Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 6811–6815, September 1988
Evolution

Gene encoding capsid protein VP1 of foot-and-mouth disease virus: A quasispecies model of molecular evolution

(RNA/genetic heterogeneity/phylogeny)

JOAQUÍN DOPAZO*, FRANCISCO SOBRINO†‡, EDUARDO L. PALMA†§, ESTEBAN DOMINGO†, AND ANDRÉS MOYA*¶

*Laboratorio de Genética, Universitat de Valencia, c/Dr. Moliner 50, 46100 Burjassot, Valencia, Spain; and †Centro de Biologia Molecular, Universidad Autónoma de Madrid, Canto Blanco, 28049 Madrid, Spain

Communicated by Francisco J. Ayala, April 8, 1988

ABSTRACT **A phylogenetic tree relating the VP1 gene of 15 isolates of foot-and-mouth disease virus (FMDV) of serotypes A, C, and O has been constructed. The most parsimonious tree shows that FMDV subtypes and isolates within subtypes constitute sets of related, nonidentical genomes, in agreement with a quasispecies mode of evolution of this virus. The average number of nucleotide replacements per site for all possible pairs of VP1 coding segments is higher among representatives of serotype A than serotype C or O. In comparing amino acid sequences, the values of dispersion index (variance/mean value) are >1, with the highest values scored when all sequences are considered. This indicates an accumulation of mutations at a limited number of residues, suggesting that distributions of sequences fluctuate around points of high stability. Evolution of FMDV follows a path very distant from that of a star phylogeny, and it has not been possible to derive conclusions on constancy of evolutionary rates with the test applied to the analysis. FMDVs, as other RNA viruses, are of limited genetic complexity and their population sizes are extremely large. Their evolution concerns complex, indeterminate mixtures of genomes rather than a single, determinate species.**

Foot-and-mouth disease virus (FMDV) is an aphthovirus of the family of Picornaviridae. Its genome is a single-stranded RNA molecule that acts as mRNA *in vivo* and *in vitro* and is translated to give a polyprotein that is proteolytically processed to yield the mature structural and nonstructural viral proteins. RNA replication occurs by means of an RNA (minus strand) complementary to the genomic (plus strand) RNA (reviews in refs. 1 and 2). Although mutation rates during FMDV RNA synthesis have not been calculated, they are such that clonal FMDV populations (derived from a single genome) are genetically heterogeneous with an estimated average of two to eight substitutions per infectious genome (3, 4). This level of genetic heterogeneity is similar to that of populations of bacteriophage $Q\beta$ (5), vesicular stomatitis virus (6, 7), and influenza virus (8), among many others (9). Also, each natural FMDV isolate, the RNA of which has been subject to nucleotide sequence sampling, has proven genetically unique (9–12). Thus it is likely that for the copying of at least a significant proportion of genomic nucleotide residues, mutation rates are high and in the range of 10^{-3} to 10^{-4} substitution per nucleotide and RNA doubling (7, 13). These exceedingly high mutation rates (10^5- to 10^8-fold the corresponding values estimated for the DNA of their host organisms; see refs. 6 and 9 for reviews) confer RNA viruses the potential for very rapid evolution (6). This, in turn, determines several features of RNA viruses not easily explainable without such high mutability: unpredictable changes of behavior (virulence, antigenic properties, etc.) upon passage of viruses in natural hosts or in cell culture, multiple pathologies associated with "one" virus (for example, measles and subacute sclerosing panencephalitis; ref. 14). Moreover, the implications of such RNA genome variation extend to the evolution of the DNA-based organisms with which RNA viruses by necessity interact (4, 6). These concepts for many RNA viruses, retroviruses, and other RNA genetic elements have been reviewed (6, 9, 15).

Table 1. Main characteristics of the FMDV VP1 genes analyzed

Virus	Serotype	Subtype*	Sequenced region	Isolation†	Ref.
AVen	A	NG	Complete	Brazil, 1976	11
A5Ww	A	A5	Complete	Germany, 1951	12
A10	A	A10	Complete	Argentina, 1961	11
A22	A	A22	Complete	URSS, 1965	11
A24	A	A24	10 to end	Brazil, 1955	11
CS8	C	C1	Complete	Spain, 1970	19
CS15	C	C1	Complete	Spain, 1981	19
C3B	C	C3	Complete	Brazil, 1971	20
C3P	C	C3	Complete	Brazil, 1978	20
C3Res	C	C3	Complete	Brazil, 1955	21
OIsr	O	NG	Complete	Israel, 1981	12
OWupp	O	NG	Complete	Germany, 1982	12
O1Bfs	O	O1	Complete	England, 1968	20
O1C	O	O1	Complete	Brazil, 1958	20
O1Kb	O	O1	Complete	Germany, 1966	22

NG, not given.
*Some recent FMDV isolates have not been assigned to a subtype.
†Place, year.

One of the apparent consequences of genetic variability of FMDV is its antigenic diversity. The many isolates of FMDV analyzed to date have been grouped in seven serological types (A, C, O, SAT1, SAT2, SAT3, Asia1) and many subtypes (1). Although the molecular basis for type and subtype classification of FMDV is not known, at least some of the relevant immunogenic properties of the virus can be assigned to capsid protein VP1 (16–18).

Nucleotide sequences of the VP1-coding region and the deduced amino acid sequences of several representatives of FMDV of each of the serotypes A, C, and O are now known to permit establishment of phylogenetic relationships among them. We have constructed a phylogenetic tree and calculated the rate of nucleotide substitutions per site to define

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.

Abbreviation: FMDV, foot-and-mouth disease virus.

‡Present address: Instituto Nacional de Investigaciones Agrarias, c/Embajadores 68, 28012 Madrid, Spain.
§Present address: Departamento de Virologia, Centro de Investigaciones en Ciencias Veterinarias, INTA-Castelar, 1712 Villa Udaondo, Castelar, Buenos Aires, Argentina.
¶To whom reprint requests should be addressed.

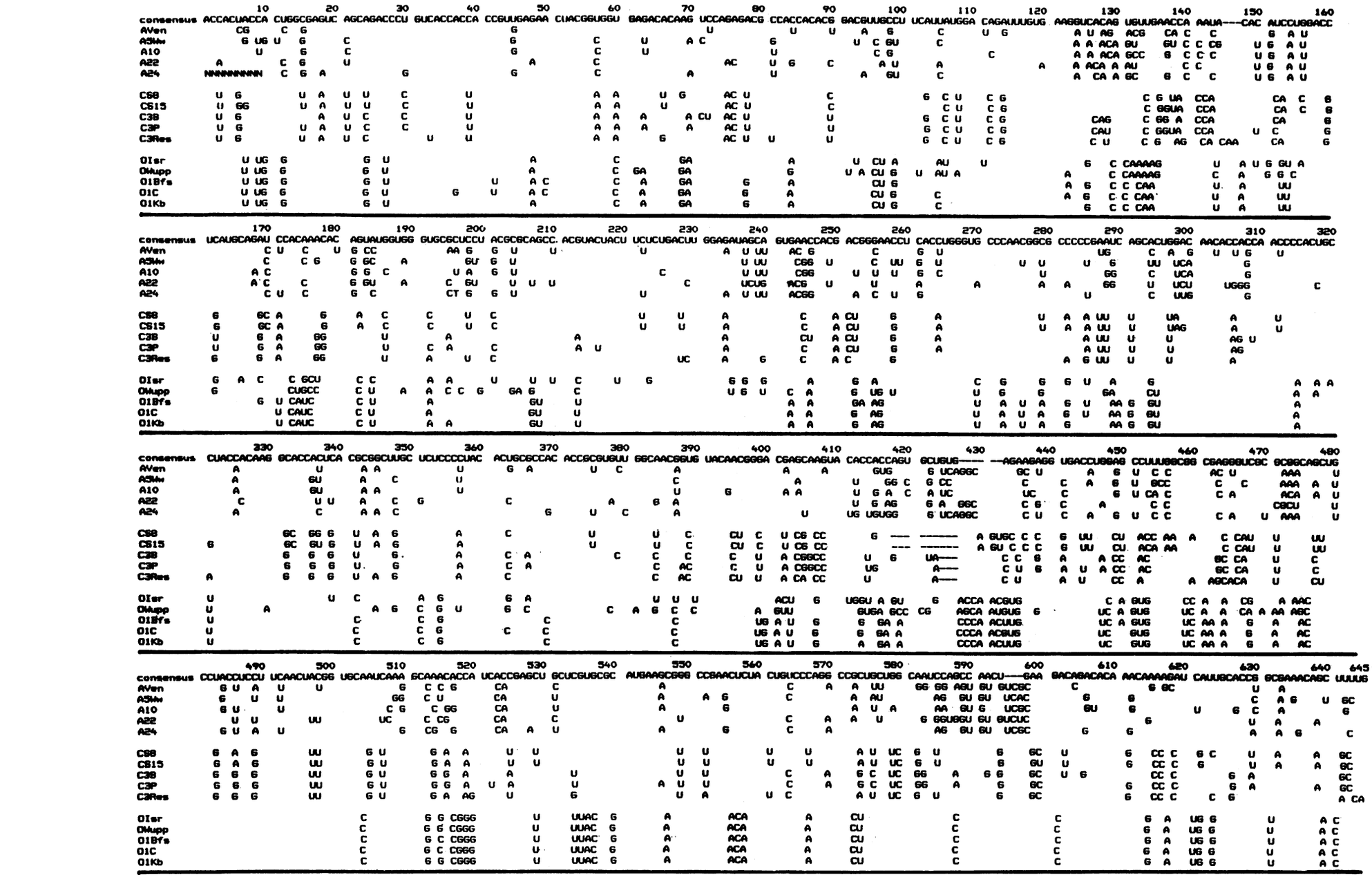

FIG. 1. Nucleotide sequence of VP1-coding segment for 15 FMDVs. N, nucleotide not determined; dash, deletion introduced to give a better sequence alignment. See also Table 1.

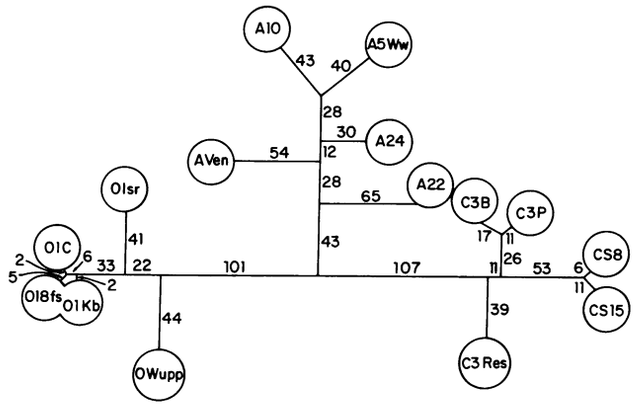

FIG. 2. Most parsimonious evolutionary tree for the 15 FMDVs inferred from analysis of RNA sequences of the VP1 genes aligned in Fig. 1.

phylogenetic relationships among the VP1-coding regions of FMDV serotypes and subtypes. The result is an extremely ramified evolutionary tree that has several implications for understanding the mode of evolution of an RNA virus.

MATERIALS AND METHODS

Viruses. The 15 FMDVs analyzed (Table 1) belong to serotypes A, C, and O. They were randomly selected among those whose entire VP1-coding region had been determined; within each serotype, viruses isolated in different years were analyzed. The conclusions reached in this report would not be altered by selecting available sequences of VP1 genes of other FMDVs for analysis (unpublished results).

Sequence Analysis. The sequence information has been analyzed following standard phylogenetic methods. We have used the program DNAPARS included in the PHYLIP package (23). Phylogenetic calculations have been performed for the complete VP1-coding region.

Kimura's parameter K (24, 25) was used to calculate the average number of nucleotide replacements per site between two genes for all possible pairs of VP1-coding segments, individually, for first plus second, for third, and for synonymous base changes in coding triplets as well as for all bases considered. Matrices of nucleotide substitutions are not given but they are available upon request.

RESULTS

Evolutionary Tree for the VP1 Gene of FMDVs of Serotype A, C, or O. The nucleotide sequences of the VP1-coding region (627–645 residues) of 15 FMDVs of serotype A, C, or O (Table 1) were aligned introducing the minimum number of deletions for maximum homology (Fig. 1). Phylogenetic analysis of the sequences shows that the most parsimonious tree contains a total of 880 mutational steps (Fig. 2). Subtypes and individual isolates within subtypes constitute sets of related, nonidentical genomes, with a higher average genetic distance between serotypes O and C.

Rate of Nucleotide Substitution. Table 2 shows the mean value of K (24, 25) between subtypes belonging to the same or to different serotype. Synonymous and third base changes are 2.9- to 5.7-fold more frequent than the sum of changes at the first and second positions, suggesting that nonsynonymous changes in the VP1 gene are subject to negative selection. Table 2 shows also that subtypes belonging to the same type are more closely related than subtypes belonging to different serotypes, probably reflecting that the serological reactivities on which the classification of FMDV into subtypes was based depended to an important extent upon VP1. The mean number of nucleotide substitutions among representatives of the A serotype is higher than the number found among representatives of serotype C or O. Moreover, the VP1 gene of serotype A holds a "central" position relative to the C and O serotypes in that the mean number of nucleotide substitutions for the third base changes, synonymous substitutions, and all positions considered is lower for A–C or A–O comparisons than for O–C (compare Fig. 2 and Table 2).

Number of Amino Acid Differences. The number of amino acid differences between VP1s (Table 3) shows a lower variance in the average number of amino acid substitutions between any two VP1s belonging to a different serotype than between VP1s within one serotype. The variance of the latter set of comparisons is about eight times higher than that of the former set, with a three times higher number of amino acid differences (67.4 ± 9.0 versus 20.9 ± 69.4; mean value of amino acid differences ± variance).

The mean number of amino acid differences along the sequence considering the antigenic types separately and together (Table 4) yields values of dispersion index (variance/mean value) always >1. The fact that the highest dispersion index is scored when all sequences are considered indicates that an accumulation of variations occurs at a limited number of residues as compared to what would be expected from a Poisson distribution of changes. Serotypes A and C as well as the three types together show a significant χ^2. Serotype O cannot be statistically tested due to the lack of degrees of freedom.

The low variance in the number of amino acid substitutions, when representatives of different serotype are compared (Table 3), and the absence of VP1s with "intermediate" amino acid sequence suggest the existence of limited points of stability for VP1 genes and the corresponding proteins. Distributions of sequences fluctuate around gravity points in the sequence space (26). Less likely, when more representatives of the many cocirculating FMDVs will be analyzed, the sequences will tend to cover the entire spectrum, blurring the differences in variance (see the *Discussion*).

Molecular Evolutionary Rates. For a situation of cocirculating, distinct genomes, the expected ratio (R) between the

Table 2. Mean number, K, and standard error of nucleotide substitutions per site between subtypes of the same or different serotypes for first plus second, third, synonymous, and all bases

Subtype	*n**	First + second	Third	Synonymous	All
Within A	10	0.089 ± 0.006	0.452 ± 0.044	0.401 ± 0.039	0.185 ± 0.013
Within C	10	0.049 ± 0.006	0.280 ± 0.039	0.255 ± 0.036	0.115 ± 0.015
Within O	10	0.054 ± 0.010	0.219 ± 0.043	0.182 ± 0.035	0.103 ± 0.020
Between A and C	25	0.239 ± 0.003	1.044 ± 0.012	0.800 ± 0.011	0.419 ± 0.003
Between A and O	25	0.281 ± 0.002	1.016 ± 0.017	0.815 ± 0.018	0.448 ± 0.002
Between O and C	25	0.262 ± 0.002	1.469 ± 0.052	1.146 ± 0.049	0.481 ± 0.004
All	105	0.204 ± 0.009	0.931 ± 0.045	0.737 ± 0.035	0.359 ± 0.015

*Number of pairs of VP1 gene segments entered into the calculation of the mean value of K (24, 25). Origins of the viral strains used for these computations are described in Table 1, and nucleotide sequences have been aligned in Fig. 1.

Table 3. Number of amino acid substitutions between any 2 of 15 VP1s

	A5Ww	A10	A22	A24	CS8	CS15	C3B	C3P	C3Res	OIsr	OWupp	O1Bfs	O1C	O1Kb
AVen	27	32	33	17	67	65	62	61	65	71	71	69	69	68
A5Ww		23	32	21	71	66	64	67	69	69	70	69	69	68
A10			33	28	67	65	66	63	66	73	74	71	71	70
A22				31	66	66	60	61	66	67	69	63	63	63
A24					69	68	63	64	68	66	66	65	65	63
CS8						9	21	20	18	67	71	68	68	66
CS15							20	21	21	67	71	67	67	65
C3B								7	16	70	74	69	70	68
C3P									13	67	72	66	67	65
C3Res										66	71	66	67	65
OIsr											21	23	22	20
OWupp												28	26	25
O1Bfs													4	5
O1C														5

Amino acid sequences are those predicted from the nucleotide sequences given in Fig. 1 and can be found in the references given in Table 1. Absence of an amino acid (deletion of a triplet upon sequence alignment) has been counted as a substitution.

variance (S^2) and the mean number of amino acid substitutions (M) should be >1. We have applied the test of Kimura (24, 25) and the one proposed by Gillespie (27) that correct for multiple substitutions of a site. For interpretation of the results (Table 5) it should be noted that the VP1-coding region did not evolve as a star phylogeny, at least when variations among subtypes are considered (Fig. 2). In this case, R tends to overestimate the true variance ratio (ref. 23, page 78). Kimura's and Gillespie's tests (Table 5) yield R values >1. The result is particularly clear when all serotypes are considered. Thus, these tests do not allow a decision as to whether rates of evolution are constant or not and whether the VP1-coding segment evolves neutrally, since we are dealing with an evolutionary system that follows a path very far from that of a star phylogeny.

DISCUSSION

FMDV VP1 Gene: A Quasispecies Mode of Evolution. Evolution of RNA viruses in nature is a complex process involving rapid emergence of mutants during RNA replication (5–10, 13–15) and frequent population bottlenecks in which probably a few particles from an infected animal are transmitted to a susceptible host where the virus is amplified to at least 10^9–10^{12} infectious units (15). Some viruses, of which FMDV is a classical example, by virtue of their wide host range, encounter multiple selective constraints (repertoire of cell types, distinct immune responses at different replication sites within one organism, etc.) that lead to differential growth rates of the continuously arising variant genomes (3, 10, 15). The result is a multitude of nonidentical FMDVs cocirculating at any given time that have been grouped serologically into serotypes and subtypes essentially by cross-reactivities with sets of standard sera that reflect antigenic properties partly embodied in VP1 (16–18). Genetic distances based on the VP1-coding region (the gene for which a sufficient number of sequences are available to provide statistically meaningful calculations) reflect the groupings of classical serology. But each group is a complex distribution of variants (3, 10, 19, 28), in agreement with the quasispecies model proposed by Eigen and colleagues to describe nucleic acid clones during early evolution of life (26, 29–32). RNA genetics deals with probability of average sequences rather than with certainties of fixed sequences (5–7, 9). High mutability, rapid generation of distributions of related, nonidentical genomes, and absence of correlation between time intervals between isolations and genetic distances are statistically documented in our phylogenetic analysis (Fig. 2). FMDVs of serotype A show genetic distances equidistant from representatives of serotypes C and O. It is possible that FMDVs A are ancestral to FMDVs C and O. The apparent higher variability previously suggested for FMDVs of type A (11) may just be a consequence of cocirculation of multiple, distinct viruses with an origin more ancient than those of type C or O.

Comparison of the VP1 gene of >50 FMDV isolates (refs. 11, 12, 19–22, and 28 and unpublished data) suggests that the number of substitutions compatible with a VP1 able to yield a viable virion may be small. Our results indicate a limited range of sequence distributions for each serotype (Fig. 2 and Table 3). We suggest that given the mutation rates and population numbers involved in FMDV replication (15), "sequence jumps" may occur that assemble a combination of substitutions compatible with a biologically competent VP1. Such rare events may be the origin of new subtypes and serotypes. Fluctuations around these main prototypes lead to the multiple variants seen in single FMDV episodes (10, 19, 21, 28, 32). Extremely ramified, "bushy" trees are expected

Table 4. Mean number and variance of amino acid differences along the VP1 protein sequence when serotypes are considered separately and together

Type	Mean value	Variance	Dispersion index
A	0.35	0.48	1.37
C	0.19	0.30	1.58
O	0.19	0.22	1.16
All	0.98	1.68	1.71

Table 5. Mean and variance estimates of amino acid substitutions according to Kimura (25) and Gillespie (27) among subtypes of a given serotype

		Kimura			Gillespie		
Serotype	n	M	S^2	R	M	S^2	R
A	5	13.85	23.84	1.72	14.79	31.42	2.12
C	5	8.30	20.53	2.47*	8.62	24.11	2.80*
O	5	8.95	66.74	7.46*	9.32	79.41	8.52*
All	15	26.95	268.82	9.27*	30.95	478.77	15.47*

M, mean number of amino acid substitutions; S^2, variance; R, variance/mean value.
*Values significantly >1. $R(n-1)$ is χ^2 distributed with $n-1$ degrees of freedom.

from the quasispecies structure (14, 15, 28–30). Palmenberg[‖] has produced such ramified trees relating different picorna viruses, as have Pallansch and Freeman[**] from sequences of the 5′ noncoding region of isolates of coxsackie B virus.

RNA Virus as a Model of Molecular Evolution. Upon application to the VP1-coding region of the procedures followed by Hayasida *et al.* (33) or by Saitou and Nei (34) a molecular clock was not observed. Different rates of substitution were observed for first plus second bases in triplets, relative to third and to synonymous substitutions (Table 2), suggesting that negative selection was acting. However, positive selection cannot be ruled out, considering the population structure for FMDV derived from our study. Although influenza virus type A NS gene has been presented as an example of a molecular clock (35), the hemagglutinin gene of influenza virus type C probably does not behave in the same way (36). Rather, its evolutionary dynamics is comparable to that of the VP1 gene of FMDV.

There are three relevant differences between viral and eukaryotic populations: (*i*) the population sizes of viruses are severalfold higher than those of eukaryotic organisms (15); (*ii*) the genome (3–30 kilobases) is much shorter; and (*iii*) mutation rates usually are 10^5- to 10^8-fold higher (9, 15). Thus, random events that may be relevant to eukaryotic populations are less likely to contribute to the evolution of viral populations. Viral genomes are small enough and population numbers are large enough to allow for many low-error mutants to be represented (5–9, 15, 29–31). In this sense we need to reconsider how deterministic is the molecular evolution of viral populations.

[‖]Palmenberg, A., Fifth Meeting of the European Group of Molecular Biology of Picornaviruses, May 31–June 6, 1987, Mallorca, Spain, abstr. P2.21.

[**]Palensch, M. A. & Freeman, C. Y., Seventh International Congress of Virology, August 9–14, 1987, Edmonton, Canada, abstr. R 16.25.

Work at the Universitat de Valéncia was supported by grants to A.M. from the Dirección General de Investigación Científica y Técnica, Project PB86-0517, and Conselleria de Cultura, Educación y Ciencia de la Communidad de Valencia (Spain). Work at the Centro de Biología Molecular, Madrid, was supported by the Comision Asesora para la Investigación Científica y Técnica, Fondo de Investigaciones Sanitarias, and Consejo Superior de Investigaciones Científicas (Spain).

1. Bacharad, H. L. (1977) in *Virology in Agriculture*, Beltsville Symposia in Agricultural Research, ed. Romberger, J. A. (Allanheld, Osmun, Montclair), Vol. 1, pp. 3–32.
2. Sangar, D. V. (1979) *J. Gen. Virol.* **45,** 1–13.
3. Sobrino, F., Dávila, M., Ortín, J. & Domingo, E. (1983) *Virology* **128,** 310–318.
4. de la Torre, J. C., Martinez-Salas, E., Díez, J., Villaverde, A., Gebauer, F., Rocha, E., Dávila, M. & Domingo, E. (1988) *J. Virol.* **62,** 2050–2058.
5. Domingo, E., Sabo, D., Taniguchi, T. & Weissmann, C. (1978) *Cell* **13,** 735–744.
6. Holland, J., Spindler, K., Horodysky, F., Grabau, E., Nichol, S. & VandePol, S. (1982) *Science* **215,** 1577–1585.
7. Steinhauer, D. & Holland, J. (1986) *J. Virol.* **57,** 219–228.
8. Parvin, J. D., Moscona, A., Pan, W. T., Leider, J. M. & Palese, P. (1986) *J. Virol.* **59,** 377–383.
9. Domingo, E. & Holland, J. (1988) in *RNA Genetics*, eds. Domingo, E., Holland, J. & Ahlquist, P. (CRC, Boca Raton, FL), Vol. 3, pp. 3–36.
10. Domingo, E., Dávila, M. & Ortín, J. (1980) *Gene* **11,** 333–346.
11. Weddell, G. N., Yansura, D. G., Dowbenko, D. J., Hoatlin, M. E., Grubman, M. J., Moore, D. M. & Kleid, D. G. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2618–2622.
12. Beck, E. & Strohmaier, K. (1987) *J. Virol.* **61,** 1621–1629.
13. Batschelet, E., Domingo, E. & Weissmann, C. (1976) *Gene* **1,** 27–32.
14. Cattaneo, R., Schmid, A., Rebmann, G., Baczko, K., Meulen, V., Bellini, W. J., Rozenblatt, S. & Billeter, M. A. (1986) *Virology* **154,** 97–107.
15. Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J. C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Breña, P., Villanueva, N., Nájera, R., VandePol, S., Steinhauer, D., DePolo, N. & Holland, J. (1985) *Gene* **40,** 1–8.
16. Strohmaier, K., Franze, R. & Adam, K. H. (1982) *J. Gen. Virol.* **59,** 295–303.
17. Bittle, J. L., Houghteh, R. A., Alexander, H., Shinnick, T. M., Sutcliffe, J. G., Lerner, R. A., Rowlands, D. J. & Brown, F. (1982) *Nature (London)* **298,** 30–33.
18. Pfaff, E., Mussgay, M., Böhm, H. O., Schultz, G. E. & Schaller, H. (1982) *EMBO J.* **1,** 869–874.
19. Sobrino, F., Palma, E. L., Beck, E., Dávila, M., de la Torre, J. C., Negro, P., Villanueva, N., Ortín, J. & Domingo, E. (1986) *Gene* **50,** 149–159.
20. Cheung, A., Delamarter, J., Weiss, S. & Kupper, H. (1983) *J. Virol.* **48,** 451–459.
21. Piccone, M. E., Kaplan, G., Giavedoni, L., Domingo, E. & Palma, E. L. (1988) *J. Virol.* **62,** 1469–1473.
22. Beck, E., Forss, S., Strebel, K., Cattaneo, R. & Feil, G. (1983) *Nucleic Acids Res.* **11,** 7873–7885.
23. Felsenstein, J. (1985) *Evolution* **39,** 783–791.
24. Kimura, M. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 454–458.
25. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
26. Eigen, M. & Biebricher, C. K. (1988) in *RNA Genetics*, eds. Domingo, E., Holland, J. & Ahlquist, P. (CRC, Boca Raton, FL), Vol. 3, pp. 211–245.
27. Gillespie, J. M. (1986) *Mol. Biol. Evol.* **3,** 138–155.
28. Martinez, M. A., Carrillo, C., Plana, J., Mascarella, R., Bergadá, J., Palma, E. L., Domingo, E. & Sobrino, F. (1987) *Gene* **62,** 75–84.
29. Eigen, M. (1971) *Naturwissenchaften* **58,** 465–523.
30. Eigen, M. & Schuster, P. (1979) *The Hypercycle* (Springer, Berlin), pp. 1–24.
31. Biebricher, C. K. (1983) in *Evolutionary Biology*, eds. Hecht, M. K., Wallace, B. & Prance, G. T. (Plenum, New York), Vol. 16, pp. 1–52.
32. Anderson, E. C., Underwood, B. O., Brown, F. & Ngichabe, C. K. (1985) *Vet. Microbiol.* **10,** 409–423.
33. Hayasida, H., Toh, H., Kikuno, R. & Miyata, T. (1985) *Mol. Biol. Evol.* **2,** 289–303.
34. Saitou, N. & Nei, M. (1986) *Mol. Biol. Evol.* **3,** 57–74.
35. Buonaugurio, D. A., Nakada, S., Parvin, J. D., Krystal, M., Palese, P. & Fitch, W. M. (1986) *Science* **232,** 980–982.
36. Buonaugurio, D. A., Nakada, S., Desselberger, U., Krystal, M. & Palese, P. (1985) *Virology* **146,** 221–232.