Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 4108–4111, June 1990
Evolution

Evolutionary origin of human and simian immunodeficiency viruses

(molecular evolution/phylogeny/recombination/primate lentiviruses)

TAKASHI GOJOBORI*, ETSUKO N. MORIYAMA*, YASUO INA*, KAZUHO IKEO*, TOMOYUKI MIURA†, HAJIME TSUJIMOTO‡, MASANORI HAYAMI†, AND SHOZO YOKOYAMA§

*National Institute of Genetics, Mishima 411, Japan; †Institute for Virus Research, Kyoto University, Kyoto 606, Japan; ‡Department of Animal Pathology, Institute of Medical Sciences, University of Tokyo, Tokyo 108, Japan; and §Department of Ecology, Ethology, and Evolution, University of Illinois at Urbana-Champaign, Champaign, IL 61801

Communicated by Motoo Kimura, March 12, 1990 (received for review February 15, 1990)

ABSTRACT **From what viruses the human immunodeficiency viruses (HIVs) originated is an extremely controversial question. To address this question, we have analyzed nucleotide sequences of simian immunodeficiency viruses (SIVs) and HIVs by using the techniques for understanding molecular evolution. In particular, we compared the nucleotide sequences of whole genomes, gene region by gene region, between a given pair of viruses, including four types of SIVs—isolated from mandrills (*Papio sphinx*), African green monkeys (*Cercopithecus aethiops*), sooty mangabeys (*Cercocebus atys*), and rhesus macaques (*Macaca mulatta*)—as well as HIVs. Phylogenetic trees for all gene regions examined showed that the present HIVs may have emerged as different variants of SIVs of Old World monkeys, possibly from recombination between viruses related to SIVs.**

AIDS, which was discovered only about 10 yr ago, has now spread over almost the entire world. Human immunodeficiency viruses type 1 and 2 (HIV-1 and -2) are the etiological agents of AIDS in humans. Although simian immunodeficiency viruses (SIVs) have been found to infect both Asian and African Old World monkeys, there have been no reports of natural infection of SIV isolated from rhesus macaques (*Macaca mulatta*) (SIV_{mac}) or from sooty mangabeys (*Cercocebus atys*) (SIV_{sm}) (1). Studies of these viruses at the molecular level have been extensive, and the nucleotide sequence data of viral genes are rapidly accumulating. Using these sequence data, several authors have studied evolutionary relationships of these viruses (2–6). However, their results could not give any definite answer to the question of the origin of HIVs.

To elucidate this subject, we have examined the phylogenetic relationships between primate lentiviruses in detail. We compared the nucleotide sequences of each gene region [3′ long terminal repeat (LTR), *gag*, *pol*, *env*, and *nef*] between a given pair of viruses. We then estimated the total number of nucleotide substitutions for each pair of isolates compared by the 6-parameter method (7). Using this number, we constructed a phylogenetic tree by the neighbor-joining method (8), which does not require the assumption of a constant substitution rate. We also constructed phylogenetic trees by the method of unweighted pair grouping (9), using the numbers of synonymous and nonsynonymous substitutions as well as the numbers of nucleotide substitutions at each position of a codon separately (data not shown).

Fig. 1*A* shows a phylogenetic tree for the 3′ LTR, constructed by the neighbor-joining method (8) by use of the number of nucleotide substitutions. The tree shows that the HIV-1 group, the HIV-2 group, SIV isolated from African green monkeys (*Cercopithecus aethiops*) (SIV_{agm}) and SIV isolated from mandrills (*Papio sphinx*) (SIV_{mnd}) all diverged from one ancestor at almost the same time, although SIV_{sm} and SIV_{mac} are definitely in the HIV-2 group. These conclusions are supported by the phylogenetic trees for all other genes we examined (Fig. 1 *B–E*). If SIV_{mnd} and SIV_{agm} emerged as variants of ancestral SIVs at some time, the phylogenetic trees shown here imply that HIV-1 and -2 emerged in exactly the same way as these SIVs. In the native habitat, 30–50% of African green monkeys have antibodies reactive to SIV_{agm} (1, 34). Moreover, genetic variation among SIV_{agm} isolates is much greater than that seen within each group of HIV-1, HIV-2, or SIV_{mac} isolates (35). These findings suggest that SIV has been present in the African green monkey population for a long time. Thus, HIV-1 and -2 might have appeared as different variants of SIVs that existed at the time.

Because the divergence of *Homo sapiens* and Old World monkeys, such as African green monkeys and mandrills, occurred much earlier than that between African green monkeys and mandrills, these results suggest that the evolution of HIV and SIV did not follow that of their host species. This fact further suggests that interspecies transmission of SIV or HIV probably occurred in the past, although this conclusion can be made only with the assumption that the rate of nucleotide substitution for primate lentiviruses did not change drastically at some time during evolution (2, 36).

Excluding SIV_{mnd} and SIV_{agm}, HIVs and SIVs can be classified into two major groups, the HIV-1 and -2 groups. The HIV-1 group contains all HIV isolates from New York, San Francisco, Haiti, and central African countries, such as Zaire. The BRU isolate, from a French patient, apparently originated from a homosexual man in New York (10). The HIV-2 group contains all viral isolates from West Africa and also SIV_{mac} and SIV_{sm}.

The evolutionary positions on phylogenetic trees of SIV_{mnd} and SIV_{agm} depend on the gene examined. The phylogenetic tree for the *gag* region is basically the same as that for LTR, except for one important point. SIV_{mnd} is located at the outside of all other clusters (Fig. 1*B*), but SIV_{agm} here is related to the HIV-2 group. In the tree for the *pol* region, SIV_{mnd} and SIV_{agm} are closely related to the HIV-1 group (Fig. 1*C*), although the branching point between these SIVs and other members of the HIV-1 group is very near the branching point between the HIV-1 and -2 groups. For *env*, the tree topology suggests that SIV_{mnd} and SIV_{agm} are related to the HIV-2 group (Fig. 1*D*). In the tree for the *nef* region, SIV_{agm} is related to the HIV-1 group, although SIV_{mnd} is located at the outside of all other clusters (Fig. 1*E*). Thus, whether SIV_{agm} and SIV_{mnd} are in the HIV-1 or -2 group depends upon the gene-region examined. This observation is

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-1 and -2, human immunodeficiency virus(es) type 1 and 2, respectively; SIV, simian immunodeficiency virus; SIV_{mnd}, SIV_{sm}, SIV_{agm}, and SIV_{mac}, SIVs isolated from mandrill, sooty mangabey, African green monkey, and rhesus macaques, respectively; LTR, long terminal repeat.

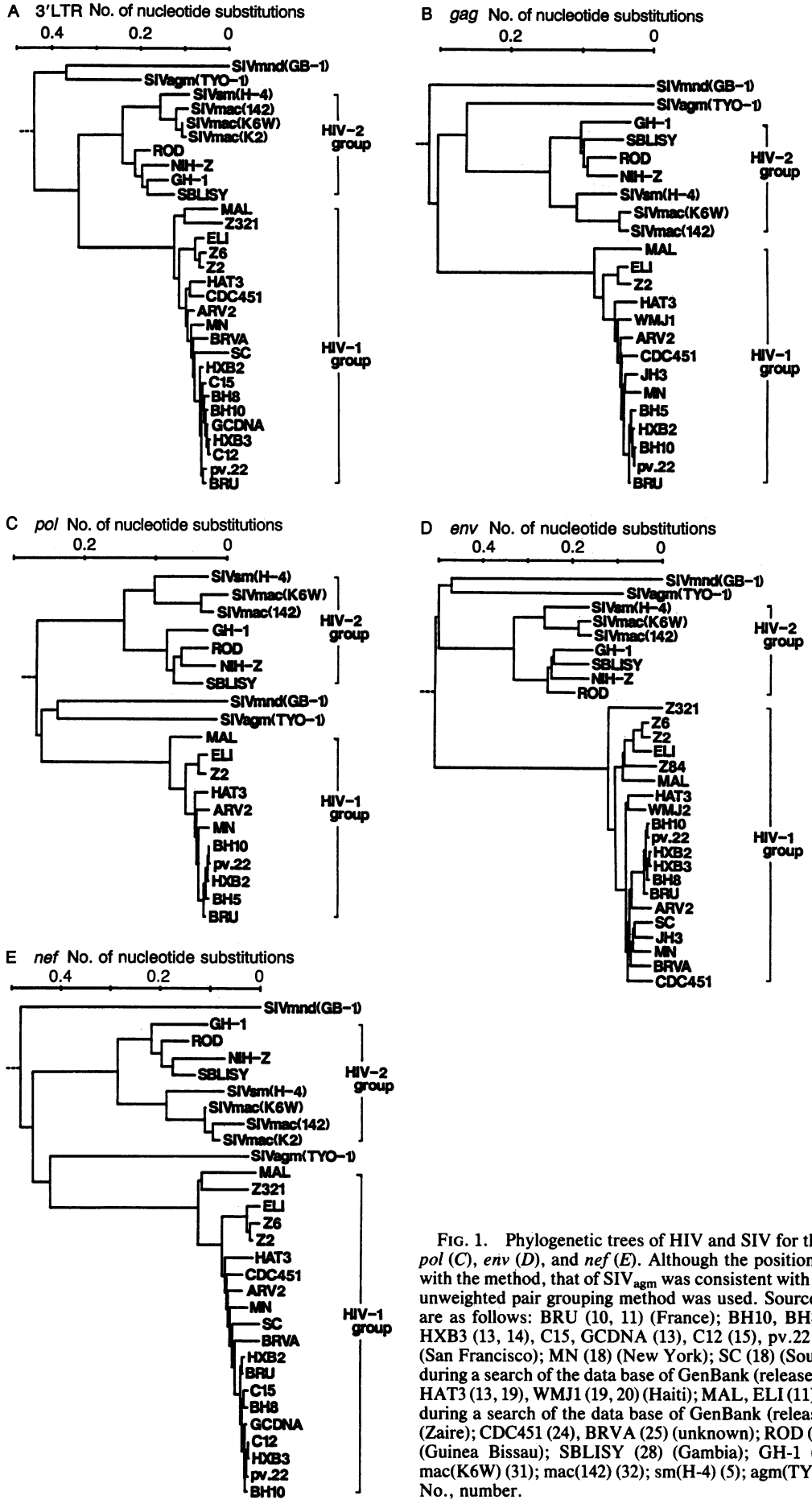

FIG. 1. Phylogenetic trees of HIV and SIV for the genes 3' LTR (*A*), *gag* (*B*), *pol* (*C*), *env* (*D*), and *nef* (*E*). Although the position of SIV_{mnd} in the tree varied with the method, that of SIV_{agm} was consistent with these results shown when the unweighted pair grouping method was used. Sources and geographical locations are as follows: BRU (10, 11) (France); BH10, BH8, BH5 (12), HXB2 (12, 13), HXB3 (13, 14), C15, GCDNA (13), C12 (15), pv.22 (16) (New York); ARV2 (17) (San Francisco); MN (18) (New York); SC (18) (Southern California); JH3 (noted during a search of the data base of GenBank (release 60.0), June, 1989 (unknown); HAT3 (13, 19), WMJ1 (19, 20) (Haiti); MAL, ELI (11), Z321 (21), Z6 (22), Z2 (noted during a search of the data base of GenBank (release 60.0), June, 1989, Z84 (23) (Zaire); CDC451 (24), BRVA (25) (unknown); ROD (26) (Cape Verde); NIH-Z (27) (Guinea Bissau); SBLISY (28) (Gambia); GH-1 (29) (Ghana); mac(K2) (30); mac(K6W) (31); mac(142) (32); sm(H-4) (5); agm(TYO-1) (33); and mnd(GB-1) (4). No., number.

supported by the phylogenetic trees that were constructed by the method of unweighted pair grouping (data not shown).

It is, therefore, possible that HIV-1 and -2 emerged from recombinational events between ancestral simian viruses. If this is the case, the recombination must have occurred several centuries ago, because the divergence time between SIV_{agm} and HIV is almost the same as the divergence time between the HIV-1 and -2 groups; this divergence time has been estimated to be >150 yr (2, 37). Because such results might arise from statistical fluctuations, however, we cannot exclude the possibility that SIV_{agm} and SIV_{mnd}, respectively, represent the third and fourth groups of primate lentiviruses, in addition to the HIV-1 and -2 groups.

To clarify relationships between HIV and other lentiviruses and to identify the root of the phylogenetic trees for HIV and SIV, we constructed a phylogenetic tree for one spumavirus, oncoviruses, and lentiviruses, including HIV and SIV. Comparing the amino acid sequences for reverse transcriptase and endonuclease with each other, we estimated the numbers of amino acid substitutions. For correction of multiple substitutions at a single amino acid site, we used Kimura's (38) equation $d_a = -\log_e(1 - p - 0.5p^2)$, where p is the proportion of different amino acids between pairs of viruses compared and d_a is the number of amino acid substitutions per site. With use of the values of d_a, a phylogenetic tree of retroviruses was constructed by the neighbor-joining method (8).

From the phylogenetic tree (Fig. 2), we found that the root of the phylogenetic tree for the primate lentiviruses is between SIV_{mnd} and the other primate lentiviruses; the root is very close to the branching points between the HIV-1 and -2 groups and between SIV_{agm} and the other members of the HIV-1 group. This suggests that viruses related to SIV_{mnd} or

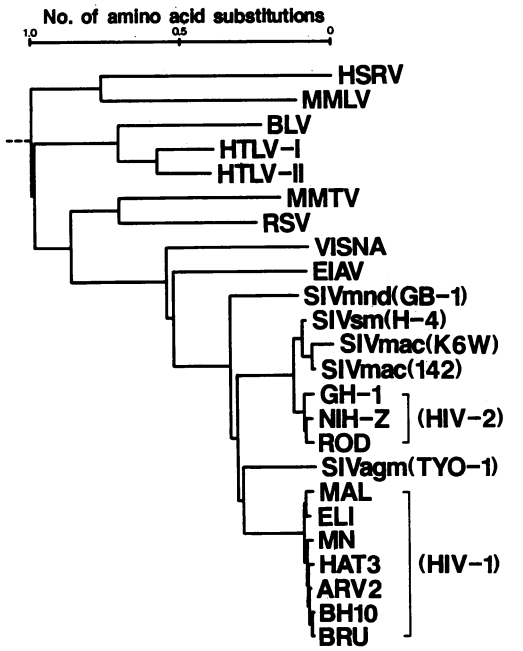

FIG. 2. Phylogenetic tree for the retrovirus family. This tree was constructed by use of the amino acid sequences of reverse transcriptase and endonuclease encoded by *pol*; this region is highly conserved among retroviruses, and only this region is comparable in different subfamilies. Sources: lentiviruses [equine infectious anemia virus (EIAV) (39) and ovine Visna lentivirus (Visna) (40)], a spumavirus [human spumaretrovirus (HSRV) (41, 42)], and oncoviruses [Moloney murine leukemia virus (MMLV) (43), bovine leukemia virus (BLV) (44), human T-lymphotropic virus types 1 and 2, respectively, (HTLV-I) (45) and HTLV-II (46), mouse mammary tumor virus (MMTV) (39), and Rous sarcoma virus (RSV) (47)]. No., number.

SIV_{agm} were an ancestral virus of HIV and SIV. We also confirmed that HIV and SIV are phylogenetically similar to equine infectious anemia virus and ovine Visna lentivirus, and that all of these viruses form a group as a lentivirus subfamily (48). Retroviruses are roughly separated into three groups, but some of the oncoviruses clustered together with the lentiviruses and spumavirus on a phylogenetic tree. This grouping suggests that recombination has taken place among the different subfamilies of retroviruses (49, 50).

These results are evidence that HIV-1 and -2 may have emerged as different variants of simian viruses. Both types of HIV may have been derived from early recombinations between ancestral SIVs. To confirm our conclusion, sequences of older SIVs or those isolated from primates in different regions of Africa may be required.

This study was supported by a Grant-in-Aid from the Ministry of Education, Science, and Culture, Japan to T.G., E.N.M. and M.H. and National Institutes of Health Grants GM-28672 and GM-4237 and National Science Foundation Grant DMB89000ZN to S.Y.

1. Ohta, Y., Masuda, T., Tsujimoto, H., Ishikawa, K., Kodama, T., Morikawa, S., Nakai, M., Honjo, S. & Hayami, M. (1988) *Int. J. Cancer* **41**, 115–122.
2. Sharp, P. M. & Li, W.-H. (1988) *Nature (London)* **336**, 315.
3. Smith, T. F., Srinivasan, A., Schochetman, G., Marcus, M. & Myers, G. (1988) *Nature (London)* **333**, 573–575.
4. Tsujimoto, H., Hasegawa, A., Maki, N., Fukasawa, M., Miura, T., Speidel, S., Cooper, R. W., Moriyama, E. N., Gojobori, T. & Hayami, M. (1989) *Nature (London)* **341**, 539–541.
5. Hirsch, V. M., Olmsted, R. A., Murphey-Corb, M., Purcell, R. H. & Johnson, P. R. (1989) *Nature (London)* **339**, 389–392.
6. Dietrich, U., Adamski, M., Kreutz, R., Seipp, A., Kühnel, H. & Rübsamen-Waigmann, H. (1989) *Nature (London)* **342**, 948–950.
7. Gojobori, T., Ishii, K. & Nei, M. (1982) *J. Mol. Evol.* **18**, 414–423.
8. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
9. Nei, M. (1975) *Molecular Population Genetics and Evolution* (North-Holland, Amsterdam).
10. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) *Cell* **40**, 9–17.
11. Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. (1986) *Cell* **46**, 63–74.
12. Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313**, 277–284.
13. Ratner, L., Starcich, B., Josephs, S. F., Hahn, B. H., Reddy, E. P., Livak, K. J., Petteway, S. R., Jr., Pearson, M. L., Haseltine, W. A., Arya, S. K. & Wong-Staal, F. (1985) *Nucleic Acids Res.* **13**, 8219–8229.
14. Crowl, R., Ganguly, K., Gordon, M., Conroy, R., Schaber, M., Kramer, R., Shaw, G., Wong-Staal, F. & Reddy, E. P. (1985) *Cell* **41**, 979–986.
15. Arya, S. K. & Gallo, R. C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2209–2213.
16. Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A. & Capon, D. J. (1985) *Nature (London)* **313**, 450–458.
17. Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D. & Luciw, P. A. (1985) *Science* **227**, 484–492.
18. Gurgo, C., Guo, H.-G., Franchini, G., Aldovini, A., Collalti, E., Farrell, K., Wong-Staal, F., Gallo, R. C. & Reitz, M. S., Jr. (1988) *Virology* **164**, 531–536.
19. 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.
20. Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo,

R. C., Parks, E. S. & Parks, W. P. (1986) *Science* **232,** 1548–1553.
21. Myers, G., Josephs, S. S., Rabson, A. B., Smith, T. F. & Wong-Staal, F., eds. (1988) *Human Retroviruses and AIDS* (AIDS Program of the National Institute of Allergy and Infectious Diseases and the U.S. Department of Energy, Washington, DC).
22. Srinivasan, A., Anand, R., York, D., Ranganathan, P., Feorino, P., Schochetman, G., Curran, J., Kalyanaraman, V. S., Luciw, P. A. & Sanchez-Pescador, R. (1987) *Gene* **52,** 71–82.
23. Yourno, J., Josephs, S. F., Reitz, M., Zagury, D., Wong-Staal, F. & Gallo, R. C. (1988) *AIDS Res. Hum. Retroviruses* **4,** 165–173.
24. Desai, S. M., Kalyanaraman, V. S., Casey, J. M., Srinivasan, A., Andersen, P. R. & Devare, S. G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8380–8384.
25. Anand, R., Thayer, R., Srinivasan, A., Nayyar, S., Gardner, M., Luciw, P. & Dandekar, S. (1989) *Virology* **168,** 79–89.
26. Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. & Alizon, M. (1987) *Nature (London)* **326,** 662–669.
27. Zagury, J. F., Franchini, G., Reitz, M., Collalti, E., Starcich, B., Hall, L., Fargnoli, K., Jagodzinski, L., Guo, H.-G., Laure, F., Arya, S. K., Josephs, S., Zagury, D., Wong-Staal, F. & Gallo, R. C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5941–5945.
28. Franchini, G., Fargnoli, K. A., Giombini, F., Jagodzinski, L., De Rossi, A., Bosch, M., Biberfeld, G., Fenyo, E. M., Albert, J., Gallo, R. C. & Wong-Staal, F. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2433–2437.
29. Hasegawa, A., Tsujimoto, H., Maki, N., Ishikawa, K., Miura, T., Fukasawa, M., Miki, K. & Hayami, M. (1989) *AIDS Res. Hum. Retroviruses* **5,** 593–604.
30. Hirsh, V., Riedel, N. & Mullins, J. I. (1987) *Cell* **49,** 307–319.
31. Franchini, G., Gurgo, C., Guo, H.-G., Gallo, R. C., Collalti, E., Fargnoli, K. A., Hall, L. F., Wong-Staal, F. & Reitz, M. S., Jr. (1987) *Nature (London)* **328,** 539–543.
32. Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P. & Sonigo, P. (1987) *Nature (London)* **328,** 543–547.
33. Fukasawa, M., Miura, T., Hasegawa, A., Morikawa, S., Tsujimoto, H., Miki, K., Kitamura, T. & Hayami, M. (1988) *Nature (London)* **333,** 457–461.
34. Kanki, P. J., Kurth, R., Becker, W., Dreesman, G., McLane, M. F. & Essex, M. (1985) *Lancet* **i,** 1330–1332.
35. Li, Y., Naidu, Y. M., Daniel, M. D. & Desrosiers, R. C. (1989) *J. Virol.* **63,** 1800–1802.
36. Yokoyama, S. & Gojobori, T. (1987) *J. Mol. Evol.* **24,** 330–336.
37. Gojobori, T., Moriyama, E. N. & Yokoyama, S. (1988) *4th Int. Conf. AIDS*, Book 1, 142.
38. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
39. Chiu, I.-M., Yaniv, A., Dahlberg, J. E., Gazit, A., Skuntz, S. F., Tronick, S. R. & Aaronson, S. A. (1985) *Nature (London)* **317,** 366–368.
40. Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A. & Wain-Hobson, S. (1985) *Cell* **42,** 369–382.
41. Flügel, R. M., Rethwilm, A., Maurer, B. & Darai, G. (1987) *EMBO J.* **6,** 2077–2084.
42. Maurer, B., Bannert, H., Darai, G. & Flügel, R. M. (1988) *J. Virol.* **62,** 1590–1597.
43. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) *Nature (London)* **293,** 543–548.
44. Sagata, N., Yasunaga, T., Tsuzuki-Kawamura, J., Ohishi, K., Ogawa, Y. & Ikawa, Y. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 677–681.
45. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3618–3622.
46. Shimotohno, K., Gold, D. W., Miwa, M., Sugiyama, T. & Chen, I. S. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1079–1083.
47. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32,** 853–869.
48. Yokoyama, S., Moriyama, E. N. & Gojobori, T. (1987) *Proc. Jpn. Acad.* **63,** 147–150.
49. Doolittle, R. F. (1989) *Nature (London)* **339,** 338–339.
50. McClure, M. A., Johnson, M. S., Feng, D.-F. & Doolittle, R. F. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2469–2473.