Cousin ($n = 14$), eight weeks after reaching independence; compared with 23.3 months on Cousin ($n = 93$, $P < 0.001$). The mean age at which young bred was 8.1 months on Aride ($n = 61$) and 8.3 months on Cousin ($n = 4$), again much sooner than on Cousin (47.3 months, $n = 44$, $P < 0.001$). None of the 105 young, old enough to be a helper (91 and 14, respectively) acted as a helper on their natal territories. It is very clear then that habitat saturation is the pre-eminent cause of cooperative breeding on Cousin.

Two years after the transfer to Aride, and a little more than one year after the transfer to Cousin, all high-quality areas on both islands have become occupied and young birds born on high-quality territories have begun to stay as helpers, even though there is still abundant space for them in lower-quality areas to establish territories (refs 11 and 14, respectively). It can be concluded that habitat saturation and territory quality are both involved in the evolution of cooperative breeding. That territory quality should be important to dispersal options by offspring was also emphasized by the 'marginal habitat' model[15], a refined version of the 'habitat saturation' hypothesis. □

Received 23 March; accepted 26 June 1992.

1. Woolfenden, G. E. & Fitzpatrick, J. W. *The Florida Scrub Jay: Demography of a Cooperative Breeding Bird* (Princeton University Press, New Jersey, 1984).
2. Brown, J. L. *Helping and Communal Breeding in Birds* (Princeton University Press, New Jersey, 1987).
3. Emlen, S. T. *Evolution of Cooperative Breeding in Birds and Mammals. Behavioural Ecology: an Evolutionary Approach* (ed. Krebs, J. R. and Davies, N. B.) 301–337 (Blackwell Scientific, Oxford, 1991).
4. Selander, R. K. *Univ. Calif. Pub. Zool.* **74,** 1–224 (1964).
5. Stacey, P. B. & Ligon, D. J. *Am. Nat.* **130,** 654–676 (1987).
6. Stacey, P. B. & Ligon, D. J. *Am. Nat.* **137,** 831–846 (1991).
7. Collar, N. J. & Stuart, S. N. *Threatened Birds of Africa and Related Islands. The ICBP/IUCN Red Data Book* (ICBP, Cambridge, UK, 1985).
8. Diamond, A. W. *ICBP, Technical Publication* **3,** 239–251 (1985).
9. Diamond, A. W. *Proc. IX Pan-African Ornithol. Congr.* 253–266 (1980).
10. Brooke, M. de L. & Houston, D. C. *J. Zool. Lond.* **200,** 779–795 (1983).
11. Komdeur, J. thesis, Univ. Cambridge, 1991).
12. Koenig, W. D. & Mumme, R. L. *Population Ecology of the Cooperatively Breeding Acorn Woodpecker* (Princeton University Press, New Jersey, 1987).
13. Komdeur, J., Bullock, I. D. & Rands, M. R. W. *Bird Conservation International* **1,** 179–188 (1991).
14. Huffstadt, A. & Prast, W. E. thesis, Univ. Amsterdam (1992).
15. Koenig, W. D., Pitelka, F. A., Carmen, W. J., Mumme, R. L. & Stanback, M. T. *Q. Rev. Biol.* (in the press).

ACKNOWLEDGEMENTS. I thank N. B. Davies for supervision and for comments on this manuscript, M. R. W. Rands (Programme Director for ICBP), who set up the Seychelles warbler research programme and arranged permission for the transfer, M. D. Komdeur for her help in the field and in processing data, and I. D. Bullock, G. Lewis, R. Mileto, G. Castle, A. Huffstadt and W. E. Prast for assistance on the Islands Aride and Cousin. The work was supported by the ICBP.

Human infection by genetically diverse SIV_{SM}-related HIV-2 in West Africa

Feng Gao*, Ling Yue*, Albert T. White†, Peter G. Pappas‡, Joseph Barchue§, Aloysius P. Hanson§, Bruce M. Greene†, Paul M. Sharp‖, George M. Shaw* & Beatrice H. Hahn*

Department of Medicine, * Divisions of Hematology/Oncology, † Geographic Medicine, and ‡ Infectious Diseases, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA
§ Liberian Institute for Biomedical Research, Robertsfield, Liberia
‖ Department of Genetics, Trinity College, Dublin 2, Ireland

OUR understanding of the biology and origins of human immunodeficiency virus type 2 (HIV-2) derives from studies of cultured isolates from urban populations experiencing epidemic infection and disease[1–8]. To test the hypothesis that such isolates might represent only a subset of a larger, genetically more diverse group of viruses, we used nested polymerase chain reactions to characterize HIV-2 sequences in uncultured mononuclear blood cells of two healthy Liberian agricultural workers, from whom virus isolation was repeatedly unsuccessful, and from a culture-positive symptomatic urban dweller. Analysis of *pol*, *env* and long terminal repeat regions revealed the presence of three highly divergent HIV-2 strains, one of which (from one of the healthy subjects) was significantly more closely related to simian immunodeficiency viruses infecting sooty mangabeys and rhesus macaques (SIV_{SM}/SIV_{MAC}) than to any virus of human derivation. This subject also harboured multiply defective viral genotypes that resulted from hypermutation of G to A bases. Our results indicate that HIV-2, SIV_{SM} and SIV_{MAC} comprise a single, highly diverse group of lentiviruses which cannot be separated into distinct phylogenetic lineages according to species of origin.

In 1989 we conducted a limited seroepidemiological survey in Liberia, West Africa, to estimate the prevalence of HIV-1 and HIV-2. Serum samples were collected from 372 healthy adults living in remote villages of northern Liberia, 944 rubber plantation workers attending outpatient clinics in rural areas of central Liberia, and 366 adults from the capital city of Monrovia and surrounding urban areas. Specimens were first screened by a recombinant HIV-1/-2 combination enzyme-linked immunosorbent assay (ELISA) and subsequently confirmed by western immunoblots specific for HIV-1 and HIV-2. Three individuals, one from Monrovia with frank AIDS and two from surrounding areas with unknown clinical status, were seropositive for HIV-1. Five other individuals, all healthy inhabitants of rural or remote villages, were seropositive for HIV-2. Two of these individuals (F0784 and 2238) could be relocated for further serologic testing, virus culture and analysis by polymerase chain reaction (PCR). Both were male workers from rubber plantations, ages 46 and 47. Their physical examinations were normal, and they had no history of sexually transmitted disease, chronic illness, blood transfusion or homosexuality. For comparison, a third HIV-2-seropositive but symptomatic man from Abidjan, Côte d'Ivoire was also studied (7312A). He was 32 years old, had lymphadenopathy, cutaneous anergy and recurrent skin abscesses, and reported frequent sexual encounters with urban prostitutes. Blood specimens from all three subjects were obtained on two separate occasions and processed for virus isolation and PCR analysis[9,10]. Mononuclear cells (PBMCs) were cultured alone, in combination with normal donor lymphocytes and macrophages, and with immortalized T-cell lines (Molt4 clone8, CEM × 174, H9, SupT1). HIV-2 was successfully isolated from subject 7312A on each of two occasions but not from subjects F0784 and 2238, despite optimal cell growth and viability.

To determine the genetic identity of viral sequences in subjects F0784, 2238 and 7312A, we used a highly sensitive nested PCR technique to amplify viral sequences directly from uncultured PBMC DNA[10,11]. Using primer pairs designed according to HIV-2/SIV_{MAC}/SIV_{SM} consensus sequences[12], 708-base-pair (bp) *pol* and 453-bp *env* fragments were amplified (Fig. 1) and sequences corresponding to a total 34,770 nucleotides were determined. This analysis revealed viral mixtures, or quasi-species[13], of varying complexity for each subject (Fig. 2). F0784 harboured the largest number of variants, most of which represented defective (prematurely terminated) viral genomes that resulted from G-to-A hypermutation[14]. Such G to A changes, which were found in two different genes (*pol* and *env*) and in blood samples obtained four months apart, accounted for 66% to 87% of all nucleotide substitutions in F0784 and resulted in sequence differences among individual *env* clones that were as high as 11.5%. *Env* and *pol* regions from 2238 and 7312A exhibited much less intrastrain variability (<0.3%) and no G-to-A hypermutation, which together with our previous studies on SIV_{AGM}[10], indicated that *Taq* polymerase or sequencing errors did not contribute significantly to the G-to-A changes observed in F0784.

To determine the phylogenetic relationships among HIV-2_{F0784}, HIV-2_{2238}, HIV-2_{7312A}, and other HIV-2/SIV_{SM}/SIV_{MAC} strains, we constructed evolutionary trees for their *pol* and *env*

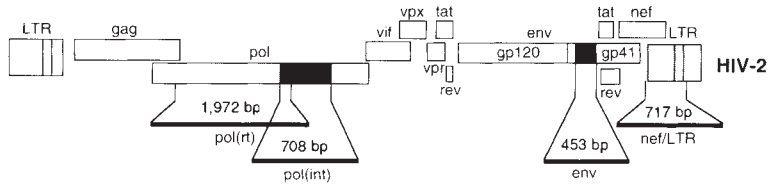


FIG. 1 Location of HIV-2 sequences amplified from uncultured PBMC DNA. Dark shaded areas highlight *pol* (integrase, int; 708 bp) and *env* (453 bp) fragments amplified from all three study subjects (F0784, 2238, 7312A). Lighter shaded areas indicate *pol* (reverse transcriptase, rt; 1,972 bp) and *nef*/LTR (717 bp) fragments amplified only from subject F0784.

METHODS. Nested PCR amplification, cloning and sequencing have been described[10]. Primer pairs were designed according to HIV-2/SIV_{MAC}/SIV_{SM} consensus sequences[12] and are numbered according to HIV-2_{ROD} (ref. 1): *env* A 5′-GCTAGGGTTCTTGGGTTTTCTCGCGACAGCAGG-3′ (nucleotides (nt) 7,691–7,723); *env* B 5′-CAAGAGGCGTATCAGCTGGCGGATCAGGAA-3′ (nt 8,415–8,444); *env* C 5′-GGGATAGTGCAGCAACAGCAACAGCTGTTG-3′ (nt 7,782–7,811); *env* D 5′-GGGAGGGGAAGAGAACACTGGCCTATA-3′ (nt 8,265–8,291); *pol* A (same as *pol* C) 5′-AATATACTAGTAGATTCACAATATGT-3′ (nt 3,857–3,882); *pol* B 5′-CTGCCTTCTCTGAAATAGACCCGAAAA-3′ (nt 4, 723–4,749); *pol* D 5′-CTTCTTTTAAAATTCATGCAATGAACTGCC-3′ (nt 4,591–4,620); reverse transcriptase (RT) A 5′-ATAGTAGGTGGAATAGGGGGTTTCATAAATA-3′ (nt 2,219–2,249); RT B 5′-TGAAGTGAAGTTGGCACCGTTGTCTGTGTG-3′ (nt 4,397–4,426); RT C 5′-ATGACAGGGGATCCCCCAATCAATATTTTTG-3′ (nt 2,309–2,339); RT D 5′-CTTCCTGTTTCCTGTGGAATGACTTCTGCT-3′ (nt 4,312–4,341); *nef* A 5′-AATATGGATGATGTAGATTCAGATGATGATGAC-3′ (nt 8,812–8,844); LTR B 5′-AAGCAGAAAGGGTCCTAACAGACCAGGGT-3′ (nt 9,739–9,767); *nef* C 5′-TGGCAATAGATATGTCACATTTTATAAA-3′ (nt 8,900–8,927); LTR D 5′-CCAGGCGGCGACTAGGAGAGATGGGAGCAC-3′ (nt 9,682–9,711). Amplification conditions were 94 °C, 1.5 min; 45 °C, 1.5 min; 55 °C, 1.5 min; 35 cycles for *env* primer pairs; 94 °C, 1 min; 40 °C, 1.5 min; 70 °C, 2 min; 30 cycles for *pol* (integrase) primer pairs; 94 °C, 1.5 min; 40° C, 1 min; 72° C, 4 min; 30 cycles for *pol* (reverse transcriptase) primer pairs, and 94 °C, 1.5 min; 40 °C, 1.5 min; 72 °C, 2 min, 30 cycles for *nef*/LTR primer pairs. The number of individual recombinant clones sequenced were as follows: F0784A *env* (11/89 PBMC), 21 clones; F0784C *env* (3/90 PBMC), 16 clones; F0784C *pol* integrase (3/90 PBMC), 5 clones; F0784C *pol* RT (3/90 PBMC), 4 clones; F0784C *nef*/LTR (3/90 PBMC), 3 clones; 2238 *env* (5/89 PBMC), 17 clones; 2238 *pol* (5/89 PBMC), 5 clones; 7312A *env* (9/90 PBMC), 3 clones; 7312A *env* (9/90 cultured), 1 clone. 7312A *pol* was sequenced directly from amplified products. A complete compilation of primary sequence data is available from GenBank/EMBL (accession numbers M87069–M87145) and the HIV database (Los Alamos National Laboratories).

sequences using both maximum parsimony[15] and neighbour-joining[16] methods (Fig. 3). In both *pol* (Fig. 3*a*) and *env* (Fig. 3*b*) regions, seven previously reported ('prototypic') isolates of HIV-2 clustered in a closely related group. In contrast, the three strains reported here branched quite differently. HIV-2_{2238} was most closely related to a single divergent HIV-2 isolate, HIV-2_{D205} (ref. 7), although these two viruses differed from each other to a greater extent than was typical for prototypic HIV-2 strains. HIV-2_{7312A} appeared to have a mosaic genome: its *pol* sequence was most closely related to HIV-2_{D205} (Fig. 3*a*), yet its *env* sequence clustered with the prototypic HIV-2 isolates (Fig. 3*b*). These discordant relationships between the HIV-2_{7312A} *pol* and *env* regions were strongly supported by bootstrap analyses[17] and probably reflect recombination between phylogenetically divergent strains.

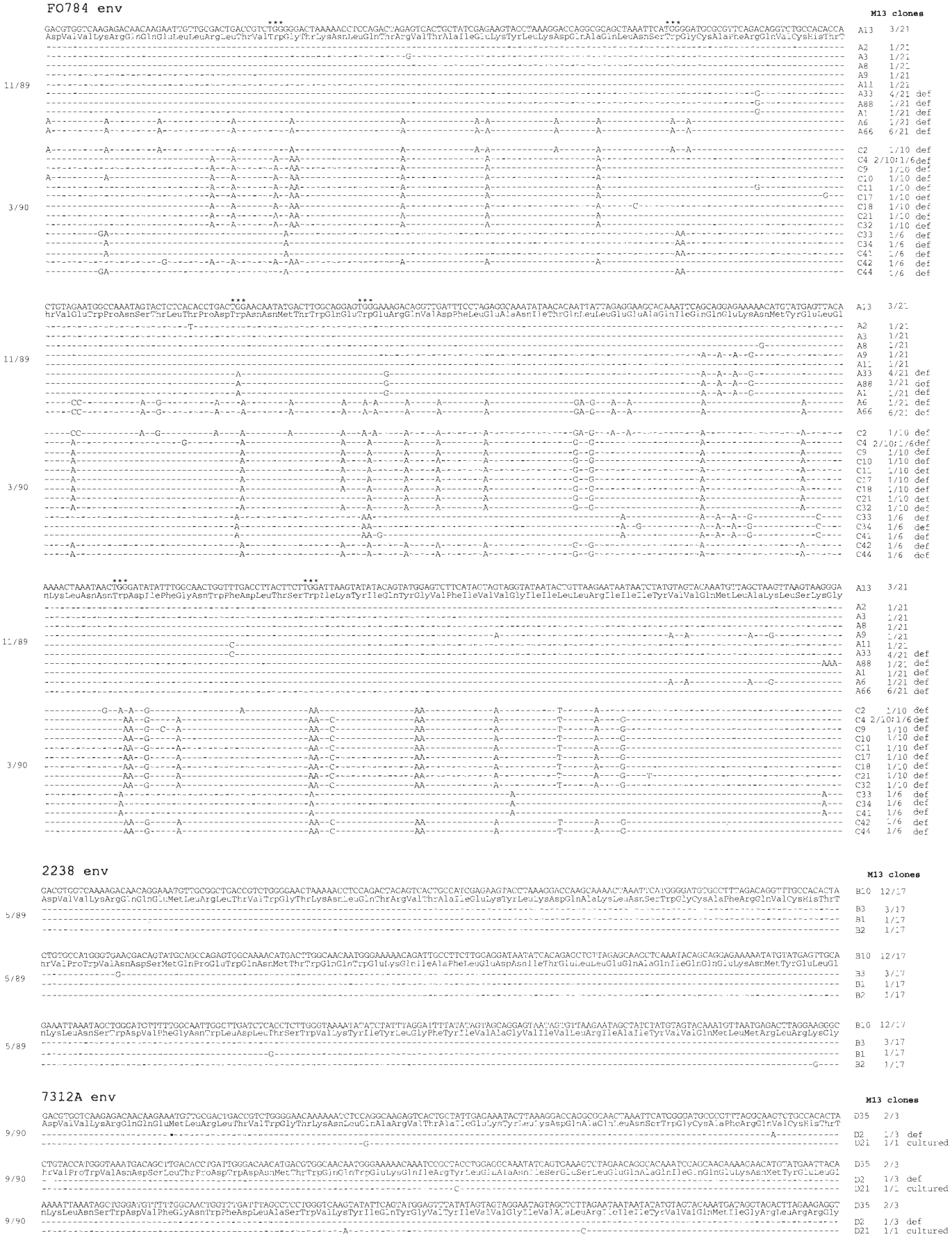
The most striking finding from the phylogenetic analyses was that HIV-2_{F0784} *pol* and *env* sequences clustered with the simian viruses, SIV_{SM} and SIV_{MAC}, rather than with other HIV-2 strains. This relationship was found in the majority, although less than 95%, of bootstrap samples, and so to substantiate the phylogenetic position of HIV-2_{F0784} we amplified and analysed two additional regions of its genome: a 1,972-bp *pol* fragment encoding reverse transcriptase and a 717-bp *nef*/long terminal repeat (LTR) fragment (Fig. 1). Phylogenetic analyses of the reverse transcriptase fragment using both maximum parsimony and neighbour-joining algorithms (Fig. 3*c*) showed clustering of HIV-2_{F0784} with the SIV lineage in more than 99% of bootstrap samples. The *nef*/LTR region was chosen for analysis because it encompasses a 40–44-bp 'signature' sequence that is present in all published clones of HIV-2 (including HIV-2_{D205}) but is absent from all known SIV_{SM}/SIV_{MAC} sequences[12]. Like the SIV genomes, HIV-2_{F0784} lacks this signature sequence (Fig. 4), which therefore can no longer distinguish between viruses that infect man and those that infect monkeys.

The results of our phylogenetic analyses, together with structural features unique to HIV-2/SIV_{SM}/SIV_{MAC} genomes[1,12,18] (that is, the presence of both *vpr* and *vpx* genes, and duplication of the LTR *trans* activation response (TAR) element), indicate that HIV-2 in man and SIV in mangabeys and captive macaques represent members of a single, albeit genetically diverse, group of viruses. Although the evolutionary origins and transmission patterns of this virus group remain to be defined, there is mounting evidence that the sooty mangabey is a natural reservoir and that human infection probably represents a zoonosis (a disease communicable from animals to man under natural conditions[19]). First, roughly 10% of wild-caught sooty mangabeys in Côte d'Ivoire and Liberia are infected with SIV_{SM} (ref. 20; and P. Fultz, personal communication). Second, the natural habitat of mangabeys coincides with the geographic pattern of HIV-2 endemicity and close contact between mangabeys and man (and between mangabeys and captive macaques) is well documented[18,20–22]. Third, SIV_{SM} generally fails to cause disease in mangabeys[20,23] but is highly pathogenic in macaques[21,22,24], as is the case for many zoonotic infections that cause less severe or even no disease in their natural hosts[19]. Definitive proof of zoonotic infection will require direct epidemiological and genetic evidence of SIV_{SM} transmission from an infected mangabey to man. Short of this, the identification of an HIV-2 strain (F0784) that is phylogenetically more closely related to simian than to human immunodeficiency viruses (Fig. 3) provides the strongest evidence yet for simian/human cross-species transmission.

It is notable that most previously studied isolates of HIV-2 were obtained by virus culture from individuals who were generally symptomatic and resided in areas where HIV-2 was spreading epidemically[1–8]. In contrast, HIV-2_{F0784} and HIV-2_{2238} could not be cultured (despite repeated attempts) from healthy inhabitants of rural areas where HIV seroprevalence was low and AIDS had not been recognized clinically. This raises the possibility that certain naturally occurring strains of HIV-2 may replicate less efficiently and cause less virulent infection. It is of interest that the majority of HIV-2 sequences from subject

FIG. 2 Intrastrain variability and G-to-A hypermutation in HIV-2. An alignment ► of 453-bp *env* sequences is shown. Individual clones are compared to a predominant, non-truncated clone with dashes indicating sequence identity and a dot indicating a single nucleotide deletion. Individual clone designations, number of clones with identical sequence versus total number of clones analysed, and the number of defective clones (def) with premature truncations are indicated on the right. Asterisks indicate premature stop codons resulting from G-to-A hypermutation.

METHODS. HIV-2 *env* fragments were amplified, cloned, and sequenced as described in Fig. 1. F0784 *env* sequences were amplified from two sequential blood specimens (11/89 and 3/90) on three different occasions resulting in 21, 10 and 6 recombinant M13 clones, as shown. Short-term PBMC cultures of subject 7312A were also analysed as indicated (cultured). Individual sequences were aligned using the program EUGENE (Baylor College of Medicine).



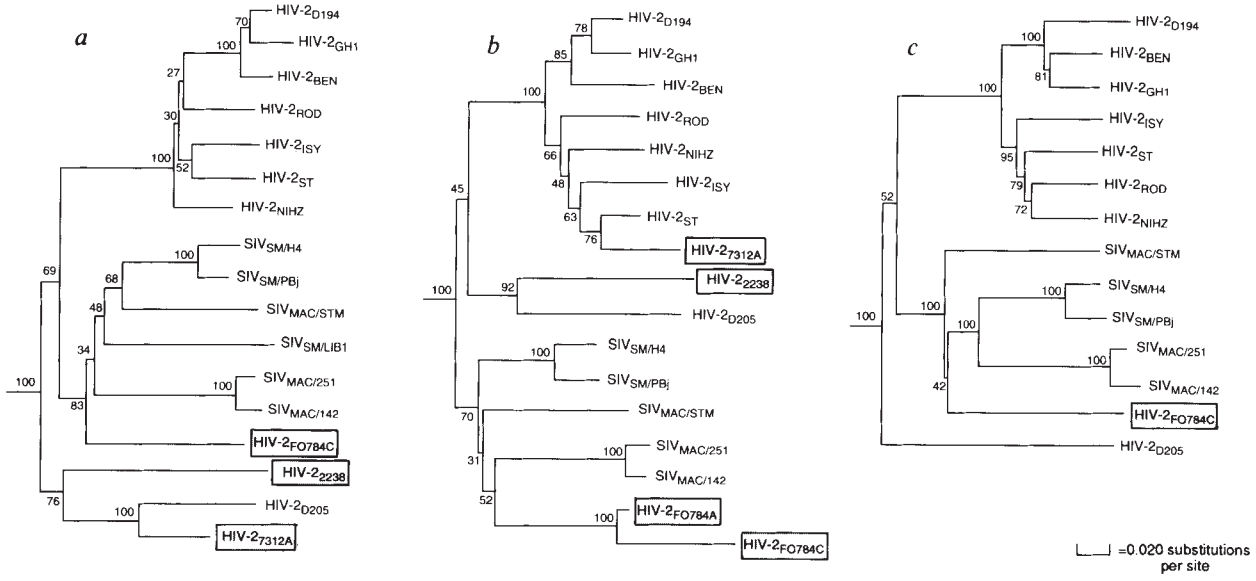


FIG. 3 Phylogenetic relationships among (*a*) *pol*-integrase, (*b*) *env*, and (*c*) *pol*-reverse transcriptase sequences from HIV-2_{F0784}, HIV-2_{2238}, and HIV-2_{7312A} (boxed) and other HIV-2, SIV_{SM}, and SIV_{MAC} isolates[12]. Horizontal branch lengths are to scale; vertical separation is for clarity only. Numbers at each node indicate the percentage of bootstrap samples (out of 5,000) in which the cluster to the right is supported. The trees are rooted using HIV-1, SIV_{AGM}, SIV_{CPZ}, and SIV_{MND} sequences (not shown); branch lengths from the nearest node to the roots of the trees shown are 0.08, 0.11 and 0.11 substitutions per site for the *pol*-integrase, *env*, and *pol*-RT trees, respectively.

METHODS. Representative *pol*-integrase (F0784pol.C12: GenBank/EMBL accession number M87110; 2238pol.B7: M87138; 7312Apol.DSU: M87145), *env* (F0784env.A13: M87069; F0784env.C9: M87090; 2238env.B10; M87118; 7312Aenv.D35; M87142), and *pol*-RT (F0784rt.C2: M87111) sequences were analysed using the neighbour-joining method[16] applied to a distance matrix of substitutions per site, corrected for multiple hits by the 2-parameter method[28], and implemented using the CLUSTALV package[29]. Phylogenetic analyses were also performed by the maximum parsimony method[15], implemented using the MULPARS option of PAUP[30] and the DNAPARS program of PHYLIP[17] (which, with repeated randomized sequence input order, gave identical results). Neighbour-joining and maximum parsimony analyses produced virtually identical topologies which differed only in non-significant aspects of the relative branching order. For example, maximum parsimony analysis of the *pol*-integrase sequences yielded four trees of length 1,422 (excluding gap mutations) compared to the neighbour-joining tree of length 1,423, and maximum parsimony analysis of reverse transcriptase sequences yielded one tree of length 3,781 compared with the neighbour-joining tree of length 3,782. In each case, the trees differed only in the relative branching order within the two major lineages (that is, prototypic HIV-2 and SIV/HIV-2_{F0784} subgroups). Analysis of *env* sequences by maximum parsimony yielded two trees of length 1,024 which differed from the neighbour-joining tree (length 1,032) by placing the HIV-$2_{D205/2238}$ cluster outside the prototype HIV-2 and SIV/HIV-2_{F0784} lineages (neither analysis could definitively resolve this aspect of the branching order). Importantly, in none of the three genomic regions examined did the topological differences affect the positions of HIV-2_{F0784}, HIV-2_{2238} or HIV-2_{7312A}. The recombinant nature of HIV-2_{7312A} and the clustering of HIV-2_{F0784} with the SIV lineage were also confirmed by testing how many additional substitutions were required to (artificially) position HIV-2_{7312A} *pol* with HIV-2_{ST} (1,499 compared with 1,422), HIV-2_{7312A} *env* with HIV-2_{D205} (1,064 compared to 1,024), and HIV-2_{F0784} RT with the prototype HIV-2 lineage or HIV-2_{D205} (3,819 and 3,841 compared with 3,781). Finally, bootstrap analyses (200 replicates) of the maximum parsimony trees (using DNABOOT from the PHYLIP package[17]) gave repeatability values similar to those from the neighbour-joining analyses. A complete description of all sequences used for the phylogenetic studies, including alignments and accession numbers, and the results of the maximum parsimony analyses are available on request from the authors.

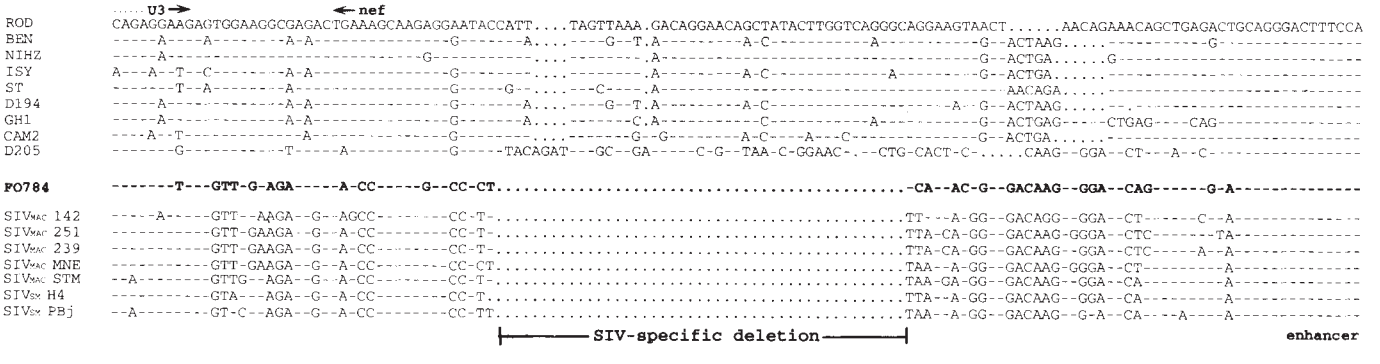


FIG. 4 Identification of an SIV_{SM}/SIV_{MAC} 'signature' sequence in HIV-2_{F0784}. HIV-2_{F0784} LTR sequences are aligned with prototype HIV-2 and SIV sequences in a region of the LTR encompassing a specific 40–44-bp insertion which distinguishes all previously reported HIV-2 viruses from SIV_{SM}/SIV_{MAC} viruses[12]. Sequences are compared to HIV-2_{ROD} (ref. 1) as a reference sequence, with dashes indicating sequence identity and dots indicating gaps introduced for optimal alignment. Positions of enhancer sequences and the *nef* termination codon are indicated. HIV-2_{F0784} (bold face), like viruses of monkey derivation, lacked the 40–44-bp insertion.

METHODS. The 717-bp *nef*/LTR fragment was amplified from uncultured F0784 PBMC DNA, cloned (F0784LTR.C5: M87115), and sequenced as described in Fig. 1.

F0784 contained multiply defective genomes and an inordinate number of G-to-A substitutions. Biased G-to-A hypermutation has previously been described for HIV-1[14,25], but not to the extent found in HIV-2_{F0784}. But in macaques experimentally infected with SIV_{SM}, extensive G-to-A hypermutation was found which correlated with reduced viral pathogenicity[26]. Studies are presently underway to elucidate the molecular basis of G-to-A hypermutation and to determine whether, in the extreme case, it could result in attenuated or even abortive HIV-2 infection.

Finally, a recent suggestion[27] that SIV_{SM} may have been accidentally transmitted to man by inoculation with infected monkey blood probably cannot explain the diversity now recognized for HIV-2. Our results re-emphasize the need to target viruses from feral monkey populations and humans living in remote areas of Africa in a search for the origins of human immunodeficiency viruses and events leading to their recent epidemic spread. □

Received 27 May; accepted 26 June 1992.

1. Guyader, M. *et al. Nature* **326,** 662–669 (1987).
2. Zagury, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **85,** 5941–5945 (1988).
3. Franchini, G. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 2433–2437 (1989).
4. Kumar, P. *et al. J. Virol.* **64,** 890–901 (1990).
5. Hasegawa, A. *et al. AIDS Res. hum. Retrovir.* **5,** 593–604 (1989).
6. Kirchhoff, F., Jentsch, K. D., Stuke, A., Mous, J. & Hunsmann, G. *AIDS* **4,** 847–857 (1990).
7. Dietrich, U. *et al. Nature* **342,** 948–950 (1989).
8. Schulz, T. F. *et al. J. Virol.* **64,** 5177–5182 (1990).
9. Clark, S. J. *et al. New Engl. J. Med.* **324,** 954–960 (1991).
10. Allan, J. S. *et al. J. Virol.* **65,** 2816–2828 (1991).
11. Mullis, K. B. & Faloona, F. A. *Meth. Enzym.* **155,** 335–350 (1987).
12. Myers, G., Berzofsky, J. A., Korber, B., Smith, T. F. & Pavlakis, G. N. *Human Retroviruses and AIDS* (Los Alamos National Laboratory, New Mexico, 1991).
13. Meyerhans, A. *et al. Cell* **58,** 901–910 (1989).
14. Vartanian, J.-P., Meyerhans, A., Asjo, B. & Wain-Hobson, S. *J. Virol.* **65,** 1779–1788 (1991).
15. Fitch, W. M. *Syst. Zool.* **20,** 406–416 (1971).
16. Saitou, N. & Nei, M. *Molec. biol. Evol.* **4,** 406–425 (1987).
17. Felsenstein, J. *Evolution* **39,** 783–791 (1985).
18. Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., Purcell, R. H. & Johnson, P. R. *Nature* **339,** 389–392 (1989).
19. T-W-Fiennes, R. N. *Zoonoses and the Origins and Ecology of Human Disease* (Academic, London, 1978).
20. Marx, P. A. *et al. J. Virol.* **65,** 4480–4485 (1991).
21. Khan, A. S. *et al. J. Virol.* **65,** 7061–7065 (1991).
22. Novembre, F. J., Hirsch, V. M., McClure, H. M., Fultz, P. N. & Johnson, P. R. *Virology* **186,** 783–787 (1992).
23. Fultz, P. N., Gordon, T. P., Anderson, D. C. & McClure, H. M. *AIDS* **4,** 619–625 (1990).
24. Letvin, N. L. *et al. Science* **230,** 71–73 (1985).
25. Li, Y. *et al. J. Virol.* **65,** 3973–3985 (1991).
26. Johnson, P. R., Hamm, T. E., Goldstein, S., Kitov, S. & Hirsch, V. M. *Virology* **185,** 217–228 (1991).
27. Gilks, C. *Nature* **354,** 262 (1991).
28. Kimura, M. *J. molec. Evol.* **16,** 111–120 (1980).
29. Higgins, D. G., Bleasby, A. & Fuchs, R. *Comp. appl. Biosci.* **8,** 189–191 (1992).
30. Swofford, D. L. *PAUP: Phylogenetic Analysis Using Parsimony* (Illinois Natural History Survey, Champaign, Illinois, 1991).

ACKNOWLEDGEMENTS. This paper is dedicated to the memory of B.M.G. We thank G. Myers and K. MacInnes for assistance with phylogenetic analyses; the Irish National Centre for Bioinformatics for their facilities; J. Hoxie for independent attempts at cultivating blood samples from subject 2238; Serologicals, Inc. (Atlanta, GA) for blood specimens from subject 7312A; R. Desrosiers, P. Fultz and D. Ho for discussion; D. Decker and M. Mixon for technical assistance; and C. Davis and A. J. Nicholson for manuscript preparation. This work was supported by grants from the NIH, the US Army Medical Research Acquisition Activity, the Life and Health Insurance Medical Research Fund, and the Birmingham Center for AIDS Research. G.M.S. is a PEW Scholar in the Biomedical Sciences.

A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin

Meera Berry, Frank Grosveld & Niall Dillon

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

IN normal humans the fetal stage-specific γ-globin genes are silenced after birth and not expressed in the adult. Exceptions are seen in cases of hereditary persistence of fetal haemoglobin (HPFH). These are clinically important because the elevated levels of γ-globin can alleviate β-thalassaemia and sickle cell anaemia. One class of mutations is associated with point mutations in the promoter of the γ-globin genes (non-deletion HPFH), whereas others seem to be caused by large deletions 3′ to the γ-globin genes[1]. To test whether the point mutation found in the Greek non-deletion HPFH[2,3] (guanine to adenine at nucleotide position −117) is the cause of the raised γ-globin levels in the adult stage and is not just a linked polymorphism, we engineered this mutation into a γ-globin gene. When this gene was introduced into mice, the presence of the −117 mutation results in persistence of γ-globin expression at a high level and a concomitant decrease in β-globin expression in fetal and adult mice. We show that these changes correlate with the loss of binding of the transcription factor GATA1 to the γ-globin promoter, suggesting that it may act as a negative regulator of the γ-globin gene in adults.

Two globin minilocus constructs were injected into fertilized mouse eggs. The first construct contained a wild-type (wt) γ-globin gene flanked by the entire locus control region (LCR) and a β-globin gene[4] (Fig. 1). The second construct was the same wild-type locus but with a single engineered point mutation at position −117 (G→A) in the promoter of the γ-globin gene. This mutation was verified by sequence analysis (not shown). The β-gene was included as a reference gene for quantitation and to allow rapid analysis of the construct without the need to establish a large number of bred lines. When the wild-type γβ minilocus was introduced into fertilized mouse eggs, five transgenic mice were obtained. Southern blots showed that two of the founders were mosaic (31 and 36) and that all contained the intact γβ minilocus, albeit at different copy numbers (Table 1, and data not shown). S1 nuclease protection analysis showed that the γ-globin gene expression was suppressed in adult mice (Fig. 1*a*, *b*). In contrast, the human β-globin gene was expressed at this stage at levels comparable to those observed for the mouse β-maj-globin genes[5] (Fig. 1*b*; Table 1). The suppression of the wild-type γ-globin gene is in agreement with results obtained when a minilocus containing only the γ-globin gene is introduced into mice[4]. Repeated phlebotomy increases the number of reticulocytes, but even under those conditions the γ-globin gene remains suppressed (Fig. 1*b*). When the −117 mutant γβ minilocus was introduced into mice, nine transgenic mice were obtained and Southern blots showed that they contained intact miniloci at different copy numbers, although a number were mosaic (Table 1, and data not shown). S1 protection analysis showed a completely different result from that obtained with the wild-type γβ minilocus. The −117 mutant γ-globin gene is now expressed at high levels in the adult stage in eight out of nine founder transgenic mice (Fig. 1*a*, *c*). Line 7, which does not express γ at the adult stage, was found to express the γ gene at the embryonic and early fetal stages (not shown). Analysis by polymerase chain reaction (PCR)[6] of the γ-globin gene promoter in line 7 showed that it still contained the −117 mutation, thus ruling out a reversion of the mutation (not shown). The line 7 γ gene was not analysed for other mutations and we have therefore, as yet, no explanation for this exception. The expression level of the γ gene in the other HPFH mice varied considerably (Fig. 1; Table 1), which may be caused by the different arrangement of the transgene loci (P. Fraser and N.D., unpublished results; for example, the three lowest γ/β expressors all have a head-to-head integration) and the fact that the developmental regulation of the γ gene is very sensitive to position effects[4,7,8].

To determine whether expression of the γ-globin gene at the adult stage in the HPFH mice leads to a partial suppression of the linked human β-globin gene, all RNA samples of the wild type and HPFH γβ minilocus-containing mice were compared