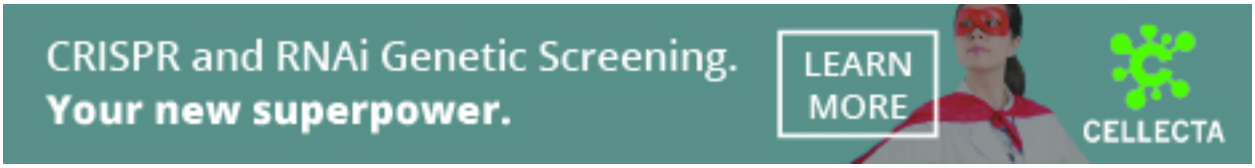
Genetic subtyping of human immunodeficiency virus using a heteroduplex mobility assay.

E L Delwart, B Herring, A G Rodrigo, et al.

Genome Res. 1995 4: S202-S216



Genetic Subtyping of Human Immunodeficiency Virus Using a Heteroduplex Mobility Assay

Eric L. Delwart,[1,2] Belinda Herring,[1,3] Allen G. Rodrigo,[1,4] and James I. Mullins[1,3]

[1]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305–5402

Present addresses: [2]Aaron Diamond AIDS Research Center, New York University School of Medicine, New York, New York 10016 USA; [3]Department of Microbiology and Immunology, University of Washington, Seattle, Washington 98195 USA; [4]Kingett Mitchell and Associates, ASDA Plaza, Takapuna, Auckland, New Zealand.

Different strains of the same "species" of microorganism often display distinctive properties. In a growing number of cases, linkage of phenotypic traits with genetic markers is allowing complicated biological assays to be replaced by genetic typing. The detection of such genetic variation has been revolutionized by the PCR, which allows fragments of even the most complex genomes to be isolated in an essentially pure form in a matter of hours. Differences between gene segments can then be determined by direct sequencing of the PCR product. To further expedite, and thereby to extend genetic screening assays to greater numbers of samples, multiple nonsequencing methods have been developed that are simpler, typically require less complex apparatus, and are of lower cost than DNA sequencing. For example, heteroduplex analysis has been used in the field of medical genetics[1–7] and for the detection of genetic polymorphisms in human populations.[8,9] These methods have involved use of nondenaturing as well as slightly denaturing gel electrophoresis conditions.[10,11] Vinyl polymer gels have also been used for the diagnosis of mutations within proto-oncogenes.[12–14]

Nonsequencing genetic analysis methods can also facilitate study of complex and rapidly evolving genetic systems such as RNA viruses. Recently described heteroduplex mobility assays,[15] now used to classify HIV-1 strains into genetic subtypes, are presented here. These assays should also be readily applicable to the analysis of other highly variable microorganisms.

SEQUENCE, STRUCTURE, AND ELECTROPHORETIC MOBILITY OF HETERODUPLEXES

Heteroduplexes are formed by simply denaturing and reannealing (usually by heating and cooling) partially complementary DNA strands. Sequence variation can then be detected by noting a reduced electrophoretic mobility of DNA heteroduplexes following electrophoresis through a polyacrylamide gel. The structural distortions of the DNA double helix caused by mismatched nucleotides (resulting from base substitution mutations) and unpaired nucleotides or "gaps" (resulting from insertions or deletions within otherwise annealed regions) reduce the mobility of the DNA through the pores of the gel. The effects of theses distortions on heteroduplex mobility are far less pronounced in larger pore, agarose gels.

Unpaired nucleotides in heteroduplexes result in greater mobility retardation than mismatched nucleotides.[1,2,16–19] The nature of unpaired nucleotides in gaps also influences DNA structure and, therefore, mobility in polyacrylamide gels. For example, the larger the gap, the slower the heteroduplex mobility.[18,20] The identity of the unpaired nucleotides also has an effect. For example, single purine insertions result in greater mobility shifts than single pyrimidine insertions.[19] The exact sequence of the base-paired nucleotides neighboring a gap also has a noticeable influence on heteroduplex mobility, indicating that the base stacking interactions close to the helix distortion impacts on the final structure and resulting mobility.[19,20] The spatial arrangement of gaps also influences structure. For example, two single nucleotide insertions separated by a half turn of the helix (5 bp) will largely cancel out the mobility retardation caused by each individual insertion, whereas mobility retardation is increased if the two insertions are separated by 10 nucleotides.[18] Such reinforcement is presumably the result of the insertion-induced bends lying on the same side of the heteroduplex molecule, resulting in a larger total curvature of the DNA fragment. The more centrally located the unpaired nucleotides are, relative to the extremities of the fragment, the greater the mobility retardation.[20] The same effects of central location and phasing of DNA bends is observed for mobility retardations induced by

4:S202–S216©1995 by Cold Spring Harbor Laboratory ISSN 1054-9805/95 $1.00

DNA-bending proteins[21–23] and DNA-bending sequences such as the poly(A) stretch found in trypanosome kinetoplasts.[21,24]

The structure of heteroduplexes has also been analyzed using chemical and enzymatic probes.[20] The resulting structural model for an unpaired gap structure is of a nonflexible kink in the double helix to accommodate the likely extrahelical nucleotides. In contrast, mispaired nucleotides are thought to result in a more flexible, bubble-like distortion of the double helix in which every nucleotide is equally accessible to chemical modification and little overall perturbation of the helix structure is detected.[20] The stronger effects on mobility of unpaired versus mispaired nucleotides and of increasing gap size were also noticed using the HIV-1 envelope DNA sequences described below[15,17] (data not shown).

HETERODUPLEXES FORMED BY HIV-1 ENVELOPE GENE SEQUENCES

Heteroduplexes derived from HIV-1 envelope genes were detected when multiple, genetically divergent, proviral DNA sequences from an individual infected more than 5 years were coamplified in the same PCR reaction and analyzed by polyacrylamide gel electrophoresis.[15,17] With the nested PCR conditions used, primer binding and extension efficiency during the last PCR cycle was low, such that rather than being duplicated, the previous amplification products reannealed randomly to form heteroduplexes.[15,17,25] The reduced migration exhibited by some of the resulting heteroduplexes (detected following electrophoresis by ethidium bromide staining and UV illumination) indicated that at least a portion of the sequence diversity seen within a single blood sample could be detected by this method. Multiple pairwise heteroduplexes formed using previously sequenced variants indicated that a semiquantitative relationship existed between their electrophoretic mobility and the level of sequence divergence between the reannealed DNA strands.[15,17] Using regions of significant base mismatch and length variation (Fig. 1), this relationship could be used to rapidly classify HIV-1 strains into envelope sequence subtypes in what we refer to as the heteroduplex mobility assay, or HMA.[15]

A variation of the HMA involves labeling products of one PCR reaction (e.g., using radioactive or fluorescent molecules) and reannealing it to a 100-fold mass excess of unlabeled DNA from another PCR reaction, such that the labeled "probe" fragments are entirely driven into heteroduplexes formed with the unlabeled PCR product (the "driver"). Following polyacrylamide gel electrophoresis and detection, only heteroduplexes between the two sets of fragments should be detected. The mobility of the labeled heteroduplexes could therefore reveal the genetic relationships between complex mixtures of sequences[15,26] as well as allow detection of specific variants within these mixtures, as indicated by fast migrating heteroduplexes.[17] This heteroduplex tracking assay (HTA) proved useful for tracking the evolution of HIV-1 within individuals,[17] in investigations of suspected HIV-1 transmission,[16] and in detecting the source of sample contamination.[27]

The World Health Organization (WHO) Network on HIV Isolation and Characterization recently conducted a pilot study comparing different methods of HIV-1 envelope gene sequence subtype determination.[28] Results obtained using HMA, RNase A cleavage of RNA/DNA heteroduplexes,[29,30] and anchored primer PCR[31,32] were then compared with results from DNA sequencing. All 54 isolates for which DNA sequence information was available were subtyped correctly by HMA,[33] and because of its speed, low cost, ease of use, and high specificity, HMA was chosen as the initial screening method to determine the HIV-1 subtypes prevalent in countries in which vaccine trials are being prepared.[28]

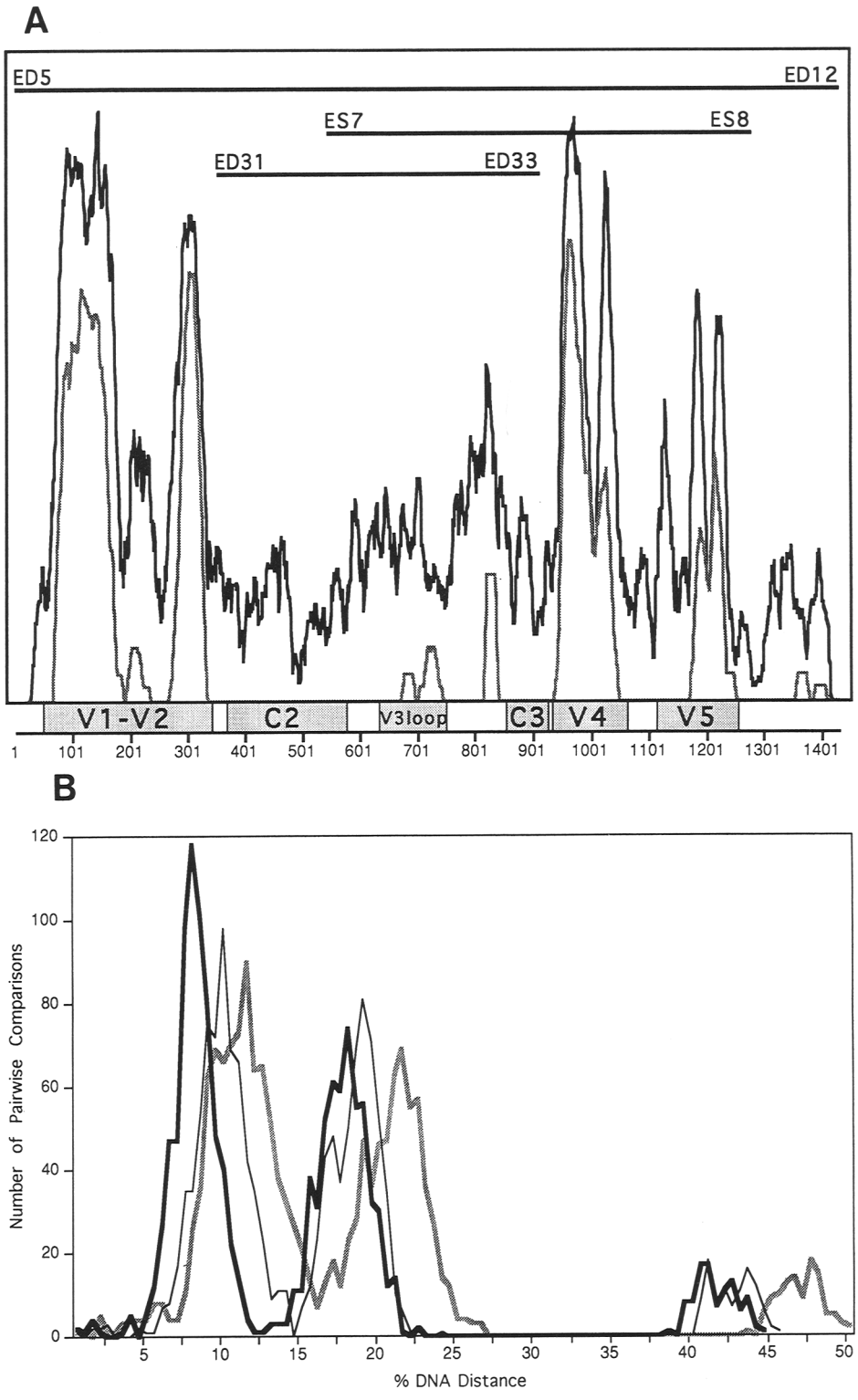


FIGURE 1 (*A*) Distribution of unpaired bases (gray line) and mismatched plus unpaired bases (black line) in aligned HIV sequences across the SU-coding sequence. The number of differences were counted between individual pairs of sequences and tabulated along the *y*-axis. Fragments amplified by the second-round primer sets used in this study are shown. (*B*) Degree of divergence between individual pairs of sequences sampled using fragments delimited by the ED5–ED12 (black line), ES7–ES8 (gray line), and ED31–ED33 (thick black line) primers. Plots were generated using viruses from within the same subtype, between different subtypes, and between the main and O groups.

METHOD DETAILS

DNA Purification

The starting material for HMA is PCR-amplified DNA fragments. Substrates for PCR can include purified DNA or detergent lysed cells from a variety of sources. Red-cell-depleted whole blood and fresh or frozen white blood cells can also be used.[34] However, it should be kept in mind that amplification directly from cellular material is less sensitive than with purified DNA because of the presence of inhibitors and thus is more noticeably dependent on the initial proviral DNA load in each sample.

PCR Amplification

Nested PCR reactions are used to generate sufficient quantities of product for HMA. For analysis of the HIV-1 *env* gene, well-conserved primer sequences were chosen. First-round reactions were conducted using ED3 [5′-TTAGGCATCTCCTATGGCAGGAAGAAGCGG; corresponding to positions 5956–5985 of the HIV-1–HXB2 genome (GenBank accession no. K03455)] and ED14 (5′-TCTTGCCTGGAGCTGTTTGATGCCCCAGAC; positions 7960–7931). These amplify an ~2-kb fragment, spanning from the first exon of *rev* and the complete surface (SU) protein coding sequence of *env*. Three sets of second round primers have been used to date, depending on the region to be analyzed (Fig. 1). They include ED5 (5′-ATGGGATCAAAGCCTAAAGCCATGTG; positions 6556–6581) and ED12 (5′-AGTGCTTCCTGCTGCTCCCAAGAACCCAAG; positions 7822–7792), used to amplify the 1.2-kb first variable (V1) through V5 coding domains of the SU protein. ED31 (5′-CCTCAGCCATTACACAGGCCTGTCCAAAG; positions 6816–6844) and ED33 (5′-TTACAGTAGAAAAATTCCCCTC; positions 7359–7380) amplify the 0.5-Kb C2 through C3 coding domains of SU. Finally, ES7 (5′-tgtaaaacgacggccagtCTGTTAAATGGCAGTCTAGC; positions 7001–7020) and ES8 (5′-caggaaacagctatgaccCACTTCTCCAATTGTCCCTCA; positions 7647–7667) amplify the 0.7-kb V3 through V5 coding domains of SU [lowercase letters are complementary to the universal M13 (ES7) or the reverse (ES8) sequencing primers; these allow direct sequencing of the PCR product but are not essential for heteroduplex analysis]. The 50-μl PCR reaction mixtures contained 200 nM primers, 50 mM KCl, 10 mM Tris (pH 8.3), 1% DMSO, 1% glycerol, 200 μM of each dNTP, and 1.25 mM $MgCl_2$ (or 1.8 mM $MgCl_2$ in the case of ES7–ES8 amplifications). First-round PCR reactions also included 0.2–2 μg of target DNA. Thermal cycling conditions were 3 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min followed by 32 cycles of 94°C for 15 sec, 55°C for 45 sec, 72°C for 1 min, with a final extension at 72°C for 5 min and storage at −20°C until use.

One hundred-microliter volume second-round reactions were initiated with 2 μl of the first-round reaction as template. Additional quantities of DNA from the uncharacterized strain can be generated by repeating the second-round amplification with another aliquot of first-round material. Ten nanograms of plasmid DNA containing HIV-1 envelope genes as envelope sequence subtype references were amplified using second-round primers only. Positive control (10 copies of pNL4-3, containing the HIV-1IIIB genome) and negative control (no template added) reactions were included and carried through the two rounds of amplification.

In some instances, standard amplification conditions, in which primers were annealed at 55°C, did not yield a visible product. In these instances, a "step-up" amplification is carried out for the first 5 cycles, during which primers were annealed at lower temperatures (as low as 37°C), followed by 30 cycles under standard conditions. If samples were still negative after the step-up amplification, an alternative set of second-round or first-round primers (e.g., ED3–ED12 with 1.4 mM $MgCl_2$) were used.

Second-round reactions are checked for amplification products on a 1% agarose gel, mixing 5 μl of the PCR reaction with 1 μl of Ficoll–Orange G loading dye (5× = 25% Ficoll, 1% Orange G) and electrophoresis at 100 V for 1 hr in 1× TBE (89 mM Tris-borate, 89 mM boric acid, 8 mM EDTA) or TAE (40 mM Tris base, 20 mM acetic acid, 2 mM EDTA) buffer. The gel is then stained with ethidium bromide (0.5 μg/ml in H_2O) for 30–60 min, and the DNA is detected by UV transillumination.

Heteroduplex Formation

The first step in the process of analysis of heteroduplexes is to assess the genetic diversity of each PCR-amplified sample. This analysis provides a baseline heteroduplex pattern (because of quasispecies diversity in vivo) with which to compare deliberately formed heteroduplexes. It is useful if this sample is run on the same gel as the heteroduplexes formed between references and the unknown, so that bands present as a result of intra-quasispecies heteroduplexes can be identified (Figs. 3 and 4, below, U lanes). PCR fragments generated from reference strains in plasmids do not result in detectable heteroduplexes, because, except for errors introduced during PCR, all duplexes should be perfectly complementary (Fig. 2). In a 500-μl Eppendorf PCR reaction tube, mix 5 μl of second-round PCR reaction (~100–250 ng of DNA) with 5 μl of H_2O and 1.1 μl of 10× heteroduplex annealing buffer (10× = 1 M NaCl, 100 mM Tris at pH 7.8, 20 mM EDTA).

Heteroduplexes are deliberately formed using approximately equal amounts of amplified fragments from the unknown and each reference. Mix, as above,

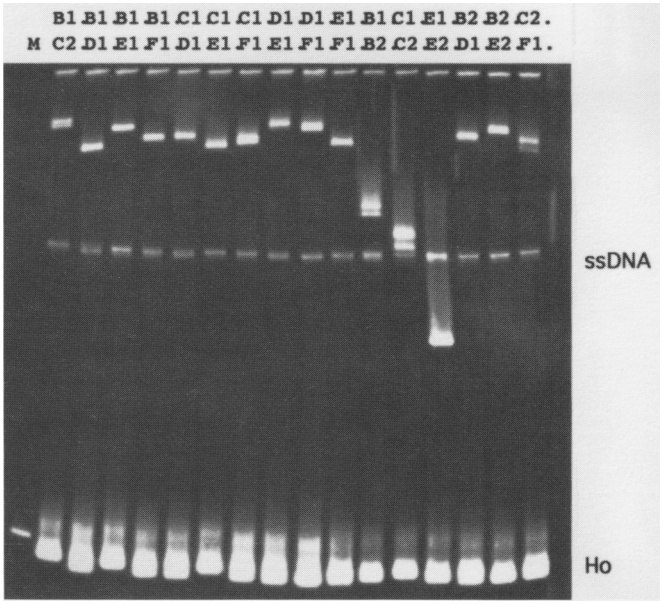


FIGURE 2 Heteroduplexes formed between pairs of DNA fragments PCR amplified from subtype references. ED12–ED5 *env* gene fragments (1.2 kb) were separated in a 5% polyacrylamide gel at 70 mA for 1000 V hr. The subtypes of the two fragments compared are noted above each lane. Numbers next to letters refer to different reference strains, listed below along with the clone name and GenBank accession number (when available): (Lane *A1*) RW20, U08794; (lane *A2*) IC144, no sequence; (lane *A3*) SF170, M66533 (Figs. 3 and 4); (lane *B1*) BR20, U08691; (lane *B2*) TH14, U08801; (lane *B3*) SF162, M65024; (lane *C1*) MA959, U08453; (lane *C2*) ZM18, L22954; (lane *C3*) IND868, U07103; (lane *C4*) BR25, U09133; (lane *D1*) UG21, U08804; (lane *D2*) UG38, U08806; (lane *D3*) UG46, U08809; (lane *E1*) TH22, U09131; (lane *E2*) TH06, U08810; (lane *E3*) CAR7, no sequence; (lane *F1*) BZ162, L22084; (lane *F2*) BZ163, L22085. (Lane M) The molecular mass marker; the 1.35-kb *Hae*III fragment from ϕX174 is visible. (ssDNA) Single-stranded DNA; (Ho) the position of homoduplex migration.

5 μl of the second-round PCR reaction from the unknown (~100–250 ng of DNA) with 5 μl of the second-round PCR reaction from the analogous fragment of *env* derived from a reference strain and 1.1 μl of the 10× annealing buffer.

Form heteroduplexes by heating the mixture to 94°C for 2 min in a thermocycler or boiling water bath. Cool tubes rapidly by transferring to wet ice. Heteroduplexes can be kept at room temperature before loading or stored at −20°C. Samples are mixed with Ficoll–Orange G loading dye prior to loading on a gel. Rapid cooling facilitates stable formation of heteroduplexes between highly divergent sequences but is less important for creating heteroduplexes between closely related sequences. Note that the apparent heteroduplex yield will generally be greater between more closely related sequences (Fig. 3).

Polyacrylamide Gel Electrophoresis

Heteroduplexes are examined after electrophoresis on a 5% polyacrylamide gel (using a 30% acrylamide, 0.8% bis-acrylamide stock) in TBE buffer. A 50-ml mixture is prepared per gel, and polymerization is initiated with the addition of 50 mg of ammonium persulfate and 33 μl of TEMED. We have used the GIBCO BRL V16 vertical gel electrophoresis apparatus and glass plates with 1.5-mm spacers. Other gel systems, including fluorescence-based automated DNA sequencers can also be used (data not shown). ED5–ED12 fragments (1.2 kb) are separated at a constant 200 V for 6 hr or at a 70-mA constant current for 1000 V hr. ES7–ES8 fragments (0.7 kb) are separated at 250 V for 3 hr. ED31–ED33 fragments (0.5 kb) are separated at 250 V for 2 hr. Urea (10% wt/vol) can also be added to gel mixtures to enhance mobility shifts.

STRATEGY FOR SUBTYPING

Nested PCR is used to generate either 1.2-, 0.7-, or 0.5-kb *env* gene fragments from uncharacterized strains of HIV-1. The same size fragments are also amplified from a series of plasmids containing HIV-1 *env* genes from different subtypes used as references. Heteroduplexes formed between the unknown sample and the most closely related reference sequences exhibit the fastest mobilities and thus indicate the likely subtype of that strain (Figs. 2–5). For subtype assignment, heteroduplexes formed with a set of references should have markedly faster mobilities than with other subtypes. Unknowns should

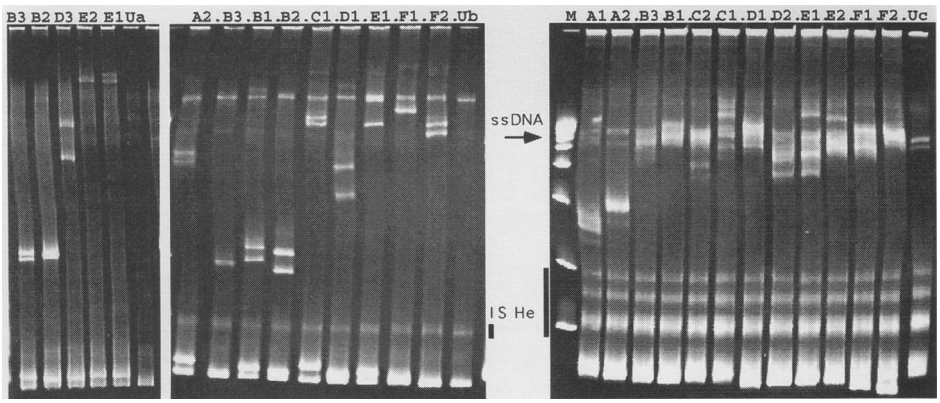


FIGURE 3 Heteroduplexes formed between DNA fragments amplified from infected blood samples and reference plasmids. ED12–ED5 fragments (1.2 kb) were separated in a 5% polyacrylamide gel at 200 V for 6 hr. Reference fragments are noted above each lane and described in the legend to Fig. 2. The subtype determined for each unknown was B (*Ua*), B (*Ub*), and A (*Uc*). ssDNA and Ho are defined in the legend to Fig. 2. IS He refers to the position of intrasample heteroduplexes, indicated by solid vertical bars next to the gels. The ssDNA is indicated with an arrow. Nonspecific amplification products in sample Ub are found just above the ssDNA indicator.

initially be compared with each of the reference strains available to obtain both clear results and to gain familiarity with the assay. Subsequently, when analyzing a geographic region suspected of having principally a single or few subtypes (e.g., subtype B in North America, subtypes B and E in Thailand), it is possible to reduce the number of reference strain comparisons required to definitively assign a subtype. In Thailand for example, where both the subtype B and E strains predominate and are still relatively homogeneous,[15,35] it has been possible to positively identify the subtype of an unknown by comparison to standards from only the B and E clades.[33] The high degree of homology within these subtypes and, therefore, the very fast mobility of their heteroduplexes provide a high degree of confidence in assigning a subtype. Even though a single subtype predominates in the United States, making subtype assignments is slightly more complex because the virus population in the United States is more diversified and U.S. subtype B heteroduplexes will not migrate as rapidly as Thai subtype B heteroduplexes. Therefore, we recommend that North American sequences be compared with multiple B reference strains in addition to at least a single strain from other subtypes (as negative controls). When the expected subtype is not known and in geographic regions where multiple subtypes are present, such as in Uganda, it is necessary to compare each unknown with a panel of two or more references from each available subtype. Comparison to only a single reference sequence from a given subtype may result in ambiguous and, in rare cases, erroneous assignments.

There is great flexibility in the choice of reference strains for each subtype. Within a given geographic region, resident HIV strains of the same subtype are more likely to be closely related to each other than to other geographically more distant strains of the same subtype. The use of locally derived references may therefore result in faster migrating heteroduplexes with an unknown, thereby permitting more rapid and confident assignment of subtype and intrasubtype relationships to be drawn. Use of closely related reference strains also allows confident subtype assignments using fewer pairwise comparisons.

Subtypes may also be assigned using smaller regions of the *env* region amplified with other second round primers [ES7–ES8 (Fig. 4) and ED31–ED33 (Fig. 5)] and are particularly useful when ambiguous results are obtained with the larger fragment. The effect of gaps versus mismatches on the mobility of heteroduplexes is relatively stronger in shorter fragments than in longer fragments, particularly in the lower range of sequence divergence such as that found within an individual.[15,17] However, normally, very few or no gaps are found in the ED31- to ED33-amplified regions (Fig. 1A).

Some samples may be difficult to subtype with confidence by HMA. This generally indicates detection of a genetic outlier within a known subtype (relative to the reference strain used), as shown in Figure 4, or the detection of a new subtype. It is also possible, however, that the difficulty stems from a large deletion or insertion in the V1–V2 or V4–V5 region or, less likely, another region of *env* (Fig. 1A). Such deletions have not affected subtype assignments to date (see a discussion of very large deletions, below), examining >300 strains (data not shown).[15,16,33] Nonetheless, the next step in characterizing an unusual strain would be to try another fragment of *env* or determine a portion of the DNA sequence. As shown in Figure 4, even though the fastest migrating ED5–ED12 heteroduplexes were found with the unknown annealed with subtype A references, their mobilities overlapped that formed with other subtype references. Use of ES7–ES8, however, permitted unambiguous assignment of the unknown as belonging to subtype A. There is no limitation in performing HMA with other segments of the viral genome. The degree of variation presently required for good discrimination of hetero-

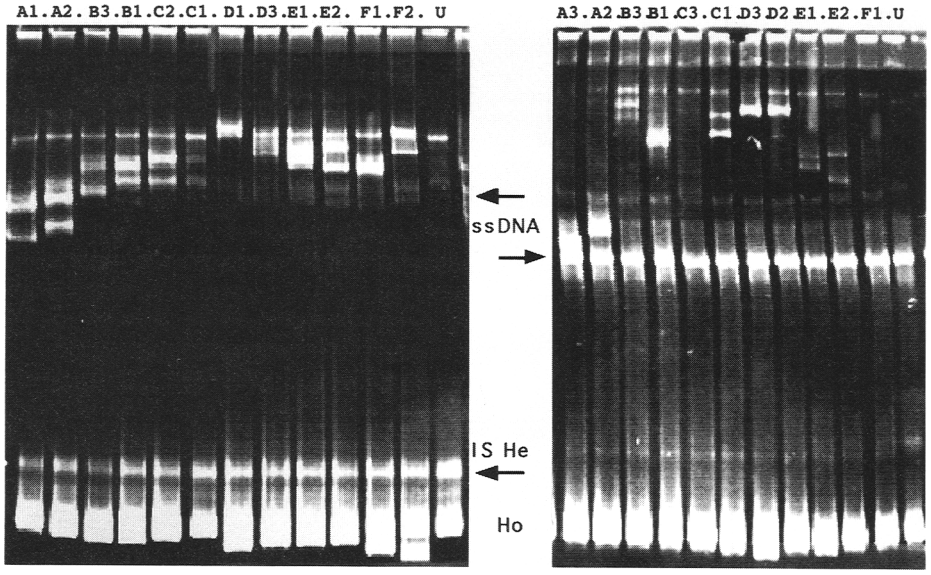


FIGURE 4 Heteroduplexes formed between different size DNA fragments from a HIV-1 subtype A unknown (U) and reference sequences. (*Left*) ED12–ED5 region (1.2 kb), 200-V electrophoresis conditions. (*Right*) ES7–ES8 region (0.7 kb), 250-V conditions. Heteroduplexes in lanes *A3* and *A2* are within the ssDNA region. Reference fragments are noted above each lane and described in the legend to Fig. 2. ssDNA, IS He, and Ho are defined in the legends to Figs. 2 and 3. Nonspecific amplification products are seen above the ssDNA indicators.

duplexes is within the range of 5–25%; hence, the degree of mismatch expected to be encountered should guide these decisions. Because the degree of DNA mismatch variation is similar in all three regions analyzed to date (Fig. 1B), the greater discrimination afforded by the larger fragments is largely the result of length variation and the associated helix-destabilizing effects of adjacent mismatched sequences.

HETERODUPLEX PATTERN ANALYSIS

The Problem of Inherent Quasispecies Complexity

When amplifying viral sequences from a plasmid or a single provirus or RNA template, no heteroduplexes can be formed, because all DNA strands are perfectly complementary (except as a result of polymerase error during amplification) and only homoduplexes are seen on the polyacrylamide gel. When such fragments are reannealed together, a simple heteroduplex pattern of two single or overlapping bands is typically observed (Fig. 2).[15–17] When amplifying sequences from a viral quasispecies, heteroduplexes can form between different, simultaneously amplified variants within the mixture.[15,17] In the latter situation, multiple heteroduplexes are seen when the PCR reaction is analyzed on polyacrylamide gels (Figs. 3 and 4, U lanes). Heteroduplexes can take the form of sharp bands or of a smear-like pattern. The complexity of the heteroduplex pattern and, by extension, the genetic diversity in a single sample can vary widely. For example, soon after infection in vivo and often following virus isolation in vitro, quasispecies typically display a very low level of heterogeneity that is reflected by the presence of primarily homoduplexes in polyacrylamide gels.[17] Conversely, uncultured proviral DNA samples from individuals infected for >5 years typically display a high level of viral sequence complexity.[17] Complex quasispecies are seen as both homoduplexes and as heteroduplexes with reduced mobility. Quasispecies can consist of multiple variants in the absence of a clearly dominant sequence, and in these cases, homoduplexes are not necessarily visible and no

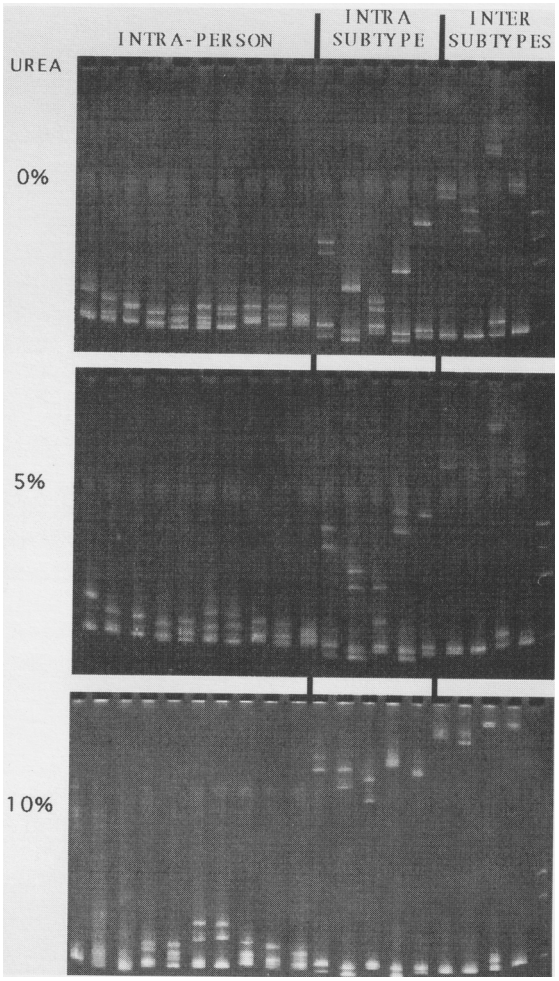


FIGURE 5 Effect of urea concentration on the mobility of using fragments derived from the 0.5-kb ED31–ED33 region. Plasmids harboring *env* genes from the same person (intraperson), same subtype, and different subtypes were used as the source of the DNA fragment pairs used. Urea concentrations present in the gel are indicated at *left*.

particularly bright heteroduplex bands are observed on the polyacrylamide gel (i.e., a smear is evident).

When locating the position of the major heteroduplexes formed between a reference (e.g., clone-derived) sequence and an uncharacterized strain, it is important to identify heteroduplexes formed by the unknown's quasispecies. For that reason, the PCR reaction of the unknown sample is heated and cooled by itself and analyzed (preferably on the same gel as the heteroduplexes formed with reference sequences). Heteroduplexes resulting from inherent quasispecies complexity can then be identified and disregarded when looking for interstrain heteroduplexes (Figs. 3 and 4).

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, two fast migrating homoduplexes (usually with indistinguishable or similar mobilities, e.g., see homoduplex bands at the bottom of Fig. 3) are seen. In addition, usually two (but occasionally comigrating) sharp heteroduplex bands are seen. In contrast, when a complex quasi-species 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 interstrain heteroduplexes may be encountered. We have found it useful to serially dilute the genomic DNA

prior to nested PCR 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. Alternatively, one sample used in forming the heteroduplexes can be labeled and used in a heteroduplex tracking assay to permit unambiguous identification of intersample heteroduplexes.

Appearance of Single-stranded DNA

Discrete single or multiple bands migrating with a mobility of ~40% that of homoduplexes are sometimes seen when examining PCR products (including that formed by references alone) (Figs. 2–4). These bands correspond to collapsed single-stranded DNA (ssDNA) fragments that failed to reanneal with a complementary strand.[25] Their uniform positioning makes them useful for visual comparison of heteroduplex mobilities.

Very Large Deletions

Detection of PCR products with reduced size in agarose gels is generally indicative of the presence of a subpopulation of viral genomes with large internal deletions, not of the type that characterizes normal variation in the V1–V2 and V4–V5 region (Fig. 1A) but rather that extends into normally conserved regions of the coding sequence and hence is derived from obviously defective genomes. Because short DNA fragments are preferentially amplified, the majority of the amplified DNA may in some cases consist of smaller than expected fragments.[36] The occasional presence of these short amplification products results in heteroduplexes formed between fragments differing greatly in size that migrate near the top of the gel. Detection of these heteroduplexes formed within individual patients' quasispecies is thus another indication of deleted genomes. To prevent amplification of the smaller products, nested PCR is repeated using serially decreasing amounts of input genomic DNA until the shorter proviruses are diluted out and only the correct size DNA fragment is amplified.[36] This dilution procedure is usually successful as deleted proviral genomes typically make up only a minority of target proviruses (data not shown).

Nonspecific Amplification Products

Another potential source of apparently very slowly migrating heteroduplexes is fragments formed by amplification of nonviral DNA. Such nonspecific amplification products can be identified by their uniform presence in all lanes containing these products (Figs. 3 and 4).

GEL ELECTROPHORESIS—GENERAL CONSIDERATIONS

Heteroduplex mobility can be greatly slowed at higher temperatures or in the presence of denaturing agents in the gel matrix, particularly when the reannealed DNA strands are from very divergent strains. Temperature increases result in local increases in duplex melting and thus further slow heteroduplex mobility. In Figure 5 the decrease in mobility with increasing concentration of the denaturant urea is greater for more divergent heteroduplexes than for the more related sequences. 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, voltage/current, and buffer conditions should be carefully adhered to in each experiment.

PHYLOGENETIC INFERENCES

Given specific electrophoretic conditions, an equation can be derived to estimate the genetic distance between two DNA fragments from their hetero-

duplex mobility.[15] The estimated genetic distances can then be used to derive fairly reliable phylogenetic relationships between multiple sequences without analyzing all of the possible pairwise heteroduplexes. For this purpose we typically evaluate 25–33% of the $[N\times(N-1)]/2$ possible comparisons (where $N=$ number of sequences being compared).

Relative mobilities of heteroduplexes are typically estimated from photographs or video-captured images of ethidium bromide stained gels. The distance between the loading well and the midpoint between the two heteroduplexes is measured and divided by a value corresponding to the distance between the loading well and the midpoint between the two homoduplexes (the latter often migrate with the same mobility) as given below. A more precise measurement would involve measuring each heteroduplex band independently. 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 has been used to estimate mobilities.

To determine DNA distances, the sequences of each reference, bounded by the second-round primers, are first aligned using one of the available computer programs (e.g., GENALIGN, Intelligenetics Inc., Mountain View, CA), including manual refinement [e.g., MASE[37]]. DNA distances are then calculated by counting mismatches after removal of unpaired sites (gaps) introduced to maintain alignment [e.g., DNADIST from the PHYLIP software package[38]]. This method was chosen to provide a comparison of HMA data to the currently most commonly used methods of HIV sequence analysis for investigation of phylogenetic origins.[39] These methods ignore gaps because there is no generally accepted means of weighting them. Weighting factors are normally available as a user definable option but have not been thoroughly investigated to date. Despite this caveat, a generally reliable relationship between heteroduplex mobility and DNA distance is possible.

To estimate DNA distances from heteroduplex mobility data, standard curves are generated for the relevant electrophoresis conditions by reannealing together pairs of DNA fragments of known sequence. Relative mobilities are then plotted against genetic distances, and the curve is approximated, for example, by an exponential function.

Using the 1.2-kb ED5–ED12 fragments and the exact electrophoresis conditions described above, constant 200-V conditions result in the following relationship: $\text{DNA distance} = -\ln[(\text{mobility}-0.106)/0.94]/7.86$. Using constant 70 mA (for 1000 V hr) results in the following relationship: $\text{DNA distance} = -\ln[(\text{mobility}-0.045)/1.14]/13.55$.

Once a calibration curve has been obtained, a matrix of distances 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, for example, the neighbor-joining[40] and Fitch–Margoliash[41] methods. If, however, only a fraction of all 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[38] [PHYLIP may be obtained by "anonymous ftp" at genetics.washington.edu. Information files and programs are located in the /pub/phylip subdirectory. Users are required to register with J. Felsenstein (e-mail: joe@genetics.washington.edu). PHYLIP is available for Macintosh, IBM-compatible, and UNIX computers]. With FITCH, the user has the option of including missing values in the distance matrix. This is done by using the subreplicate option. The documentation file included with the software 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 PAUP 3.1.1[42] and MacClade.[43]

WHY DETERMINE HIV-1 SUBTYPES?

There are several reasons to anticipate that HIV-1 genetic variation will impact on vaccine efficacy.[44,45] The few mutations observed between subclones of the same HIV-1 strain[46] or rapidly generated following passage in vitro[47] can strongly influence recognition by monoclonal antibodies or neutralization with immune sera. Some cross-neutralization experiments using sera and viral isolates from individuals infected with different subtypes indicated that neutralization titers were generally higher using viruses and sera of the same subtype than of different subtypes.[47] Difficulties in establishing protection against slightly heterogeneous strains in animal vaccine models[48–50] suggest that the high level of genetic variation found between the major subtypes of HIV-1 will also affect vaccine efficacy. It is therefore possible that vaccines will induce levels of protection against challenge strains that at least are in part proportional to their level of genetic relatedness to the vaccine strain. However, evidence of convergence of protein structure in the V3 region of Env in viruses from different HIV-1 subtypes has been presented recently,[51] suggesting that some cross-protection may be possible. Nonetheless, if phylogenetically defined groups broadly correspond to antigenic groups, vaccination using the most appropriate (antigenically related) strain will require information regarding the genetic subtypes of the most likely challenge strains found regionally.

A large breadth of HIV-1 sequence variation has occurred world-wide.[39,52] A growing number of highly divergent subtypes are being documented,[52–54] including recent documentation of an outlier group (group O,[55,56]), divergent from the previously recognized "main" group of HIV-1 strains by as much as 45% in *env* (Fig. 1B). In Thailand, prevalent subtypes have been shown to vary according to risk group within the same city.[35,57] The predominance of phylogenetically clustered strains in some geographic areas may reflect introduction by single individuals into high risk groups.[15,26,35,58] The ensuing rapid transmission of these variants may have then obscured an otherwise more gradual continuum of sequence variation present in the donor population. Strong founder effects as well as uneven geographic and temporal sampling may have all contributed to the phylogenetic clustering predominantly seen today.

The main group of HIV-1 isolates can to date be grouped within one of eight distinct genetic subtypes (A–H) whose envelope sequences differ by up to 30%. These subtypes may therefore have evolved from a common human precursor along with the fitful nature of global virus spread. When the increase in HIV-1 genetic variation was analyzed within a geographic area in which a single subtype was introduced, a gradual increase in genetic diversity was observed over time.[59] Alternatively, distinct zoonotic transmissions from a yet undetermined source, perhaps chimpanzees, may account for the various subtypes.[60] Multiple zoonotic transmissions for HIV seem particularly likely when considering HIV-2, which is now recognized to be phylogenetically indistinguishable from SIV,[61,62] and the highly divergent HIV-1 type O group.

HIV-1 SUBTYPING KIT

An HIV-1 envelope gene subtyping kit, based on the heteroduplex mobility assays presented here, has been developed in collaboration with the WHO Network on HIV Isolation and Characterization and is available freely

through the National Institutes of Health (NIH) AIDS Research and Reference Reagent Program in the United States. It currently consists of 23 reference plasmids from eight HIV-1 subtypes, a detailed protocol, and the four sets of PCR primers described here. These protocols are subject to improvement, particularly with respect to the genomic regions analyzed and the gel conditions used. The rapid communication of technical improvements and the submission to the NIH and MRC AIDS reagent programs of subclones of new HIV subtypes for distribution for heteroduplex analysis will greatly improve later versions of the kit and are encouraged.

ACKNOWLEDGMENTS

We thank Dr. Eugene G. Shpaer for help in developing the computational analyses presented here. This work was supported by U.S. Public Health Service grant R01 AI32885 and by a grant from the World Health Organization.

REFERENCES

1. Cai, S.-P., B. Eng, Y.W. Kan, and D.H.K. Chui. 1991. A rapid and simple electrophoretic method for the detection of mutations involving small insertion and deletion: Application to β-thalassemia. *Hum. Genet.* **87:** 728–730.
2. Farrar, G.J., P. Kenna, S.A. Jordan, R. Kumar-Singh, M.M. Humphries, E.M. Sharp, D.M. Sheils, and P. Humphries. 1991. A three-base-pair deletion in the peripherin-RDS gene in one form of retinitis pigmentosa. *Nature* **354:** 478–480.
3. Paw, B.H., P.T. Tieu, M.M. Kaback, J. Lim, and E. Neufeld. 1990. Frequency of three Hex A mutant alleles among Jewish and non-Jewish carriers identified in a Tay-Sachs screening program. *Am. J. Hum. Genet.* **47:** 698–705.
4. Rommens, J., B.-S. Kerem, W. Greer, P. Chang, L.-C. Tsui, and P. Ray. 1990. Rapid nonradioactive detection of the major cyctic fibrosis mutation. *Am. J. Hum. Genet.* **46:** 395–396.
5. Clay, T.M., J.L. Bidwell, M.R. Howard, and B.A. Bradley. 1991. PCR-fingerprinting for selection of HLA matched unrelated marrow donors. *Lancet* **337:** 1049–1052.
6. Gyllensten, U. and M. Allen. 1991. PCR-based HLA class II typing. *PCR Methods and Applic.* **1:** 91–98.
7. Sorrentino, R., I. Cascino, and R. Tosi. 1992. Subgrouping of DR4 alleles by DNA heteroduplex analysis. *Hum. Immunol.* **33:** 18–23.
8. Ruano, R., A.S. Deinard, S. Tishkoff, and K.K. Kidd. 1994. Detection of DNA sequence variation via deliberate heteroduplex formation from genomic DNAs amplified en masse in "population tubes." *PCR Methods Applic.* **3:** 225–231.
9. Ruano, G., and K.K. Kidd. 1992. Modeling of heteroduplex formation during PCR from mixtures of DNA templates. *PCR Methods Applic.* **2:** 112–116.
10. Borresen, A.-L., E. Hovig, B. Smith-Sorensen, D. Malkin, S. Lystad, T.I. Andersen, J.M. Nesland, K.J. Isselbacher, and S.H. Friend. 1991. Constant denaturant gel electrophoresis as a rapid screening technique for p53 mutations. *Proc. Natl. Acad. Sci.* **88:** 8405–8409.
11. White, M.B., M. Carvalho, D. Derse, S.J. O'Brien, and M. Dean. 1992. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12:** 301–306.
12. Keen, J., D. Lester, C. Inglehearn, A. Curtis, and S. Bhattacharya. 1991. Rapid detection of single-base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.* **7:** 5.
13. Soto, D. and S. Sukumar. 1992. Improved detection of mutations in the p53 gene as single-stranded conformational polymorphs and double-stranded heteroduplex DNA. *PCR Methods Applic.* **2:** 96–98.
14. Perrey, D.J. and R.W. Carrel. 1992. Hydrolink gels: A rapid and simple approach to the detection of DNA mutations in thromboembolic disease. *J. Clin. Path.* **45:** 158–160.
15. Delwart, E.L., E.G. Shpaer, F.E. McCutchan, J. Louwagie, M. Grez, H. Rübsamen-Waigmann, and J.I. Mullins. 1993. Genetic relationships determined by a DNA heteroduplex mobility assay: Analysis of HIV-1 *env* genes. *Science* **262:** 1257–1261.
16. Delwart, E.L., M.P. Busch, M.L. Kalish, J.W. Mosley, and J.I. Mullins. 1995. (in prep.).
17. Delwart, E.L., H.W. Sheppard, B.D. Walker, J. Goudsmit, and J.I. Mullins. 1994. HIV-1 evolution *in vivo* tracked by DNA heteroduplex mobility assays. *J. Virol.* **68:** 6672–6683.
18. Hsieh, C.H. and J.D. Griffith. 1989. Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: Possible relevance to the folding of single-stranded nucleic acids. *Proc. Natl. Acad. Sci.* **86:** 4833–4837.
19. Wang, Y.-H. and J. Griffith. 1991. Effects of bulge composition and flanking sequence on the kinking of DNA by bulged bases. *Biochemistry* **30:** 1358–1363.
20. Bhattacharyya, A. and D.M.J. Lilley. 1989. The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res.* **17:** 6821–6840.

21. Wu, H.-M. and D.M. Crothers. 1984. The locus of sequence-directed and protein-induced DNA bending. *Nature* **308:** 509–513.
22. Zahn, K. and F.R. Blattner. 1987. Direct evidence for DNA bending at the lambda replication origin. *Science* **236:** 416–422.
23. Kerppola, T.K. and T. Curran. 1991. Fos-Jun heterodimers and Jun homodimers bend DNA in opposite orientations: Implications for transcription factor cooperativity. *Cell* **66:** 317–326.
24. Zinkel, S.S. and D.M. Crothers. 1987. DNA bend direction by phase sensitive detection. *Nature* **328:** 178–181.
25. Jensen, M.A. and N. Straus. 1993. Effect of PCR conditions on the formation of heteroduplex and single-stranded DNA products in the amplification of bacterial ribosomal DNA spacer regions. *PCR Methods Applic.* **3:** 186–194.
26. Louwagie, J., E.L. Delwart, J.I. Mullins, F.E. McCutchan, G. Eddy, and D.S. Burke. 1994. Genetic analysis of HIV-1 isolates from Brazil reveals the presence of two distinct genotypes. *AIDS Res. Hum. Retrovir.* **10:** 561–567.
27. Sabino, E.C., E. Delwart, T.-H. Lee, A. Mayer, J.I. Mullins, and M.P. Busch. 1994. Identification of low-level contamination of blood as basis for detection of human immunodeficiency virus (HIV) DNA in anti-HIC-negative specimens. *J. AIDS* **7:** 853–859.
28. Osmanov, S., L. Belsey, W. Heyward, J. Esparza, J. Bradac, B. Galvao-Castro, P. Van de Perre, E. Karita, S. Sempala, B. Tugume, B. Biryanwaho, C. Wasi, H. Rübsamen-Waigmann, H. von Briesen, U. Esser, M. Grez, H. Holmes, F.E. McCutchan, J. Louwagie, P. Hegerich, C. Lopez-Galindez, J.I. Mullins, E.L. Delwart, M.H. Bachmann, J. Goudsmit, F. deWolf, B. Hahn, F. Gao, L. Yue, S. Saragosti, G. Schochetman, M. Kalish, R. George, J. Weber, R. Cheingson-Popov, P. Kaleebu, P. Nara, E.M. Fenyo, J. Albert, G. Myers, and B. Korber. 1994. HIV-1 variation in WHO-sponsored vaccine evaluation sites: Genetic screening, sequence analyses and preliminary biological characterization of representative viral strains. *AIDS Res. Hum. Retrovir.* **10:** 1327–1343.
29. Sanchez-Palomino, S., J.M. Rojas, M.A. Martinez, E.M. Fenyo, R. Najera, E. Domingo, and C. Lopez-Galindez. 1993. Dilute passage promotes expression of genetic and phenotypic variants of human immunodeficiency virus type 1 in cell culture. *J. Virol.* **67:** 2938–2943.
30. Lopez-Galindez, C., J.M. Rojas, R. Najera, D.D. Richman, and M. Perucho. 1991. Characterization of genetic variation and 3′-azido-3′-deoxythymidine-resistance mutations of human immunodeficiency virus by the RNase A mismatch cleavage method. *Proc. Natl. Acad. Sci.* **88:** 4280–4284.
31. McCutchan, F.E., B.L. Ungar, P. Hegerich, C.R. Roberts, A.K. Fowler, S.K. Hira, P.L. Perine, and D.S. Burke. 1992. Genetic analysis of HIV-1 isolates from Zambia and an expanded phylogenetic tree for HIV-1. *J. AIDS* **5:** 441–449.
32. McCutchan, F.E., E. Sanders-Buell, C.W. Oster, R.R. Redfield, S.K. Hira, P.L. Perine, B.L. Ungar, and D.S. Burke. 1991. Genetic comparison of human immunodeficiency virus (HIV-1) isolates by polymerase chain reaction. *J. AIDS* **4:** 1241–1250.
33. Bachmann, M.H., E.L. Delwart, E.G. Shpaer, P. Lingenfelter, R. Singal, J.I. Mullins, and the WHO Network on HIV isolation and characterization. 1994. Rapid genetic characterization of HIV-1 from four WHO-sponsored vaccine evaluation sites using a heteroduplex mobility assay. *AIDS Res. Hum. Retrovir.* **10:** 1343–1351.
34. Mercier, B., C. Gaucher, O. Feugaes, and C. Mazurier. 1990. Direct PCR from whole blood, without DNA extraction. *Nucleic Acids Res.* **18:** 5908.
35. McCutchan, F.E., P.A. Hegerich, T.P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P.W. Berman, A.M. Gray, A.K. Fowler, and D.S. Burke. 1992. Genetic variants of HIV-1 in Thailand. *AIDS Res. Hum. Retrovir.* **8:** 1887–1895.
36. Edmonson, P.L. and J.I. Mullins. 1992. Efficient amplification of half-genome sized fragments of human immunodeficiency virus from infected tissue samples. *Nucleic Acids Res.* **20:** 4933.
37. Faulkner, D.V. and J. Jurka. 1988. Multiple aligned sequence editor (MASE). *Trends Biochem. Sci.* **13:** 321–322.
38. Felsenstein, J. 1989. PHYLIP—Phylogeny inference package. *Cladistics* **5:** 164–166.
39. Myers, G., B. Korber, S. Wain-Hobson, R.F. Smith, and G.N. Pavlakis. 1993. Human retroviruses and AIDS 1993. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences, Los Alamos National Laboratory, NM.
40. Saitou, N. and M. Nei. 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4:** 406–425.
41. Fitch, W.M. and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* **155:** 279–284.
42. Swofford, D.L. 1993. *PAUP 3.1.1: Phylogenetic analysis using parsimony*, 3.1.1 ed. Illinois Natural History Survey, Champaign, IL.
43. Maddison, W.P. and D.R. Maddison. 1992. MacClade 3.01. 3.01 ed. Sinauer and Associates, Sunderland, MA.
44. Cheng-Mayer, C. 1993. HIV-1 variation: Consequences for disease progression and vaccine strategies. *Trends Microbiol.* **1:** 353–355.

45. Sonigo, P., M. Girard, and D. Dormont. 1990. Design and trials of AIDS vaccines. *Immunol. Today* **11:** 465–471.
46. Moore, J.P., H. Yoshiyama, D.D. Ho, J.E. Robinson, and J. Sodroski. 1993. Antigenic variation in gp120 from molecular clones of HIV-1 LAI. *AIDS Res. Hum. Retrovir.* **9:** 1179–1187.
47. Mascola, J.R., J. Louwagie, F.E. McCutchan, C.L. Fischer, P.A. Hegerich, K.F. Wagner, A.K. Fowler, J.G. McNeil, and D.S. Burke. 1994. Two antigenetically distinct subtypes of HIV-1: Viral genotype predicts neutralization serotype. *J. Infect. Dis.* **169:** 48–54.
48. Alter, J.J., J.W. Eichberg, H. Masur, W.C. Saxinger, R.C. Gallo, H.C. Lane, and A.S. Fauci. 1984. Transmission of HTLV-III from human plasma to chimpanzees: An animal model for AIDS. *Science* **226:** 549–552.
49. Issel, C.J., D.W. Horohov, D.F. Lea, W. Adams Jr., S.D. Hagius, J.M. McManus, A.C. Allison, and R.C. Montelaro. 1992. Efficacy of inactivated whole-virus and subunit vaccines in preventing infection and disease caused by equine infectious anemia virus. *J. Virol.* **66:** 3398–3408.
50. Luke, W., G. Voss, C. Stahl-Hennig, C. Coulibaly, P. Putkonen, H. Petry, and G. Hunsmann. 1993. Protection of cynomolgus macaques (*Macaca fascicularis*) against infection with the human immunodeficiency virus type 2 strain ben (HIV-2ben) by immunization with the virion-derived envelope glycoprotein gp130. *AIDS Res. Hum. Retrovir.* **9:** 387–394.
51. Korber, B.T., K. MacInnes, R.F. Smith, and G. Myers. 1994. Mutational trends in V3 loop protein sequences observed in different genetic lineages of human immunodeficiency virus type 1. *J. Virol.* **68:** 6730–6744.
52. Louwagie, J., F.E. McCutchan, M. Peeters, T.P. Brennan, E. Sanders-Buell, G.A. Eddy, G. van der Groen, K. Fransen, G.-M. Gershy-Damet, R. Deleys, and D.S. Burke. 1993. Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes. *AIDS* **7:** 769–780.
53. Morgado, M.G., E.C. Sabino, E.G. Shpaer, V. Bongertz, L. Brigido, M.D.C. Guimaraes, E.A. Castilho, B. Galvao-Castro, J.I. Mullins, R.M. Hendry, and A. Mayer. 1994. V3 region polymorphisms in HIV-1 from Brazil: Prevalence of subtype B strains divergent from the North American/European prototype and detection of subtype F. *AIDS Res. Hum. Retrovir.* **10:** 569–576.
54. Dumitrescu, O., M.L. Kalish, S.C. Kliks, C.I. Bandea, and J.A. Levy. 1994. Characterization of human immunodeficiency virus type 1 isolates from children in Romania: Identification of a new envelope subtype. *J. Infect. Dis.* **169:** 281–288.
55. Gürtler, L.G., P.H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J.M. Tsague, and L. Kaptue. 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. *J. Virol.* **68:** 1581–1585.
56. Vanden Haeselvelde, M., J.-L. Decourt, R.J. De Leys, B. Vanderborght, G. van der Groen, H. van Heuverswijn, and E. Saman. 1994. Genomic cloning and complete sequence analysis of a highly divergent African Human Immunodeficiency Virus Isolate. *J. Virol.* **68:** 1586–1596.
57. Ou, C.Y., Y. Takebe, B.G. Weniger, C.C. Luo, M.L. Kalish, W. Auwanit, S. Yamazaki, H.D. Gayle, N.L. Young, and G. Schochetman. 1993. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* **341:** 1171–1174.
58. Grez, M., U. Dietrich, P. Balfe, H. von Briesen, J.K. Maniar, G. Mahambre, E.L. Delwart, J.I. Mullins, and H. Rübsamen-Waigmann. 1994. Genetic analysis of human immunodeficiency virus type 1 and 2 (HIV-1 and HIV-2) mixed infections in India reveals a recent spread of HIV-1 and HIV-2 from a single ancestor for each of these viruses. *J. Virol.* **68:** 2161–2168.
59. Kuiken, C., G. Zwart, E. Baan, R. Coutinho, J. Anneke, R. van den Hoek, and J. Goudsmit. 1993. Increasing antigenic and genetic diversity of the HIV-1 V3 domain in the course of the AIDS epidemic. *Proc. Natl. Acad. Sci.* **90:** 9061–9065.
60. Huet, T., R. Cheynier, A. Meyerhans, G. Roelants, and S. Wain-Hobson. 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. *Nature* **345:** 356–359.
61. Jin, M.J., J. Rogers, J.E. Phillips-Conroy, J.S. Allan, R.C. Desrosiers, G.M. Shaw, P.M. Sharp, and B.H. Hahn. 1994. Infection of a Yellow Baboon with simian immunodeficiency virus from African Green Monkeys: Evidence for cross-species transmission in the wild. *J. Virol.* **68:** 8454–8460.
62. Gao, F., L. Yue, A.T. White, P.G. Pappas, J. Barchue, A.P. Hanson, B.M. Greene, P.M. Sharp, G.M. Shaw, and B.H. Hahn. 1992. Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. *Nature* **358:** 495–499..