giyama, in *Bioorganic Photochemistry, Photochemistry and the Nucleic Acids*, H. Morrison, Ed. (Wiley, New York, 1990), vol. 1, pp. 317–378.
2. H. Weintraub, *Cold Spring Harbor Symp. Quant. Biol.* **38**, 247 (1973); S. Y. Lin and A. D. Riggs, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 947 (1974). B. M. Barbier, M. Charlier, J.-C. Maurizot, *Biochemistry* **23**, 2933 (1984); R. Ogata and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4973 (1977); M. Katouzian-Safadi, B. Blazy, M. Charlier, *Photochem. Photobiol.* **53**, 611 (1991); H. Wolfes, A. Fliess, F. Winkler, A. Pingoud, *Eur. J. Biochem.* **159**, 267 (1986); K. Khalili, J. Rappaport, G. Khoury, *EMBO J.* **7**, 1205 (1988); M. Katouzian-Safadi *et al.*, *Nucleic Acids Res.* **19**, 4937 (1991).
3. K. L. Wick and K. S. Matthews, *J. Biol. Chem.* **266**, 6106 (1991); T. D. Allen, K. L. Wick, K. S. Matthews, *ibid.*, p. 6113.
4. E. E. Blatter, Y. W. Ebright, R. H. Ebright, *Nature* **359**, 650 (1992).
5. N. K. Tanner, M. M. Hanna, J. Abelson, *Biochemistry* **27**, 8852 (1988).
6. J. M. Gott, M. C. Willis, T. H. Koch, O. C. Uhlenbeck, *ibid.* **30**, 6290 (1991).
7. The mechanism for photocoupling of the 5-iodouracil chromophore to amino acid residues has not been established, but it may be similar to that of the 5-bromouracil chromophore [T. M. Dietz, R. J. von Trebra, B. J. Swanson, T. H. Koch, *J. Am. Chem. Soc.* **109**, 1793 (1987)]. A model study was performed to establish similarity between 5-iodouracil and 5-bromouracil photocrosslinking. Irradiation of an aqueous solution of iodouracil and *N*-acetyltyrosine *N*-ethylamide at 308 nm gave a photoadduct identical to that from irradiation of bromouracil and *N*-acetyltyrosine *N*-ethylamide [T. M. Dietz and T. H. Koch, *Photochem. Photobiol.* **46**, 971 (1987)].
8. M. C. Willis, O. C. Uhlenbeck, T. H. Koch, in preparation.
9. A previous measurement of the quantum yield of photocrosslinking of RNA 1 to the R17 coat protein was lower (*6*). Our technique for performing this measurement has improved substantially, and we believe the new value to be correct.
10. Emission from a HeCd laser is advantageous for crosslinking because at 325 nm absorption by other chromophores in the nucleoprotein complex such as thymine, guanine, indole, and phenol is not observed. In the region of 325 nm only the 5-iodouracil chromophore absorbs appreciably; the molar extinction coefficient is 1.63×10^5 cm^2/mol. For comparison, bromouracil and iodouracil absorb at 308 nm with extinction coefficients of 3.85×10^5 and 2.64×10^6 cm^2/mol, respectively. The HeCd laser also eliminates the possibility of two-photon excitation because it is a continuous wave laser as opposed to the pulsed XeCl excimer laser.
11. A 45% crosslinking yield to R17 coat protein was obtained with IU–RNA 2 and polystyrene-filtered emission from a 312-nm, broad-band Spectroline transilluminator. For comparison, only a 20% crosslinking yield was obtained with BrU–RNA 1 and the transilluminator.
12. The diversity of bands generated on excitation at 308 nm is due to photocrosslinking of different amino acids in the protein to at least three different nucleotides which, when substituted with BrdU, result in enhancement of three distinct bands within the gel pattern (*13*).
13. B. J. Hicke, M. C. Willis, T. H. Koch, T. R. Cech, in preparation.
14. RNAs 1, 2, and 3 were prepared by in vitro transcription from synthetic DNA templates by T7 RNA polymerase with [α-^{32}P]CTP [J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **15**, 8783 (1987)]; IUTP was from Sigma. RNA fragments were purified by 20% denaturing polyacrylamide gel electrophoresis (PAGE). The desired fragment was eluted from the polyacrylamide and ethanol-precipitated in the presence of 0.3 M sodium acetate. R17 bacteriophage was propagated in *Escherichia coli* strain S26, and the coat protein was purified according to the procedure described by J. Carey, P. T. Lowary, and O. C. Uhlenbeck [*Biochemistry* **22**, 4723 (1983)].
15. DNAs 4 and 5 were prepared on a DNA synthesizer (IdU phosphoramidite from Glen Research) and ^{33}P end-labeled. *Oxytricha nova* telomere protein subunits were expressed and purified according to the procedure described by G. Fang, J. T. Gray, and T. R. Cech [*Genes Dev.* **7**, 870 (1993)]. The protein was purified by cation exchange chromatography followed by dialysis against MNG 20/20/20 [20 mM MOPS (pH 7.5), 20 mM NaCl, 20% glycerol] or spin dialysis (Centricon 30 filter, Amicon) into MNG 20/20/20. Protein prepared in this fashion was analyzed by SDS-PAGE and its purity and concentration estimated by comparison to serial dilutions of protein purified according to the procedure described by J. T. Gray, D. W. Celander, C. M. Price, and T. R. Cech [*Cell* **67**, 807 (1991)].
16. A constant, low concentration of ^{32}P-labeled RNA was mixed with a series of coat protein concentrations between 0.06 nM and 1 μM in 10 mM magnesium acetate, 80 mM KCl, bovine serum albumin (BSA) (80 μg/ml), and 100 mM tris-HCl (pH 8.5 at 4°C) (TMK buffer). These were the same conditions used in the crosslinking experiments. After incubation at 4°C for 45 to 60 min, the mixture was filtered through a nitrocellulose filter and the amount of complex retained on the filter determined by liquid scintillation counting.
17. ^{32}P-Labeled RNAs 1 and 2 (5 nM) and R17 coat protein (120 nM) were incubated on ice in 100 mM tris-HCl (pH 8.5 at 4°C), 80 mM KCl, 10 mM magnesium acetate, and BSA (80 μg/ml) for 15 to 25 min before irradiation. These are conditions under which the RNA is fully bound to the coat protein. The RNAs were heated in water to 85°C for 3 min and quick-cooled on ice before use to ensure that the RNAs were in a hairpin conformation [D. R. Groebe and O. C. Uhlenbeck, *Nucleic Acids Res.* **16**, 11725 (1988)]. A Lambda Physik EMG-101 excimer laser was used for irradiations at 308 nm. The output was directed unfocused toward a 4 mm wide by 1 cm path length quartz cuvette containing the RNA-protein complex. The laser was operated in the range of 60 mJ per pulse at 10 Hz; however, only about 25% of the laser beam was incident on the reaction cell. Crosslinked RNA was separated from uncrosslinked RNA by PAGE, and the yields were quantitated with a Phosphor Imager.
18. Supported by the Council for Tobacco Research (M.C.W. and T.H.K.), NIH grants GM-36944 (O.C.U.) and GM-28039 (T.R.C.), an NSF graduate fellowship (B.J.H.), the Howard Hughes Medical Institute, (T.R.C.), and an American Cancer Society professorship (T.R.C.).

11 June 1993; accepted 22 September 1993

Genetic Relationships Determined by a DNA Heteroduplex Mobility Assay: Analysis of HIV-1 *env* Genes

Eric L. Delwart, Eugene G. Shpaer, Joost Louwagie, Francine E. McCutchan, Manuel Grez, Helga Rübsamen-Waigmann, James I. Mullins*

The genetic diversity of human immunodeficiency virus (HIV) is a major concern thought to impact on immunologic escape and eventual vaccine efficacy. Here, simple and rapid methods are described for the detection and estimation of genetic divergence between HIV strains on the basis of the observation that DNA heteroduplexes formed between related sequences have a reduced mobility in polyacrylamide gels proportional to their degree of divergence. Reliable phylogenetic subtypes were assigned for HIV-1 strains from around the world. Relationships between viruses were closest when derived from the same or epidemiologically linked individuals. When derived from epidemiologically unlinked individuals, the relationships between viruses in a given geographic region correlated with the length of time HIV-1 had been detected in the population and the number of strains initiating widespread infection. Heteroduplex mobility analysis thus provides a tool to expedite epidemiological investigations by assisting in the classification of HIV and is readily applicable to the screening and characterization of other infectious agents and cellular genes.

Human immunodeficiency viruses, like other RNA viruses, exist within their hosts as pools of related genetic variants, often referred to as quasispecies (*1–3*). Within infected individuals the HIV-1 surface envelope (*env*) glycoprotein coding sequences have been found to vary by up to 8%, with an unusually high ratio of nonsynonymous to synonymous mutations, indicative of strong selection for viral surface change, and numerous small in-frame nucleotide deletions and insertions (*4–7*). The highly variable and continuously evolving nature of HIV-1 within individuals accounts for the rapid emergence of viral variants resistant to neutralizing antibodies, cytotoxic T lymphocytes, and antiviral drugs (*8–13*), and contributes to the high level of genetic diversity observed between viral strains identified worldwide. The difficulty in elic-

E. L. Delwart, E. G. Shpaer, J. I. Mullins, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA 94305–5402.
J. Louwagie and F. E. McCutchan, Henry M. Jackson Foundation for the Advancement of Military Medicine, Rockville, MD 20850.
M. Grez and H. Rubsamen-Waigmann, Georg-Speyer-Haus Chemotherapeutisches Forschungsinstitut, Paul-Ehrlich-Straße 42-44, Frankfurt 700251, Germany.
*To whom correspondence should be addressed.

iting cellular and humoral immunity against heterologous viral strains suggests that the development of effective vaccines for protection against lentiviruses such as HIV-1 will present a considerable challenge (*14–16*). Vaccines that include antigens closely related to those likely to be presented by an infectious challenge are therefore thought to elicit greater protective immunity and may require geographically and temporally specific formulation.

The aforementioned insights into the genetic variation of HIV-1 have been achieved partly through large-scale DNA sequence analyses. The large number of basic questions remaining about HIV genomic variation as it relates to virus transmission, disease progression, tissue specificity, vaccine development, and drug resistance portends a need for further, extensive genetic analysis of HIV. Rapid assays were therefore developed for estimating the degree of genetic similarity between viruses without costly and laborious large-scale DNA sequencing.

Nested polymerase chain reaction (PCR) (*17*) with two 35 primer-extension cycles was used to amplify a 1.8-kb and then a 0.7-kb internal fragment of the HIV-1 *env* (encoding the V3-V5 region of the surface protein) directly from the peripheral blood mononuclear cells (PBMC) of an asymptomatic HIV-1 seropositive man (*18–20*). The conditions used permitted detection of a single molecule of target DNA (*21*). When analyzed on an agarose gel, a single band of the expected apparent molecular weight was observed (Fig. 1A). However, when the DNA was analyzed on a neutral 5% polyacrylamide gel, additional, prominent bands of higher apparent molecular weight were observed (Fig. 1B). Given the known variability of HIV-1 within PBMC populations, the possibility that the slower migrating bands were composed of heteroduplexes formed between divergent DNA molecules during the last denature and reanneal (heat/cool) cycle of PCR was investigated.

Procedures expected to reduce or eliminate heteroduplex formation were performed. We resolved heteroduplexes into homoduplexes by subjecting a fraction of the amplification products to a single additional round of PCR amplification using fresh Taq polymerase and an excess of primers (Resolution +; Fig. 1B, lane 7) (*22*). Additionally, the number of divergent proviral genotypes amplified was reduced to a single target molecule by serial lowering of the amount of input PBMC DNA serving as first PCR round template (Fig. 1B, lanes 1 to 5) (*23*). Both procedures resulted in the loss of the slowly migrating DNA bands. When the DNA that had been resolved was melted and reannealed (in the presence of EDTA to block polymerase activity), the same series of slowly migrating bands reappeared (Fig. 1B, lane 8). We reasoned that heteroduplexes would be formed more efficiently at a lower primer/templates ratio and at high product concentration, for example, during later rounds of PCR. Accordingly, 25 cycles of amplification in the second round reduced the level of slower migrating DNA (as compared to 35 cycles) (Fig. 1B, lane 9).

Analysis of heteroduplexes formed by denaturing and reannealing mixtures of DNA fragments from divergent HIV-1 *env* genes of known sequence confirmed that distortions of the double helix reduced fragment mobility in polyacrylamide gels (*24, 25*). When two divergent 0.7-kb fragments were mixed and reannealed, nearly comigrating homoduplex bands plus two additional slower migrating bands were observed (Fig. 2A inset), whereas mixtures of three different sequences yielded six heteroduplex bands (*25*). Thus, each possible heteroduplex is formed and the strand-

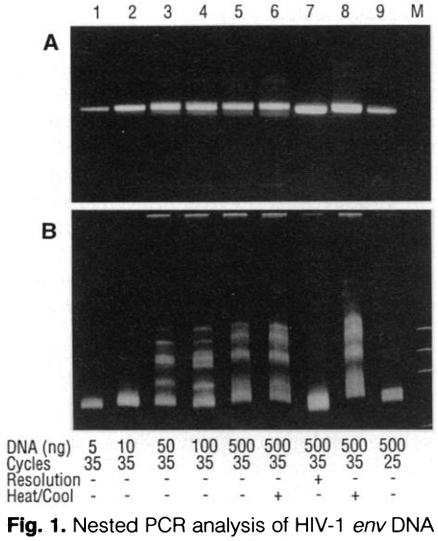

Fig. 1. Nested PCR analysis of HIV-1 *env* DNA from uncultured human PBMC analyzed on a 2.5% agarose (**A**) or a 5% polyacrylamide (**B**) gel. PBMC were isolated and DNA was purified and amplified as described in (*21*). DNA (ng) corresponds to the amount of PBMC DNA in the 35-cycle, first round of PCR. Cycles indicates the number of anneal/extend cycles in the second round of PCR. Resolution, formation of heteroduplexes (Heat/Cool) and conditions of gel electrophoresis are as described (*21, 22*). Molecular weight markers (M) are ϕX174 DNA digested with Hae III.

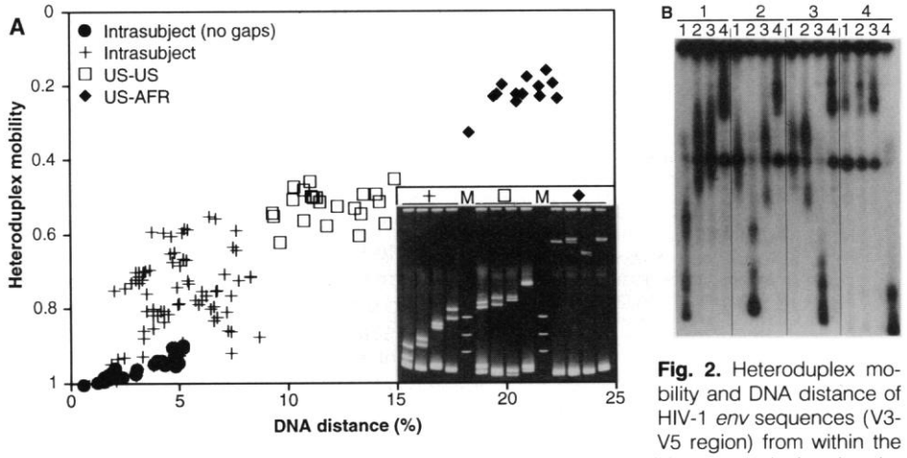

Fig. 2. Heteroduplex mobility and DNA distance of HIV-1 *env* sequences (V3-V5 region) from within the same and between different infected individuals. (**A**) Heteroduplex mobilities were calculated as the average distance of migration of the two heteroduplex bands divided by the distance of migration of the homoduplex bands, with plasmid subclones as starting material for PCR amplification of *env* gene fragments. Genetic distances were determined by the standard method of maximum likelihood in the program DNADIST, counting only mismatches between aligned sequences, discounting unpaired bases caused by insertions and deletions (*20, 27*). Relative mobilities are plotted against divergence for four sets of sequence comparisons: Intrasubject (+) comparisons were made on four sets of sequences, one each from two AIDS patients and two time points from one asymptomatic individual. Intrasubject comparisons for which no gaps appear within heteroduplexes are also shown (●). US-US comparisons (□) were made with six sequences from epidemiologically unlinked individuals from North America [NL4-3, SF2, SF162, MA5 (*29*), and BU01 and PE01 (*34*)]. US-AFR (◆) comparisons were made between two African [NDK, MAL (*29*)] and the North American virus sequences. The NDK-MAL comparison displayed the fastest mobility and least sequence divergence in this group. (**Inset**) Ethidium-stained 5% polyacrylamide gel with representative heteroduplexes formed from the intrasubject, intersubject, and US-AFR groups, respectively. (**B**) Relationships between viral quasispecies from epidemiologically unlinked individuals. PCR amplified viral DNA from the PBMC of four infected individuals (quasispecies) was generated. The ^{32}P-labeled mixture from one individual (upper number) was melted and reannealed in the presence of a 100-fold excess of unlabeled DNA from another individual (lower number) (*21*). Subjects 1 through 3 were infected in North America, subject 4 was infected in Africa. An autoradiograph of these products separated on a 5% polyacrylamide gel is shown. The common band seen approximately half way down the gel is single-stranded DNA.

specific composition of mismatched and unpaired nucleotides (gaps) affects their mobility. Heteroduplexes with mismatched nucleotides but without gaps caused by in-frame insertions and deletions displayed reduced mobilities when the degree of divergence exceeded 1 to 2% and the mobility generally decreased with the degree of mismatch (Fig. 2A, filled circles), whereas a single 3–base pair (bp) gap induced a noticeable mobility retardation in 3.2-kb heteroduplexes (*25*). The greater reduction in mobility caused by gaps relative to mismatched nucleotides probably reflects the sharp bend in the double-stranded DNA molecule required to accommodate the extra nucleotides and the resulting impairment of migration through the gel matrix (*26*).

The corresponding 0.7-kb fragment of *env* was amplified from each of 39 cloned HIV-1 strains of known sequence. These included 31 molecules obtained directly from the PBMC DNA of three North American subjects [one long-term asymptomatic, two with acquired immunodeficiency syndrome (AIDS)], six independent HIV-1 isolates from North America, and two isolates from Zaire. Heteroduplexes were formed by pairwise combination of amplified products from the same individual and between epidemiologically unrelated North American and Zairian isolates. The relative mobility of heteroduplexes (to homoduplexes) was then compared to the known genetic divergence between the fragments (Fig. 2A), the latter determined by the standard method of counting only mismatches between aligned sequences, discounting unpaired bases caused by insertions and deletions (*20*, *27*).

Three groupings were observed by both mobility and sequence analysis: Large mobility shifts and divergence were observed for heteroduplexes formed between Zairian and between Zairian and North American strains; intermediate shifts and divergence were observed for the comparison of independent North American strains; and the smallest shifts and divergence were observed for the comparisons of sequences derived from the same subject (Fig. 2A). Heteroduplexes displaying the full range of mobilities observed are shown in the Fig. 2A inset. For some sequences derived from the long term–infected asymptomatic individual the mobility shifts were atypically high and partially overlapped those observed between independent North American isolates (Fig. 2A). When heteroduplexes differing by only mismatches were analyzed, a better correlation between mobility and divergence was observed, but with a lower slope (Fig. 2A, filled circles). The scatter of values observed in the intrasubject range is thus largely attributable to the large influence of gaps on electrophoretic mobility (gaps not included when DNA distance was determined).

Estimates of genetic distance between virus strains can also be obtained without isolation of individual molecules, by tracking heteroduplexes formed between entire pools of viral genetic variants present within individuals. Given the large affect of gaps on heteroduplex mobility, this may provide a more reliable estimate of overall genetic distance between quasispecies than comparison of any given pair of sequences from these pools. Nested PCR products were derived from PBMC DNA from four individuals containing at least 20 to 50 template molecules of HIV-1 DNA, as measured by end point dilution (*28*)

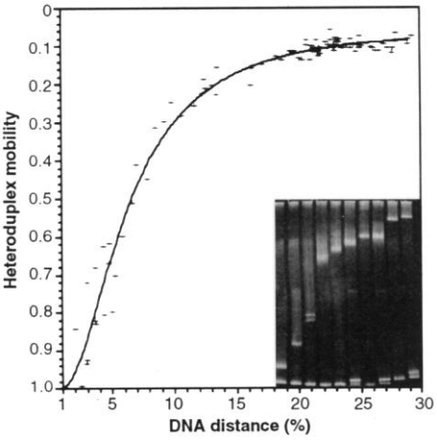

Fig. 3. Correlation between DNA heteroduplex mobilities and DNA distances. Heteroduplexes were generated between PCR fragments derived from the viruses shown in Fig. 4B. Their relative mobilities and DNA distances were determined as described in Fig. 2. Ninety-four of the 190 possible pairwise combinations were evaluated. A curve approximating the relationship shown was calculated by the least-squares method and is given by the equation, DNA distance = −ln[(mobility − 0.045)/1.14]/13.6. Estimated DNA distance values above 30% were not reliable and were kept at that maximum limit. (**Inset**) Ethidium bromide–stained polyacrylamide gel showing the range of heteroduplex mobilities observed in this study.

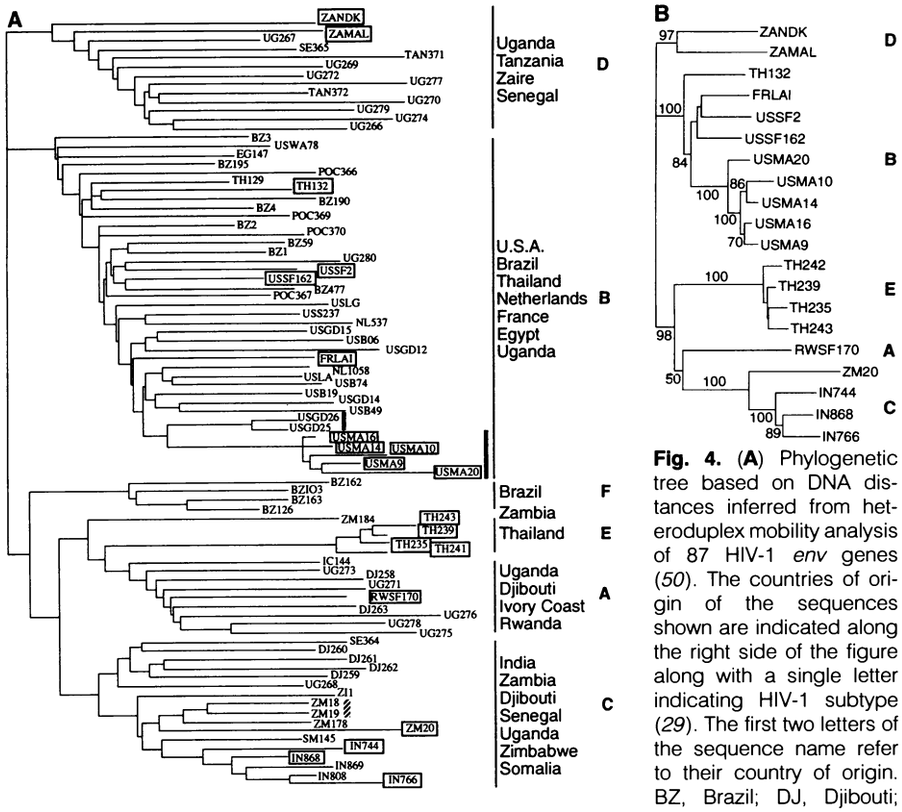

Fig. 4. (**A**) Phylogenetic tree based on DNA distances inferred from heteroduplex mobility analysis of 87 HIV-1 *env* genes (*50*). The countries of origin of the sequences shown are indicated along the right side of the figure along with a single letter indicating HIV-1 subtype (*29*). The first two letters of the sequence name refer to their country of origin. BZ, Brazil; DJ, Djibouti; EG, Egypt; FR, France; IN, India; IC, Ivory Coast; NL, Netherlands; RW, Rwanda; SE, Senegal; SM, Somalia; TA, Tanzania; TH, Thailand; UG, Uganda, US, United States; ZA, Zaire; ZM, Zambia; and ZI, Zimbabwe. POC samples (within subtype B) refer to U.S. armed forces servicemen found to be infected after tours of duty outside the United States. Sequence designations within boxes correspond to those found in the DNA sequence tree shown in Fig. 4B. Solid vertical bars indicate multiple sequences derived from single individuals from the United States. The striped vertical bar at the bottom of the figure indicates sequences derived from sexual partners from Zambia. (**B**) Phylogenetic tree based on DNA distance determined from 20 HIV-1 *env* genes for which sequences are known and subgroups assigned (*29*). The numbers adjacent to branches refer to the frequency (of 100 repetitions) with which the sequences to the right cluster together after bootstrap analysis (*51*).

to ensure a representative sampling of the quasispecies. Radioactively labeled, tracer amounts of amplified products from one quasispecies were mixed, denatured, and annealed with a 100-fold excess of unlabeled, amplified DNA from a second individual. Under these conditions, all radiolabeled DNA strands formed heteroduplexes with unlabeled strands, and the electrophoretic mobility of the radioactive tracer DNA reflects the genetic distance between the two quasispecies. When quasispecies from epidemiologically unlinked individuals within the United States were compared, large mobility shifts were observed (Fig. 2B), as expected from the analysis of individual molecules (Fig. 2A). The mobility shifts were most extreme when samples from the African-acquired infection were hybridized with those from North America. The simultaneous comparison of pools of variants afforded by this heteroduplex tracking analysis (HTA) thus allows estimation of genetic divergence between entire quasispecies.

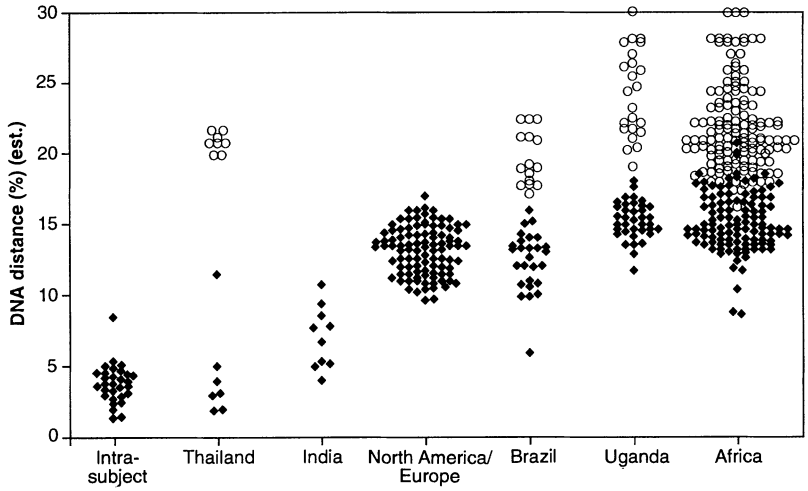

Fig. 5. Genetic diversity of HIV strains within different populations. Pairwise genetic distances were inferred by heteroduplex mobility as described in Fig. 3 and plotted according to the country or continents of origin. Intrasubject corresponds to estimated DNA distances between virus clones obtained from within five infected subjects. (◆) intrasubtype comparisons; (○) intersubtype comparisons.

To develop a quantitative screening assay for determining the genetic relationship between HIV strains, we also used 1.3-kb DNA fragments spanning the five hypervariable regions of the HIV-1 *env* surface glycoprotein gene (*21*). Using the longer fragments, we obtained a more predictable relationship between heteroduplex mobility and genetic distance throughout the range of sequence divergence examined (Fig. 3). A representative gel showing the range of heteroduplex mobilities observed is shown in the Fig. 3 inset. An exponential equation was derived to describe a curve fitting the experimental data from pairwise analysis of 20 envelope genes of known sequence (Fig. 3). This equation was used to estimate the genetic distance between HIV envelope genes from an additional 67 viruses examined by pairwise heteroduplex mobility analysis (HMA). A phylogenetic tree was then constructed from deduced DNA distances (Fig. 4A). All of the major and many of the minor branch nodes were the same as in a tree generated by an analogous distance matrix method and bootstrapping for analysis of the subset of 20 known sequences (Fig. 4B).

Five subtypes of HIV-1 *env* genes have been recognized by DNA sequence analysis as belonging to clusters that differ by more than 20% divergence over a 300-bp region analyzed (*29*). Each of these subtypes also segregated by HMA, and in addition a sixth subtype, referred to as F, was found in Brazil.

Within the group of 67 sequences assigned to subtypes by HMA, 34 of a group of 36 were assigned to the same subtype by anchored PCR typing or DNA sequencing of the viral *gag* gene (*30–32*). In one of the discordant cases, the subtype of the *env* gene determined by sequence analysis from the same isolate did indeed match that determined by HMA. This virus therefore appears to represent a recombinant formed between viruses from two subtypes. In the second discordant case the *env* gene was not sequenced to allow comparison to the HMA result. All of 26 subtype assignments made by *env* gene sequence analysis from the group of 67 (*33*) were assigned to the matching subtype by HMA. Hence, we found 100% correspondence between HIV-1 subtype assignments made by HMA and by DNA sequence analysis.

Among the closest genetic relationships inferred by HMA were, as expected, HIV genes derived from the same infected person (Fig. 4A, the GD25-GD26 and MA groups, respectively). A more distant yet linked relationship was noted for viruses from sexual partners (ZB18 and ZB19, Fig. 4A). It is notable that the four epidemiologically unlinked subgroup E sequences from (Northern) Thailand showed a level of sequence similarity comparable to that obtained from a single individual [Figs. 4 and 5; similar results were obtained for an additional 11 isolates from the same region (*34*)]. In addition, all five epidemiologically unlinked sequences from Bombay, India (subgroup C), were closely related, as were an additional five isolates from Bombay and Goa examined by DNA sequence analysis (*35*). Widespread HIV-1 infection in Thailand and in Bombay occurred within 4 to 5 years from the time these samples were taken, and infection levels in some high-risk groups in these areas have now reached 15 to 50% (*36–38*). The unusually high level of sequence conservation in these groups may reflect the recent and rapid dissemination of HIV from a single source of entry into the Bombay and Northern Thailand populations (that is, a founder effect). The high level of viremia (*39, 40*) together with the low level of viral genetic diversity found early in infection (*25, 41–44*) may contribute to the effective spread of highly conserved viral genotypes in the initial stage of HIV spread in a given geographic region.

Whereas a more extensive molecular epidemiological analysis will be required to substantiate the breadth of subtype representation in individual regions, distinct geographic clustering was noted in multiple instances. All isolates from North America (n = 15) and Europe (n = 3) belonged to subtype B, representatives of which were also found in Brazil (8/12), Thailand (2/6), Egypt (1/1), and Uganda (1/15). Subtype C was found principally along the South and East coast of Africa and the West coast of India, in agreement with a previous link made between HIV-1 strains from South Africa and India (*45*). In contrast, the 36 strains examined from Africa belonged to four subtypes. In addition, isolates related in their *gag* sequence to the E and F *env* subtype isolates found in Thailand and Brazil, respectively, were also found in Africa (*32*).

In the analysis shown in Fig. 5, we plot the pairwise relationships between viruses within a given geographic region. A gradient of increase in divergence is noted, with the level correlating with the time the virus has been thought to have infected each population in large numbers. The predominance of a single, uniformly divergent subtype in North America and Europe may reflect spread from a single point of introduction in past decades followed by a gradual increase in sequence diversity. In contrast, the pattern noted in Thailand reflects the recent introduction of virus from two divergent sources. The divergence noted

within the Brazilian population similarly reflects the presence of two subtypes, but also suggests that the virus has been spreading there for a time similar to its residence in North America. The high degree of viral genetic diversity seen within the African sample sets is consistent with the presence of multiple subtypes and the longer time of residence of HIV-1 on that continent (*46, 47*).

Heteroduplex mobility and tracking analyses provide simple screening tools for determining relationships between HIV-1 strains as well as other variable viruses, microorganisms, and cellular genes. Each can be rapidly classified into sequence homology groups after PCR amplification of variable regions and HMA, comparing unknown sequences against themselves or standard reference sequences. The DNA sequence of genetically common or rare variants could therefore be determined on a selective rather than random basis. On a routine basis the identity and purity of HIV-1 strains and PCR-amplified DNA can be simply and rapidly monitored. Heteroduplex analysis can also be used for tracking specific sequence variants within individuals (*25*) and assisting in establishing epidemiological linkages between individuals (*48*). We demonstrate here that a quantitative assay can be developed for rapid and accurate subtyping useful in evaluating the patterns of infectious agent transmission within large populations. The quantitative assay can also assist identification of commonly circulating strains for temporally and geographically specific vaccine formulations as well as estimating the breadth of the infectious agent challenges likely to be encountered within a vaccine trial setting. The low level of genetic diversity seen in regions of recent HIV-1 spread present rapidly closing windows of opportunity for vaccination against still relatively homologous viral challenge.

REFERENCES AND NOTES

1. J. J. Holland, J. C. De La Torre, D. A. Steinhauer, *Curr. Top. Microbiol. Immunol.* **176**, 1 (1992).
2. E. Domingo, *Curr. Opin. Genet. Dev.* **2**, 61 (1992).
3. A. Meyerhans *et al.*, *Cell* **58**, 901 (1989).
4. P. Simmonds, P. Balfe, C. A. Ludlam, J. O. Bishop, A. J. L. Brown, *J. Virol.* **64**, 5840 (1990).
5. L. A. Brown and P. Monaghan, *AIDS Res. Hum. Retroviruses* **4**, 399 (1988).
6. M. Goodenow *et al.*, *J. Acquired Immune Defic. Syndr.* **2**, 344 (1989).
7. P. Balfe, P. Simmonds, C. A. Ludlam, J. O. Bishop, A. J. L. Brown, *J. Virol.* **64**, 6221 (1990).
8. J. Albert *et al.*, *AIDS* **4**, 107 (1990).
9. R. E. Phillips *et al.*, *Nature* **354**, 453 (1991).
10. M. H. St Clair *et al.*, *Science* **253**, 1557 (1991).
11. M. Arendrup *et al.*, *J. Acquired Immune Defic. Syndr.* **5**, 303 (1992).
12. J. A. McKeating *et al.*, *AIDS* **3**, 777 (1989).
13. J. M. Coffin, *Curr. Top. Microbiol. Immunol.* **176**, 143 (1992).
14. M. P. Cranage *et al.*, *AIDS Res. Hum. Retroviruses* **9**, 13 (1993).
15. R. Cheingsong-Popov *et al.*, *J. Infect. Dis.* **165**, 256 (1992).
16. C. J. Issel *et al.*, *J. Virol.* **66**, 3398 (1992).
17. K. B. Mullis and F. A. Faloona, *Methods Enzymol.* **155**, 335 (1987).
18. D. D. Ho *et al.*, *Ann. Intern. Med.* **103**, 880 (1985).
19. D. D. Ho *et al.*, *N. Engl. J. Med.* **313**, 1493 (1985).
20. K. Kusumi *et al.*, *J. Virol.* **66**, 875 (1992).
21. Peripheral blood mononuclear cells were isolated by Ficoll-Hypaque density gradient centrifugation, and DNA was extracted with the IsoQuick isolation kit (MicroProbe Corp., Garden Grove, CA). PCR used a nested series of reactions, with 2 μl of the first reaction product added to a second round of PCR with internally annealing primers. Positions correspond to those from the IIIB isolate BH10 sequence [L. Ratner *et al.*, *Nature* **313**, 277 (1985)]. First round primers were ED3 (5′-TTAGGCATCTCCTATGGCAGGAAGAAGCGG at positions 5537 to 5566) and ED14 (5′-TCTTGCCTGGAGCTGCTTGATGCCCCAGAC at positions 7509 to 7538). The second round primers were ES7 (5′-TGTAAAACGACGGCCAGTCTGTTAAATGGCAGTCTAGC, corresponding to the complement of the M13 forward sequencing primer followed by positions 6579 to 6598) and ES8 (5′-CAGGAAACAGCTATGACC-CACTTCTCCAATTGTCCCTCA, corresponding to the complement of the M13 reverse sequencing primer followed by the complement of positions 7225 to 7245), yielding a product of approximately 700 bp of which approximately 627 bp are target dependent (the exact size depending on the number of deletions and insertions within the target HIV-1 molecule). Second round primers ED5 (ATGGGATCAAAGCCTAAAGCCATGTG at positions 6134 to 6159) and ED12 (AGTGCTTCCTGCTGCTCCCAAGAACCCAAG at positions 7359 to 7388) yielded a template-dependent product of approximately 1250 bp. Each PCR reaction used variable amounts of template DNA (up to 1 μg), 1.25 mM (ED3-ED14 and ED5-ED12) or 1.8 mM (ES7 and ES8) $MgCl_2$, 0.2 μM of each primer in 50 mM KCl, 10 mM tris-HCl (pH 8.3), 200 μM of each deoxynucleotide triphosphate (dNTP), 2.5 U of Taq DNA polymerase (Perkin-Elmer Cetus, Emeryville, CA), and 10% glycerol in a final volume of 50 μl. In some experiments, PCR products were radiolabeled by addition of 10 μCi of [^{32}P]deoxycytidine triphosphate (dCTP), with only 30 μM of each dNTP in the second round of nested PCR. Amplifications were carried out in a Perkin-Elmer Thermocycler for 25 to 35 cycles with 1-s ramp times between steps of 94°C for 1 min, 55°C for 45 s, and 72°C for 1 min. To demonstrate single molecule template sensitivity, we used pNL4-3 (HIV-1 Lai) plasmid DNA [A. Adachi *et al.*, *J. Virol.* **59**, 284 (1986)] and ACH-2 cell DNA (containing a single defective HIV-1 genome) [T. M. Folks *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2365 (1989)] as control templates in the presence of a total of 1 μg of human genomic DNA. The DNA fragments were separated in 5% polyacrylamide gels (30:0.8 acrylamide:Bis) in TBE buffer (0.088 M tris-borate, 0.089 M boric acid, 0.002 M EDTA) at 250 V for 3 hours (for the 700-bp fragments) or 70 mA for 1100 volt-hour (for the 1200-bp fragments) in a BRL V16 vertical gel apparatus. It should be noted that mobilities of heteroduplexes are strongly affected by the high gel temperature reached (approximately 60°C) during electrophoresis at 70 mA. Agarose gels (2.5%) were run at 100 V for 1.5 hours in TAE buffer (0.04 M tris-acetate, 0.001 M EDTA).
22. Heteroduplex resolution was carried out by transfer of 10 μl of the second round PCR reaction to 90 μl of fresh standard PCR reaction mix (with 100 pmol of each primer) followed by 1 min at 94°C, 1 min at 57°C, and a single 5-min extension at 72°C.
23. P. Simmonds *et al.*, *J. Virol.* **64**, 864 (1990).
24. For formation of heteroduplexes, 4.5 μl from two second round PCR reactions were combined and 1 μl of 10× annealing buffer added [1 M NaCl, 100 mM tris-HCl (pH 7.8), 20 mM EDTA]. DNA was denatured at 94°C for 2 min and annealed by rapid cooling on ice.
25. E. L. Delwart, H. W. Sheppard, B. D. Walker, J. Goudsmit, J. I. Mullins, in preparation.
26. C. H. Hsieh and J. D. Griffith, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4833 (1989).
27. J. Felsenstein, *Annu. Rev. Genet.* **22**, 521 (1988).
28. Estimates of proviral DNA load in cells is necessary to ensure that multiple copies of proviral DNA are being co-amplified. Semi-quantitative estimates of HIV-1 provirus copy number were done by duplicate serial fivefold end point dilution of PBMC DNA. The lowest concentration of infected cell DNA to yield a positive PCR signal was used to estimate the input proviral DNA load.
29. G. Myers, B. Korber, J. A. Berzofsky, R. F. Smith, G. N. Pavlakis, Los Alamos National Laboratory, Human Retroviruses and AIDS Database (1992).
30. F. E. McCutchan *et al.*, *J. Acquired Immune Defic. Syndr.* **5**, 441 (1992).
31. J. Louwagie *et al.*, *AIDS Res. Hum. Retroviruses* **8**, 1467 (1992).
32. J. Louwagie *et al.*, *AIDS* **7**, 769 (1993).
33. F. E. McCutchan, J. Louwagie, E. L. Delwart, J. I. Mullins, unpublished data.
34. E. L. Delwart *et al.*, unpublished data.
35. M. Grez *et al.*, in preparation.
36. H. Rubsamen-Waigmann *et al.*, *Lancet* **337**, 550 (1991).
37. F. E. McCutchan *et al.*, *AIDS Res. Hum. Retroviruses* **8**, 1887 (1992).
38. R. Singal, P. Gilada, J. I. Mullins, unpublished observations.
39. E. S. Daar, T. Moudgil, R. D. Meyer, D. D. Ho, *N. Engl. J. Med.* **324**, 961 (1991).
40. S. J. Clark *et al.*, *ibid.*, p. 954.
41. S. Pang *et al.*, *AIDS* **6**, 453 (1992).
42. L. Q. Zhang *et al.*, *J. Virol.* **67**, 3345 (1993).
43. T. F. Wolfs, G. Zwart, M. Bakker, J. Goudsmit, *Virology* **189**, 103 (1992).
44. T. Zhu *et al.*, *Science* **261**, 1179 (1993).
45. U. Dietrich *et al.*, *AIDS* **7**, 23 (1993).
46. A. Srinivasan *et al.*, *AIDS Res. Hum. Retroviruses* **5**, 121 (1989).
47. G. Myers, K. MacInnes, B. Korber, *ibid.* **8**, 373 (1992).
48. E. L. Delwart, M. L. Kalish, M. Busch, L. Romero, J. I. Mullins, in preparation.
49. J. Felsenstein, *Cladistics* **5**, 164 (1989).
50. Approximate DNA distances were calculated from the equation described in Fig. 3 for each of the total of 794 pairs of heteroduplexes generated using molecular end points or plasmid subclones (21% of the total possible combinations). A triangle matrix of these distances was generated and a phylogenetic tree constructed with the program FITCH (*49*), with global rearrangements to improve branching order.
51. SEQBOOT was used to generate 100 unique sets of DNA sequences from the input of aligned sequences, each differing by a minor reshuffling of a portion of the nucleotides (*49*). Because of the requirement for the use of a DNA distance matrix for analysis of heteroduplex mobility data (versus aligned sequences when DNA sequences are known), we then generated DNA sequence–based trees using a method parallel to that described in Fig. 4A. DNADIST and FITCH were used to generate 100 trees and CONSENSE used to find the consensus tree and the frequencies of branch clusters. We calculated branch lengths for the consensus tree using FITCH. Without the use of SEQBOOT to shuffle input sequences, the entire branching order shown was preserved in each of 100 repetitions of FITCH (*34*).

52. We thank B. D. Walker, M. S. Hirsch, H. W. Sheppard, M. Busch, A. Mayer, C. Cheng-Mayer, E. Sabino, M. Morgado, F. Bex, J. Goudsmit, D. A. Katzenstein, M. Holodniy, L. Corey, J. Levy, J.-C. Chermann, and the AIDS Research and Reference Reagent Program for some of the viral DNA samples used in this study. We also thank J. Sninsky and Roche Molecular Systems for the gift of Taq polymerase and G. Myers and B. Korber for comments on the manuscript. Supported by grants from the Public Health Service (AI85007 and AI32885), the Pediatric AIDS Foundation (500170-11-PG), the Stanford Program in Molecular and Genetic Medicine, and the World Health Organization (A20/181/305). E.L.D. was supported during the early stages of this study by NIH National Research Service Award AI07328.

8 June 1993; accepted 27 September 1993