- ---------- ---------- --------- 1150
LBV21-G2  .....C.... .......... A........T ..G...... 1261
VI525-G3  .....C.... .......... A......... ......... 1291
VI557-H2  .....T.... .......... .......... ..G......  768
SE9173-J1 .......... .......... ....A..... ......... 1249
SE9280-J2 .......... .......... ....A..... .....T... 1243
```