Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5062–5065, June 1993
Medical Sciences

Origin of JC polyomavirus variants associated with progressive multifocal leukoencephalopathy

TAKAKO IIDA*†, TADAICHI KITAMURA‡, JING GUO§, FUMIAKI TAGUCHI*, YOSHIO ASO¶, KAZUO NAGASHIMA‖, AND YOSHIAKI YOGO§**

*Department of Microbiology, School of Hygienic Sciences, Kitasato University, Sagamihara, Kanagawa 228, Japan; ‡Department of Urology, Branch Hospital, Faculty of Medicine, The University of Tokyo, Mejirodai, Bunkyo-ku, Tokyo 112, Japan; §Department of Viral Infection, The Institute of Medical Science, The University of Tokyo, Shirokanedai, Minato-ku, Tokyo 108, Japan; ¶Department of Urology, Faculty of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan; and ‖Department of Pathology, Hokkaido University School of Medicine, Kita-ku, Sapporo 060, Japan

Communicated by Susumo Ohno, March 8, 1993

ABSTRACT **JC polyomavirus (JCV) DNAs from the urine of nonimmunocompromised individuals (designated archetypal isolates) regularly contain a regulatory sequence that may have generated various regulatory sequences of JCV isolates derived from the brain of patients with progressive multifocal leukoencephalopathy (PML). In this report, we constructed a phylogenetic tree for 14 isolates (7 archetypes and 7 PML types) from DNA sequence data on the VP1 (major capsid protein) gene. According to the phylogenetic tree, the 14 isolates diverged into types A and B, each of which contained archetypal and PML-type isolates. Each type further diverged into several groups containing archetypal and PML-type isolates. We conclude that PML-type isolates are polyphyletic in their origin and do not constitute a unique lineage. This conclusion suggests that PML-type JCV isolates are generated from archetypal strains during persistence in the hosts. Furthermore, the present phylogenetic analysis indicates that an ancestral JCV carried the archetypal regulatory sequence and that this structure has been conserved in the course of JCV evolution.**

JC polyomavirus (JCV) is a ubiquitous symbiote in the human population (1), infecting children asymptomatically and then persisting in the renal tissue (2, 3). Yet, this virus causes a fatal demyelinating disease of the central nervous system, known as progressive multifocal leukoencephalopathy (PML), in patients with decreased immune competence (4). Among JCV isolates from the brain of PML patients (designated herein as PML-type isolates), there occurs a remarkable diversity in the sequences of the regulatory regions in the viral genome, although the other regions are apparently conserved (5–13). It has been suggested that this variation may have been the consequence of JCV's adaptation to growth in brain tissue (7, 8, 12, 14). Recently, we had an opportunity to clone JCV DNAs from the urine of nonimmunocompromised patients as well as healthy volunteers and analyzed them for the structures of their regulatory regions (15). We found that the regulatory sequences of all cloned DNAs were identical, albeit a few nucleotide substitutions. Yet, this regulatory sequence was remarkably different from those of PML-type isolates. We reasoned that various regulatory sequences of PML-type isolates must have been derived from the archetypal sequences noted above (15).

In this study, we attempted to elucidate the origin of PML-type isolates by molecular phylogenetic analysis. DNA sequences of the VP1 (major capsid protein) genes are compared among 14 isolates.†† Seven of them derived from urine of asymptomatic individuals have either the archetypal regulatory region or a regulatory region that deviated a little from the archetype, whereas the other seven were PML-type isolates carrying rearranged regulatory sequences. It was hoped that comparison of genetic distances among these isolates will provide the answer to the crucial question of whether or not various PML-type isolates constituted a unique pathogenic strain that became independent of the archetype. Alternatively, each PML-type might have been derived directly from the archetypal strain in an independent manner on sporadic occasions.

MATERIALS AND METHODS

JCV Isolates. Origins of 14 JCV isolates used in this study are shown in Table 1. Sequences of their regulatory regions are diagrammatically shown in Fig. 1. A urine isolate (C1) had a regulatory region that deviated from the archetype by a 5-nucleotide (nt) deletion (16). We designate this isolate and those carrying the archetypal regulatory region (G2, G3, N1, N4, CY, and MY) as archetypal isolates for convenience. PML-type isolates M1, H1, M8, NY, T1, and S1 shared two common features, duplication of domain A and deletion of domain B, whereas M11 had a unique regulatory region very different from the other PML-type isolates.

Sequence Determination. An *Acc* I fragment comprising a region from nt 412 to nt 2835 [the nucleotide numbering system was that of Frisque *et al.* (8)] was blunt-ended and inserted into pUC118 at the *Sma* I site. Nested sets of deletion mutants were generated from recombinant plasmids that contained the *Acc* I fragment in both orientations. The deletion mutants were prepared by the "deletion kit for kilo-sequencing" (Takara Shuzo, Kyoto). *Escherichia coli* MV1184 containing deletion mutants was infected with helper phage (M13 K07) and extracellular recombinant phages were used to prepare single-stranded DNAs, which were purified and sequenced by chain termination (19). The sequences of the VP1 genes of most isolates except M1 and NY were determined in this study. The published sequence data for M1 (8) was used. The sequence of NY will be described elsewhere (17). Although the sequences of the VP1 genes and also those of their flanking regions were determined, only those of the VP1 genes were used to compare isolates.

Phylogenetic Tree. VP1 sequences were compared between all pairs of isolates and nucleotide difference scores were transformed into a matrix. This was then used to construct an

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: JCV, JC polyomavirus; PML, progressive multifocal leukoencephalopathy; PML-type isolates, JCV isolated from the brain of PML patients.

†Present address: Kanagawa Prefectural Public Health Laboratories, Yokohama, Kanagawa 241, Japan.
**To whom reprint requests should be addressed.

††The sequences reported in this paper have been deposited in the GenBank data base [accession nos. D11356 (G2), D11357 (G3), D11358 (N1), D11359 (N4), D11360 (C1), D11361 (CY), D11362 (MY), D11363 (Mad8-Br), D11364 (Mad11-Br), D11365 (Her1-Br), D11366 (NY-1B), D11367 (Sap-1), and D11368 (Tokyo-1)].

Table 1. Origin of JCV isolates studied

Isolate	Abbreviation	Origin of clone: Country	Origin of clone: Source	Cloning methods*	Ref.
G2	G2	Germany	Urine	DC	16
G3	G3	Germany	Urine	DC	16
N1	N1	Netherlands	Urine	DC	16
N4	N4	Netherlands	Urine	DC	16
CY	CY	Japan	Urine	DC	15
MY	MY	Japan	Urine	DC	15
C1	C1	China (Taiwan)	Urine	DC	16
Mad1-TC	M1	United States	PML brain	TC	8
Her1-Br	H1	United States	PML brain	DC	10
Mad8-Br	M8	United States	PML brain	DC	9
Mad11-Br	M11	United States	PML brain	DC	10
NY-1B	NY	United States	PML brain	DC	17
JCT-Br	T1	Japan	PML brain	DC	11
Sap-1	S1	Japan	PML brain†	DC	13

*DC, direct molecular cloning; TC, viral culturing followed by molecular cloning.
†Cerebellum (other PML brain samples were from the cerebrum).

evolutionary tree by the unweighted pair-group method with arithmetic averages (20). An evolutionary tree was also constructed using the neighbor-joining method (21) (data not shown).

RESULTS

Sequence Variation. To analyze the phylogenetic relationship among various isolates, we compared the nucleotide sequences of a 1065-nt region encoding VP1 (major capsid protein). In Fig. 2, only those nucleotides at the positions of variations are shown. In total, 58 nucleotide changes were detected at 57 positions (at position 1805, 2 different nucleotide changes were detected). The nucleotides at 15 positions (boxed in Fig. 2) differed between types A and B, which we previously identified on the basis of restriction analysis (16). At the other positions, one or more isolates had unique nucleotides that differed from those of the other isolates of both types or from those of type A or B. Of the 58 nucleotide differences, 13 (20%) caused amino acid substitutions in the VP1.

Phylogeny. From the data given in Fig. 2, the number of nucleotide differences was calculated for each pairwise comparison among the 14 isolates. Using the unweighted pair-group method with arithmetic averages (20), an evolutionary tree was constructed based upon the observed nucleotide differences (Fig. 3). According to this phylogenetic tree, the 14 isolates diverged into types A and B, which coincided with the classification of JCV isolates based upon restriction analysis (16). Archetypal and PML-type isolates fell into type A or B. Each type further diverged into several groups, with branch lengths of <3 nt, which contained archetypal and PML-type isolates. For example, PML-type isolate NY was clustered with archetypes G3 and N1. PML-types M1 and M11 were clustered with archetypal G2. PML-type M8 was clustered with archetypal N4. Finally, PML-types S1 and T1 were clustered with archetypal MY. (PML-type isolate H1 was not paired with any archetype, but because the number of archetypal isolates compared was relatively small, we might have missed the archetypal counterpart of H1.) We also analyzed the data by the neighbor-joining method (21). According to this method, four groups can be inferred unambiguously: (M1, M11, G2), (G3, N1, NY), (N4, M8), and (MY, S1, T1). Thus, it can be concluded that PML-type isolates are polyphyletic in their origin and do not constitute a unique lineage.

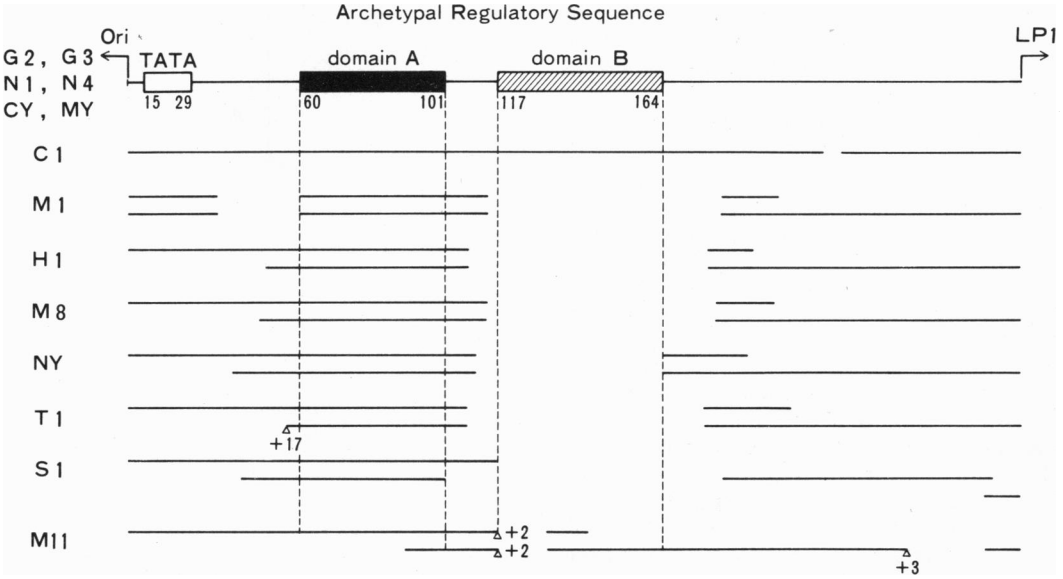

FIG. 1. Representation of the regulatory sequences of the JCV isolates studied. The structure of the archetypal regulatory region is schematically shown at the top. TATA (open box) indicates a TATA sequence (8). Domain A (closed box) indicates a sequence duplicated in most PML-type isolates except M11; domain B (hatched box) indicates a sequence deleted in most PML-type isolates except M11. [Although a short sequence named domain C (nt 181–189) is also duplicated in many PML-type isolates (18), it was not duplicated in a recent PML-type isolate (S1) (13) and is therefore omitted. Furthermore, domains A and B are 5 bp shortened compared to those indicated previously (18), because of the structures of the S1 and NY regulatory regions (13, 17)]. Numbers below each box are nucleotide numbers indicating end locations [the nucleotide numbering system is that of Frisque *et al.* (8)]. The origin of DNA replication (Ori) and the start site of the late leader protein (LP1) are shown. Isolates G2, G3, N1, N4, CY, and MY carry the archetypal regulatory sequence, as indicated. The regulatory sequences of urine isolate C1 and of 7 PML-type isolates are shown to represent deletions relative to the archetype as gaps. On reading from left to right, when a repeat is encountered, the linear representation is displaced to the line below and to a position corresponding to the sequence of the archetype. Arrowheads represent the insertion of sequences whose sizes are indicated in bp. Most regulatory sequences except NY are drawn according to the sequence data reported previously (8, 10–13, 15, 16). The regulatory sequence of NY will be described elsewhere (17).

Type	Isolate	1550	1594	1654	1664	1689	1692	1756	1771	1786	1805	1813	1818	1837	1843	1850	1870	1912	1933	1936	1940	1948	1993	2011	2161	2177	2224	2227	2239	2245
A	M1	C	C	A	G	A	G	C	C	G	A	G	G	T	G	A	G	G	G	G	C	T	C	C	C	T	A	C	T	C
	G2	C	C	A	G	A	G	C	C	G	A	G	G	T	G	A	G	G	G	G	C	T	C	C	C	T	A	C	T	C
	M11	C	C	A	C	A	G	C	C	G	A	G	G	T	G	A	G	G	G	G	C	T	C	C	C	T	A	C	T	C
	G3	C	C	A	G	G	A	C	C	G	A	G	G	T	T	G	G	G	G	G	G	T	C	C	C	T	A	C	T	C
	N1	C	C	A	G	G	A	C	C	G	A	G	G	T	T	G	G	G	G	G	G	T	C	C	C	T	A	C	T	C
	NY	C	C	A	G	G	A	C	C	G	A	G	G	T	T	G	G	G	G	G	G	T	C	C	C	T	A	C	T	C
	H1	C	C	A	G	A	A	C	C	G	A	G	G	T	G	A	G	G	G	G	C	T	C	C	C	T	A	C	T	C
	N4	T	C	T	G	A	A	C	A	T	A	G	C	C	T	G	A	G	A	A	G	T	A	C	C	C	G	T	T	T
	M8	C	C	T	G	A	A	C	A	T	A	G	C	C	T	G	A	G	A	A	G	T	A	C	C	C	G	T	T	T
	C1	C	C	T	G	A	A	T	A	G	T	G	C	T	T	A	A	G	G	G	G	T	A	A	C	C	G	T	T	T
B	CY	C	T	T	G	A	A	C	A	G	A	A	C	T	T	A	A	G	G	G	G	C	A	A	T	C	G	T	T	T
	MY	C	C	T	G	A	A	C	A	G	A	G	C	T	T	A	A	G	G	G	G	T	A	A	T	C	G	T	T	T
	S1	C	C	T	G	A	A	C	A	G	C	G	C	T	T	A	A	G	G	G	G	T	A	A	T	C	G	T	T	T
	T1	C	C	T	G	A	A	C	A	G	C	G	C	T	T	A	A	C	G	G	G	T	A	A	C	C	G	T	T	T

Type	Isolate	2251	2260	2266	2268	2269	2274	2293	2296	2311	2317	2320	2326	2338	2356	2369	2371	2386	2404	2416	2428	2429	2455	2464	2467	2494	2502	2518	2524
A	M1	C	A	G	C	T	C	C	C	G	A	G	G	C	C	T	A	T	G	C	A	G	A	G	T	C	A	G	A
	G2	C	A	G	C	T	C	C	C	T	A	G	G	C	C	T	A	T	G	C	A	G	A	G	T	C	A	G	A
	M11	C	A	G	C	T	C	C	C	T	A	G	G	C	C	T	A	T	G	C	A	G	A	G	T	C	A	G	A
	G3	C	T	G	C	T	C	C	T	T	A	G	G	C	C	T	A	T	G	C	A	G	A	G	T	C	G	G	A
	N1	C	T	G	C	T	C	C	C	T	A	G	A	T	C	T	A	T	G	C	A	G	A	G	T	C	G	G	A
	NY	C	T	G	C	T	C	G	C	T	A	G	G	C	C	T	A	T	G	C	A	G	A	G	T	T	A	G	A
	H1	C	T	G	T	A	C	C	C	T	A	G	G	C	C	C	A	T	G	C	A	G	A	G	T	T	A	G	A
	N4	C	T	A	C	T	C	G	C	T	A	A	G	C	T	T	A	C	G	C	G	G	G	A	T	T	G	G	G
	M8	C	T	A	C	T	T	G	C	T	A	A	G	C	T	T	A	C	G	C	G	G	G	A	T	T	G	G	G
	C1	C	T	A	C	T	C	G	C	T	A	A	G	C	C	T	A	C	G	C	A	G	G	A	C	T	G	G	G
B	CY	C	T	A	C	T	C	G	C	T	A	A	G	C	C	T	A	C	G	C	G	G	G	A	T	T	G	G	G
	MY	A	T	A	C	T	C	G	C	T	C	A	G	C	C	T	G	C	G	T	G	G	G	A	T	T	G	A	G
	S1	A	T	A	C	T	T	G	C	T	C	A	G	C	C	T	G	C	G	T	G	G	G	A	T	T	G	A	G
	T1	A	T	A	C	T	C	G	C	T	C	A	G	C	C	T	G	C	A	T	G	G	G	A	T	T	G	A	G

FIG. 2. Nucleotide variations in the VP1 gene among 14 JCV isolates. VP1 gene sequences (nt 1469–2533) were compared in JCV isolates. Nucleotides shown are those at positions where differences were found. Nucleotides common to all type A or B isolates examined in this study are boxed. Nucleotide numbering is that of Frisque *et al.* (8).

DISCUSSION

On the basis of our molecular phylogenetic analysis of the VP1 (capsid protein) base sequences of 14 JCV isolates, we have concluded that PML-type isolates are not members of a unique lineage. This conclusion suggests that PML-type isolates are independently generated from archetypal strains during persistence in the hosts.

Yet, all PML-type isolates directly cloned from the brain of patients with PML thus far carried rearranged regulatory sequences without exception (8, 10–13). On the other hand, isolates derived from urine and renal tissue (3, 12, 15, 16, 18, 22, 23) contained the archetypal regulatory sequence, with two minor exceptions (24). In the most instructive PML case in which isolates from the brain and kidney were compared, that of the brain had a typically rearranged regulatory sequence, whereas the kidney isolate had a regulatory sequence essentially identical with the archetype (12). These facts, together with the current phylogenetic analysis, imply that there is a correlation between the etiology of PML and structural changes of the JCV regulatory region. It remains to be determined whether the structural changes from the archetype to the PML-type enhance the growth ability of JCV in the central nervous system.

The finding that the archetypal regulatory region is not restricted to a cluster of closely related isolates but occurs in various distantly related isolates indicates that an ancestral JCV carried the archetypal regulatory region and that this structure has been conserved in the course of JCV evolution. The stabilization of the archetypal regulatory region during JCV evolution is striking, since changes in the JCV regulatory region are apparently prerequisite for the generation of PML-type JCV isolates. It is likely that, during the life cycle of JCV as a symbiote in general human hosts, a mechanism operates to prevent alteration of the archetypal regulatory region. On the other hand, it is conceivable that in PML risk groups (e.g., immunocompromised hosts), the archetypal

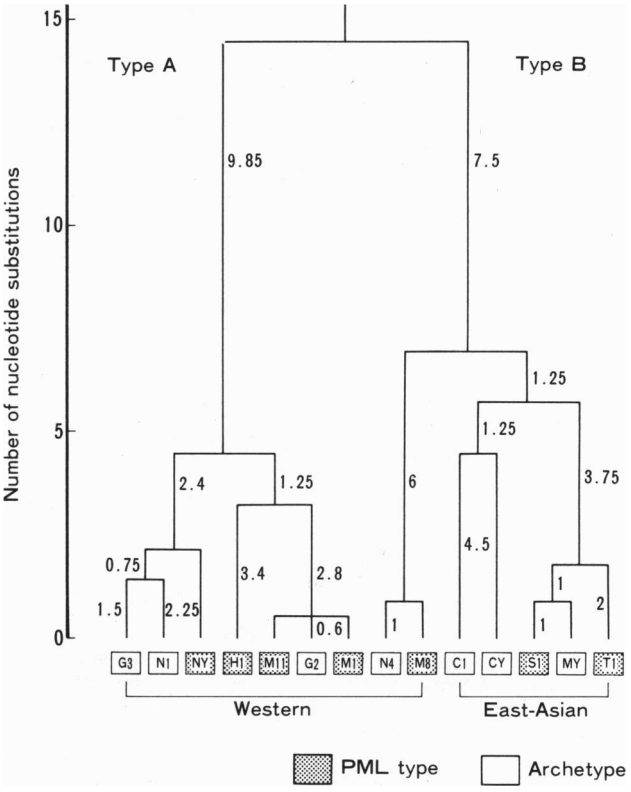

FIG. 3. Evolutionary tree relating 14 JCV isolates analyzed. From observed nucleotide differences calculated from Fig. 2, an evolutionary tree was constructed using the unweighted pair-group method with arithmetic averages (20). Numbers indicate branch lengths shown in nt.

regulatory region may become unstable and undergo changes to generate PML-type regulatory regions.

In summary, from the current phylogenetic analysis that used the VP1 coding region as a marker, we have concluded that PML-type isolates are polyphyletic in their origin. This conclusion provides strong support for the hypothesis that PML-type isolates are generated from archetypal strains during persistence in the hosts. Furthermore, this study has shown that the archetypal regulatory sequence has been conserved in the course of JCV evolution.

We are grateful to the late Dr. H. Shibuta for encouragement throughout this study, Dr. T. Fuwa for helpful suggestions, Dr. T. Shioda for critical reading of the manuscript, and Dr. R. J. Frisque for recombinant DNAs.

1. Padgett, B. L. & Walker, D. L. (1976) *Prog. Med. Virol.* **22,** 1–35.
2. Chesters, P. M., Heritage, J. & McCance, D. J. (1983) *J. Infect. Dis.* **147,** 676–684.
3. Tominaga, T., Yogo, Y., Kitamura, T. & Aso, Y. (1992) *Virology* **186,** 736–741.
4. Walker, D. L. (1985) in *Handbook of Clinical Neurology*, eds. Vinken, P. J., Bruyn, G. W., Klawans, H. L. & Koetsier, J. C. (Elsevier, Amsterdam) Vol. 47, pp. 503–524.
5. Rentier-Delrue, F., Lubiniecki, A. & Howley, P. M. (1981) *J. Virol.* **38,** 761–769.
6. Grinnell, B. W., Padgett, B. L. & Walker, D. L. (1983) *J. Virol.* **45,** 299–308.
7. Dörries, K. (1984) *Virus Res.* **1,** 25–38.
8. Frisque, R. J., Bream, G. L. & Cannella, M. T. (1984) *J. Virol.* **51,** 458–469.
9. Martin, J. D. & Foster, G. C. (1984) *J. Gen. Virol.* **65,** 1405–1411.
10. Martin, J. D., King, D. M., Slauch, J. M. & Frisque, R. J. (1985) *J. Virol.* **53,** 306–311.
11. Matsuda, M., Jona, M., Yasui, K. & Nagashima, K. (1987) *Virus Res.* **7,** 159–168.
12. Loeber, G. & Dörries, K. (1988) *J. Virol.* **62,** 1730–1735.
13. Takahashi, H., Yogo, Y., Furuta, Y., Takada, A., Irie, T., Kasai, M., Sano, K., Fujioka, Y. & Nagashima, K. (1992) *Acta Neuropathol.* **83,** 105–112.
14. Chuke, W.-F., Walker, D. L., Peitzman, L. B. & Frisque, R. J. (1986) *J. Virol.* **60,** 960–971.
15. Yogo, Y., Kitamura, T., Sugimoto, C., Ueki, T., Aso, Y., Hara, K. & Taguchi, F. (1990) *J. Virol.* **64,** 3139–3143.
16. Yogo, Y., Iida, T., Taguchi, F., Kitamura, T. & Aso, Y. (1991) *J. Clin. Microbiol.* **29,** 2130–2138.
17. Yogo, Y., Guo, J., Iida, T., Satoh, K., Taguchi, F., Takahashi, H., Hall, W. W. & Nagashima, K., *Virus Genes*, in press.
18. Yogo, Y., Kitamura, T., Sugimoto, C., Hara, K., Iida, T., Taguchi, F., Tajima, A., Kawabe, K. & Aso, Y. (1991) *J. Virol.* **65,** 2422–2428.
19. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
20. Sneath, P. H. A. & Sokal, R. R. (1973) *Numerical Taxonomy* (Freeman, San Francisco).
21. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
22. Flaegstad, T., Sundsfjord, A., Arthur, R. R., Pedersen, M., Traavik, T. & Subramani, S. (1991) *Virology* **180,** 553–560.
23. Markowitz, R.-B., Eaton, B. A., Kubik, M. F., Latorra, D., McGregor, J. A. & Dynan, W. S. (1991) *J. Virol.* **65,** 4515–4519.
24. Myers, C., Frisque, R. J. & Arthur, R. R. (1989) *J. Virol.* **63,** 4445–4449.