Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 7365–7369, August 1992
Genetics

Development of a quasispecies of human immunodeficiency virus type 1 *in vivo*

(genetic variability/virus evolution/V3 loop)

K. CICHUTEK*, H. MERGET*, S. NORLEY*, R. LINDE†, W. KREUZ†, M. GAHR‡, AND R. KURTH*§

*Paul-Ehrlich-Institut, Paul-Ehrlich-Strasse 51-59, D-6070 Langen, Federal Republic of Germany; †Centre for Pediatrics, University Clinics, Theodor-Stern-Kai 7, D-6000 Frankfurt/Manto, Federal Republic of Germany; and ‡University Children's Clinic, Robert-Koch-Strasse, 40, D-3400 Göttingen, Federal Republic of Germany

Communicated by Maurice R. Hilleman, April 27, 1992 (received for review January 8, 1992)

ABSTRACT **During treatment with one specific batch of blood clotting factor IX, a number of hemophilia B patients in Germany recently became infected with human immunodeficiency virus type 1 (HIV-1). The nucleotide sequences of cloned HIV-1 envelope gene regions including the variable V3 loop and the V4 region derived from short-term virus cultures and directly from peripheral blood cells of these patients were shown to be highly homologous. Based on the assumption that the corresponding consensus sequence (termed HIV-1_{MBK}) was identical to the genotype of the initially infecting virus, we were able to construct phylogenetic trees of the developing quasispecies in two patients studied in detail. True intermediates between input and multiply mutated genotypes were found in individual blood samples. Except for the initially infecting variant HIV-1_{MBK}, variants found at 11 months postinfection had replaced those seen at 5 months postinfection. Variability early after infection was shown to cluster in two small regions located 3′ of the V3 loop (i.e., outside the loop) and within the V4 region. This communication therefore describes the evolution of an HIV-1 quasispecies in humans starting from a single genotype.**

Like all lentiviruses, human immunodeficiency virus type 1 (HIV-1) is not present in a patient as a single wild-type virus but as a virus population of variants constantly undergoing mutation. By analysis of HIV-1 subgenomic regions in the *gag*, *tat*, and *env* genes after polymerase chain amplification (PCR) and sequencing of individual molecular clones, multiple virus variants have been shown to exist in a virus population at a specific time after natural infection (1–5). Variability was mainly restricted to the variable regions V1 to V5 of the HIV-1 *env* gene coding for the external glycoprotein gp120 (5–10).

We previously reported that 9 out of 10 HIV-1 variants from three infected hemophiliacs were homologous in the V3 loop (11). These patients had become infected during treatment with an identical contaminated batch of factor IX in 1989–1990. Of a total of at least 48 hemophiliacs known to have been treated with this batch, 9 seroconverted, suggesting a very low load of contaminating HIV. From the frequency of seroconversions and the doses of clotting factor received, one can calculate that the batch contained between 0.2 and 3 human infectious doses of HIV-1 per liter. If so, statistically each infected patient received a single infectious particle, i.e., a biologic clone. Not very surprisingly, we were able neither to rescue HIV-1 directly from the remaining 200 ml of the batch nor to amplify HIV DNA or RNA by PCR from small (10 μl) clotting factor volumes. The batch was p24 HIV-1-antigen positive. Here we provide a detailed study of variant HIV-1 genotypes analyzed directly from peripheral blood mononuclear cell (PBMC) DNA or obtained after short-term (i.e., 4 weeks) culture in cord blood lymphocytes (CBLs).

MATERIALS AND METHODS

Blood and Culture Samples. Fresh citrate-treated blood from patients M and K in Frankfurt and patient B in Göttingen was used for isolation of PBMCs by Ficoll gradient centrifugation. For virus isolation, primary cells were cocultivated with interleukin 2/phytohemagglutinin-stimulated CBLs and kept in culture for up to 8 weeks (11).

DNA Preparation and PCR Amplification. DNA was prepared from 10^6 cells and PCR-amplified by published procedures (11, 12). One to 4 μg of PBMC DNA was added to a PCR mixture containing (final concentrations) 500 mM KCl, 2.5 mM $MgCl_2$, gelatin (200 μg/ml) and 100 mM Tris/HCl (pH 8.0), 800 μM dNTPs, 1.5 units of *Taq* polymerase (Perkin-Elmer), 1 μM oligonucleotide primers, and water to make 100 μl. Five precycles (94°C, 45 sec; 37–45°C, 2 min; 72°C, 3 min) and 40 main cycles (94°C, 45 sec; 55–60°C, 2 min; 72°C, 3 min) were run. Ten microliters of the reaction product was added to 90 μl of a second PCR mixture and amplified under the conditions given above. For the first round of PCR amplification, primers were HI-env-6430(+) (5′-TGT-CCA-AAG-GTA-(T/A)(C/A)T-TTG-A-3′) and HI-env-7621(−) (5′-GTT-TTC-CAG-AGG-AAC-CCC-3′). Primers used for the second amplification round were HI-SalI-env-6447(+) (5′-CG-ATC-GTC-GAC-TTT-GA(G/T)-CCA-ATT-CCC-ATA-CA-3′) and HI-XhoI-env-7578(−) (5′-CG-ATC-CTC-GAG-TAG-GTA-TCT-TTC-CAC-AGC-CAG-3′). The oligonucleotide primers were chosen from conserved sequences of the *env* gene.

To check the mutation rate introduced by the PCR amplification procedure, we diluted plasmid pNL4-3 (8) to 10^5 molecules and PCR-amplified and sequenced 10 molecular clones as described above. None of the 10 clones showed deletions, and the error rate of *Taq* polymerase was found to be 0.1%. Therefore, the vast majority of mutations detected during the analysis of HIV-1 variants in the patients' blood samples were authentic—i.e., not due to the PCR amplification.

Molecular Cloning and Sequencing. PCR-amplified products were digested with *Sal* I and *Xho* I and separated by size in a 1.2% agarose gel (11). The specific 1.1-kilobase-pair fragment was purified (Qiaex, Stratagene) and ligated with *Sal* I/*Xho* I-digested plasmid vector pBluescript (Stratagene) or pUC (New England Biolabs). For sequencing, plasmid DNA was purified from minipreparations (Qiagen, Studio City, CA) and primers HI-SalI-env-6447(+) (see above) and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV, human immunodeficiency virus; SIV, simian immunodeficiency virus; PBMC, peripheral blood mononuclear cell; CBL, cord blood lymphocyte.

§To whom reprint requests should be addressed.

HI-env-merk1(−) (5′-TGG-GAG-GGG-CAT-ACA-TTG-CTT-3′) were used in a standard sequencing reaction.

RESULTS

Definition of the Initially Infecting Genotype HIV-1_{MBK} in PBMC DNA of Patients M and B. To investigate HIV-1 quasispecies development directly—i.e., without the possible influence of selective pressure in tissue culture—fresh PBMCs from patients M and B were purified for DNA isolation followed by PCR amplification and molecular cloning, allowing nucleotide sequencing of the region starting 5′ of the V3 loop and extending into V4 of HIV-1 *env*. Fig. 1 illustrates the very high amino acid sequence homology of the 54 sequenced HIV-1 *env* regions of the variants detected 5 months (clones B5 and M5) and 11 months (clones M11) after infection. A consensus nucleotide and amino acid sequence (HIV-1_{MBK}) could easily be derived by comparing the bases at each nucleotide position, as all bases were identical at the vast majority of positions. These results further suggest that

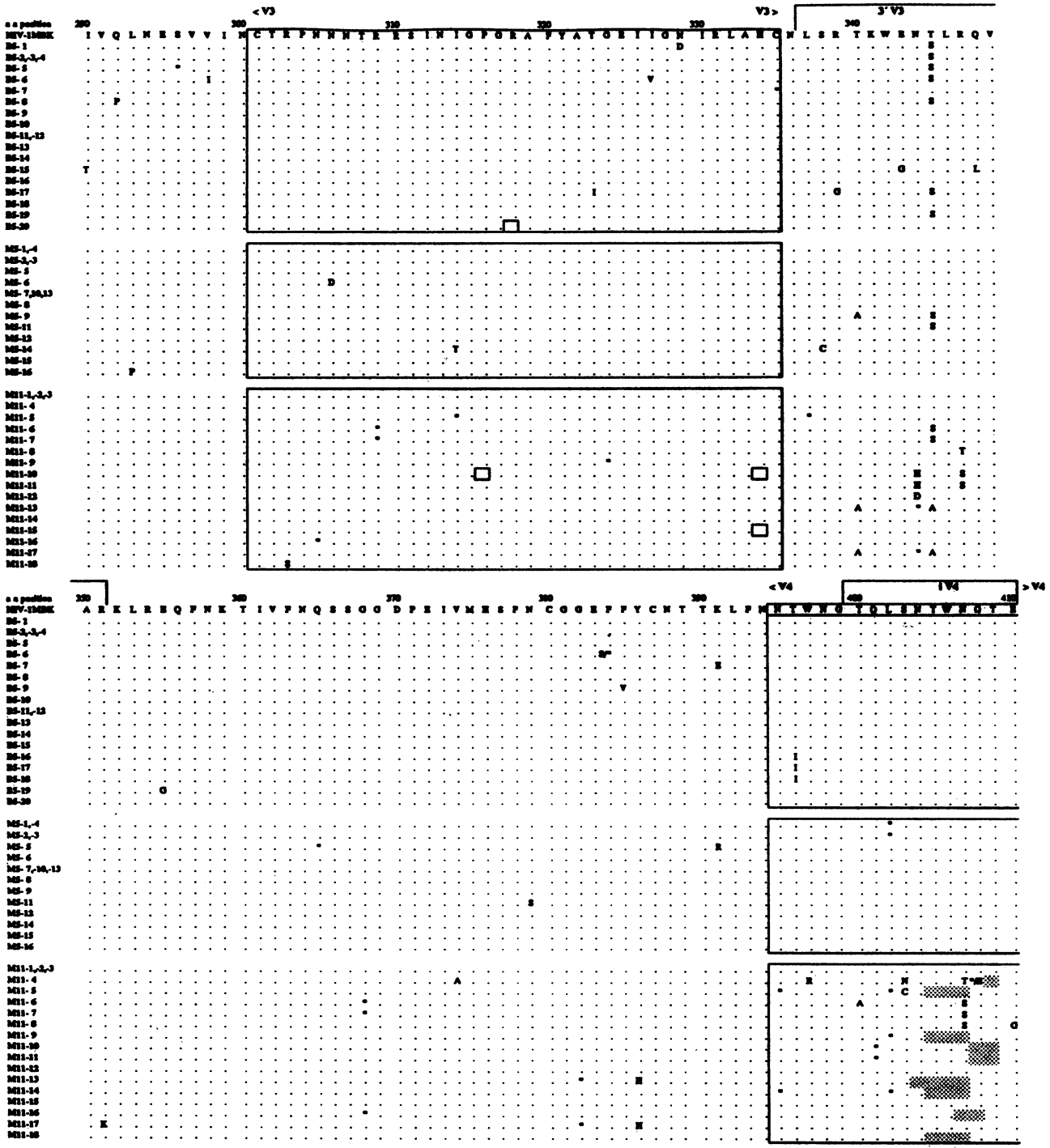

FIG. 1. Amino acid sequence comparison of HIV-1_{MBK} and variants in human PBMCs. Shown are independently cloned subgenomes of HIV-1 prepared from PBMC DNA of patients B and M 5 months postinfection (clones B5-1 to -20 and M5-1 to -16) and of patient M 11 months postinfection (clones M11-1 to -18). Numbering starts with amino acid (aa) 227 of the HIV-1_{LAI} *env* gene (8) and continues according to the HIV-1_{MBK} consensus sequence. A representative sequence comparison is shown for amino acids 290–410. Complete sequence data are available from the authors upon request. Small dots represent amino acids identical with the consensus sequence HIV-1_{MBK}. Amino acid changes are depicted for each single clone. Larger dots indicate silent mutations. Stippled boxes, in-frame deletions; open boxes, stop codons.

the hemophilia patients M and B (and K, see below) were initially infected by a biologic HIV-1 clone.

The HIV-1 quasispecies present in patient B 5 months after infection (HIV-1_{B5}) displayed >99% identity to the HIV-1_{MBK} consensus sequence (Fig. 1, clones B5-1 to -20). The subgenomic clones B5-11 and -12 were identical to HIV-1_{MBK} and nine additional clones differed only by a single base change. Three of these clones (B5-2, -3, and -4) were identical to each other. Clones were therefore highly homologous and no major variants were detected. Two subgenomes, B5-8 and -20, were defective within the analyzed region due to single nucleotide deletions.

Similar to HIV-1_{B5}, the virus population of patient M 5 months postinfection (HIV-1_{M5}; clones M5-1 to -16) exhibited an overall nucleotide sequence identity with HIV-1_{MBK} of ≥99.5%. Clones M5-7, -10, and -13 were identical to HIV-1_{MBK}. M5-1 and -4 were identical to each other, as were M5-2 and -3.

To examine whether major populations fluctuate over time as has been postulated (3, 4, 7), we analyzed in detail the HIV-1 population in the blood of patient M 11 months after infection. With the exception of three subgenomes identical to HIV-1_{MBK} (M11-1 to -3), each clone was different from the others but still displayed 98.6% nucleotide sequence identity with HIV-1_{MBK}. Two subgenomes, M11-10 and -15, were apparently rendered defective by single base deletions. Eleven of 18 variants displayed more than three mutations. In comparison, all 16 subgenomes analyzed from the blood of the same patient at 5 months postinfection had three or fewer mutations. Thus, infection by HIV-1_{MBK} resulted in an increasing number of mutations within the analyzed *env* gene region over time.

High Homology of Cultured Viruses from Patients M, B, and K. In Fig. 2, the nucleotide sequence of three independently cloned subgenomes of cultured virus from patients M (HIV-1_M), B (HIV-1_B), and K (HIV-1_K) at 11 months postinfection are compared with the nucleotide sequence of HIV-1_{LAI}. All three isolates were classified as subtypes belonging to the North American/European class of HIV-1 but represented viruses distinct from the known prototypes HIV-1_{MN} and HIV-1_{LAI} (8). Except for four nucleotide changes at positions 727 (G → A), 762 (T → C), 942 (A → T), and 1011 (T → C) in clone HIV-1_M, one nucleotide change at position 1216 (T → C) in clone HIV-1_B, and one nucleotide change at position 958 (T → C) in clone HIV-1_K, all subgenomes were identical to one another. The sequences within the V3 loop of the three subgenomes differed by only one nucleotide change each in HIV-1_M and HIV-1_K, whereas HIV-1_B was completely homologous to the HIV-1_{MBK} consensus genotype.

Detection of Intermediate Genotypes. Fig. 3, a phylogenetic tree constructed by the maximum-parsimony procedure, shows the evolutionary relationship of the subgenomic clones from patients B and M, assuming initial infection with HIV-1_{MBK}. Mutations leading to premature stop codons within the subgenomes were considered to be terminal mutations within the molecular evolution of a subgenome. Mutations common to several clones were taken as mutations in early precursor genotypes. Clones with identical nucleotide sequence within the analyzed subgenomic region were considered to represent the same virus variant. The variants represented by clones B5-18 and B5-2 to -4 could be classified as true intermediate variants from which one and six other genotypes, respectively, must have evolved. The evolutionary tree of HIV-1_{M5} is shown in Fig. 3. Again, clones M5-2 and -3 belonged to an intermediate genotype which subsequently suffered a one-base deletion resulting in a defective variant represented by the two identical subgenomes M5-1 and -4. In the HIV-1_{M11} population, variants M11-7, -11, and -14 belonged to intermediate genotypes appearing during the development of the variants represented by clones M11-6, -10, and -5, respectively. To our knowledge, this is the first detection of evolutionary intermediates in HIV-1 quasispecies development.

Early Genetic Variation of HIV-1 *in Vivo* Is Preferentially Located in Two Regions 3′ of V3 and Within V4. From comparison of all clones (Fig. 1), it was apparent that few mutations were found within the V3 loop. Instead, mutations clustered within two small areas (particularly in the HIV-1_{M11} population): a small region 3′ of the V3 loop (amino acids 337–351; nucleotides 1009–1051) and a region within V4 (amino acids 400–410; nucleotides 1198–1230). We termed these regions 3′V3 and iV4. Within region 3′V3, 8 of 10 (80%) of all mutations were nonsilent, as were 12 of 15 (80%) of the mutations in iV4 (Fig. 3). In contrast, only 65% of all mutations over the complete HIV-1_{M11} *env* region analyzed were nonsilent. Therefore, selection of early virus variants with altered amino acid sequence could be found only for regions 3′V3 and iV4.

DISCUSSION

Molecular cloning of biologically active lentiviruses [e.g., the simian immunodeficiency virus (SIV) strains SIV_{mac} and SIV_{agm}] has recently made it possible to analyze the *in vivo* development of genetic variability starting from a single clone (12, 13). This communication describes a clonal infection of humans allowing a similar investigation of HIV-1 evolution.

Studies using cloned SIV revealed that closely related genotypes evolved within months from a single parental genotype in the inoculated animals. The overall fixation rate of mutations within the SIV *env* genes was shown to be 7.4 mutations per kilobase per year for SIV_{mac239} in rhesus macaques (13) and 7.7 mutations per kilobase per year for SIV_{agm3} in African green monkeys (12). The overall fixation rate calculated from virus population HIV-1_{M11} is 5.6 mutations per kilobase per year, comparable with the fixation rate reported for the simian lentiviruses.

In humans, previous studies following seroconversion in blood transfusion cases and during treatment with contaminated preparations of blood clotting factors indicated very little sequence heterogeneity in infected recipients (14, 15). In the only other quantitative analysis performed (15), HIV-1 quasispecies in patients infected with the same contaminated batch revealed a nucleotide divergence of only 4.2% within the V3 loop. The mean nucleotide distance between patients with unrelated infections was 17% within the same *env* gene region. Unfortunately, the study was based mainly on blood samples obtained ≥3 years after infection, when intrapatient diversity of HIV-1 had already increased significantly. In the study reported here, the mean nucleotide distance of the V3 loops of HIV-1_{MBK} variants was <1% and <2%, respectively, 5 and 11 months after infection.

The genetic variability of HIV-1_{MBK} subgenomes was extremely low within the *env* region analyzed, a region that comprises one of the most variable regions of the whole viral genome (2, 7, 8). Subgenomes completely or almost identical in nucleotide sequence over the entire region were detected both in virus reisolates and in PBMC DNA analyzed directly from the patients. The variants were therefore almost identical to the consensus sequence defined by comparing all subgenomes. It is therefore highly likely that the consensus nucleotide sequence HIV-1_{MBK} represents the genotype of the virus that initially infected patients M, B, and K, supporting the hypothesis that infection occurred via a clonal HIV-1 substrain. Because lentivirus genotypes may differ after each single round of replication, such a substrain in the insufficiently inactivated batch of factor IX must have consisted of a single predominant and very homologous population with at most one nucleotide change in the 552 nucleotides within the constant and variable envelope regions analyzed (Fig. 2).

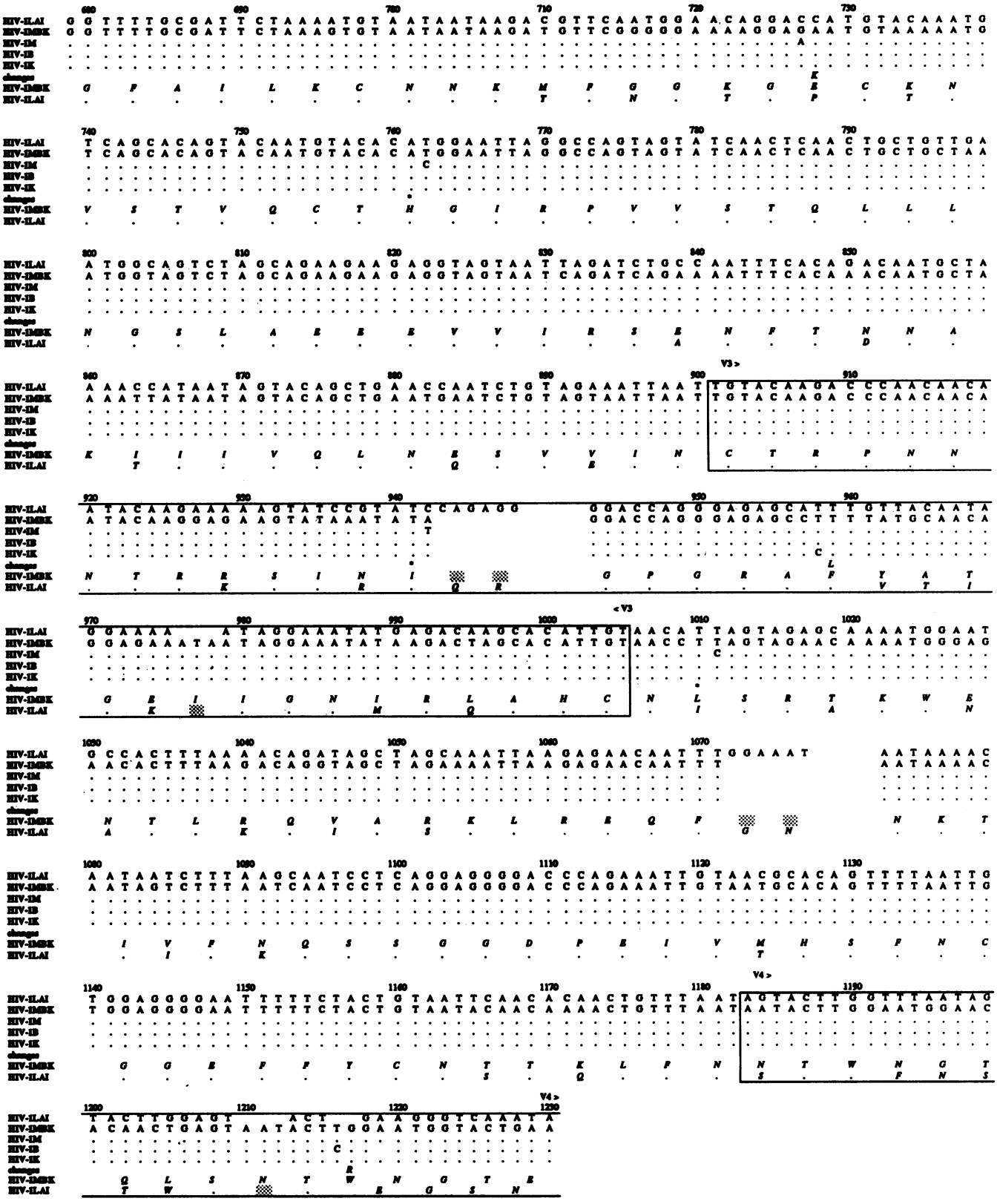

FIG. 2. Nucleotide and amino acid (italic type) sequence comparison of cultured HIV-1_{MBK} variants. Independently cloned subgenomes of HIV-1 variants from PBMCs of patients B, M, and K isolated 11 months postinfection and grown in CBL are termed HIV-1_M, HIV-1_B, and HIV-1_K, respectively. Numbering starts with nucleotide 679, the first base of codon 227 (see Fig. 1 and ref. 8). Dots represent nucleotides identical with the consensus sequence HIV-1_{MBK}. Nucleotide and amino acid changes are shown. Large dots, silent mutations; stippled boxes, in-frame deletions.

HIV-1_{MBK} showed an overall frequency of 80% of nonsynonymous mutations within the early variable regions 3'V3 and iV4, indicative of selection of altered virus variants during HIV-1_{MBK} replication *in vivo*. Selection pressure on SIV evolution was comparably strong, as 81% of the mutations within the variable regions of SIV_{mac} *env* and 92% of changes within the variable regions of SIV_{agm} *env* were nonsynonymous.

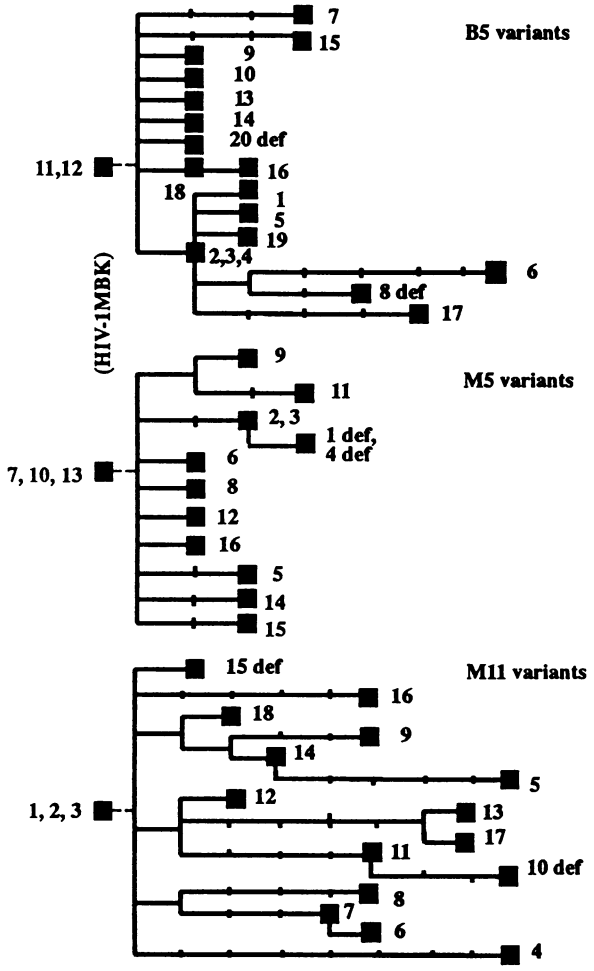

FIG. 3. Evolution of the quasispecies of HIV-1 in patients M and B. Assuming initial infection with consensus genotype HIV-1_{MBK}, evolutionary trees were constructed by comparing independent subgenomic clones at 5 and 11 months postinfection for patients M and B. Mutations common to several clones were assumed to have taken place before divergence of these clones. Mutations listed by nucleotide position are available from the authors upon request. Filled squares represent genotypes of one or more clones (numbers indicated).

In contrast to SIV_{agm} (12), development of HIV-1_{MBK} variants early after infection was not localized in regions previously defined as hypervariable (7, 9). The high variability of the HIV-1 *env* gene variable regions previously defined by comparison of various isolates from different patients and geographic areas may require longer time periods to develop. Clusters of preferred heterogeneity early after infection were located in a region 3′ of the V3 loop and within a short segment of the variable region V4. The high number of subgenomes with single in-frame deletions within V4 in one patient 11 months postinfection was surprising. However, in-frame deletions within variable regions have also been reported for HIV-1 populations in the blood of naturally infected patients (16) and in African green monkeys infected with molecularly cloned SIV_{agm} (12).

We thank C. Garber and M. Popp for excellent technical assistance, B. Brandi for competent editorial assistance, and M. Vogel, M. Lang, M. Dittmar, and I. Treinies for constructive discussions. This work was supported by the Federal Ministry of Health (F.R.G.).

1. Goodenow, M., Huet, T., Saurin, W., Kwok, S., Sninsky, J. & Wain-Hobson, S. (1989) *J. AIDS* **2**, 344–352.
2. Li, Y., Kappes, J. C., Conway, J. A., Price, R. W., Shaw, G. M. & Hahn, B. H. (1991) *J. Virol.* **65**, 3973–3985.
3. Martins, L. P., Chenciner, N., Asjö, B., Meyerhans, A. & Wain-Hobson, S. (1991) *J. Virol.* **65**, 4502–4507.
4. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjö, B. & Wain-Hobson, S. (1989) *Cell* **58**, 901–910.
5. Smith, T. F., Srinivasan, A., Schochetman, G., Marcus, M. & Myers, G. (1988) *Nature (London)* **333**, 573–575.
6. Starcich, B. R., Hahn, B. H., Shaw, G. M., McNeely, P. D., Modrow, S., Wolf, H., Parks, E. S., Parks, W. P., Josephs, S. F., Gallo, R. C. & Wong-Staal, F. (1986) *Cell* **45**, 637–648.
7. Modrow, S., Hahn, B. H., Shaw, G. M., Gallo, R. C., Wong-Staal, F. & Wolf, H. (1987) *J. Virol.* **61**, 570–578.
8. Myers, G., Smith, R. F., Berzofsky, J. A. & Pavlakis, G. N., eds. (1991) *Human Retroviruses and AIDS* (Theoretical Biology and Biophysics Group, Los Alamos Natl. Lab., Los Alamos, NM).
9. Nara, P. L., Smit, L., Dunlop, N., Hatch, W., Merges, M., Waters, D., Kelliher, J., Gallo, R. C., Fischinger, P. J. & Goudsmit, J. (1990) *J. Virol.* **64**, 3779–3791.
10. LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dressman, G. R., Boswell, R. N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E. A. & Putney, S. (1990) *Science* **249**, 932–935.
11. Cichutek, K., Norley, S., Linde, R., Kreuz, W., Gahr, M., Löwer, J., von Wangenheim, G. & Kurth, R. (1991) *AIDS* **5**, 1185–1187.
12. Baier, M., Dittmar, M. T., Cichutek, K. & Kurth, R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 8126–8130.
13. Burns, D. W. & Desrosiers, R. C. (1991) *J. Virol.* **65**, 1843–1854.
14. McNearney, T., Westervelt, P., Thielan, B. J., Trowbridge, D. B., Garcia, J., Whittier, R. & Ratner, L. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1917–1921.
15. Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O. & Brown, A. J. L. (1990) *J. Virol.* **64**, 6221–6233.
16. Simmonds, P., Balfe, P., Ludlam, C. A., Bishop, J. O. & Brown, A. J. L. (1990) *J. Virol.* **64**, 5840–5850.