J Mol Evol (1992) 35:223–229

Journal of Molecular Evolution


# Does the VP1 Gene of Foot-and-Mouth Disease Virus Behave as a Molecular Clock?

Santiago F. Elena, Fernando González-Candelas, and Andrés Moya

Departament de Genètica and Servei de Bioinformàtica, Facultat de Ciències Biològiques, Universitat de València, Dr. Moliner 50, Burjassot, 46100 València, Spain

**Summary.** We have carried out a phylogenetic study of the evolution of the VP1 gene sequence from different serological types and subtypes of foot-and-mouth disease virus (FMDV). The maximum-likelihood method developed by Hasegawa and coworkers (Hasegawa et al. 1985) for the estimation of evolutionary parameters and branching dates has been used to decide between alternative models of evolution: constant versus variable rates. The results obtained indicate that a constant rate model, i.e., a molecular clock, seems to be the most plausible one. However, additional information suggests the possibility that the appearance of serotype CS has been accompanied by an episode of rapid evolution (Villaverde et al. 1991). We discuss the possibility that this evolution of RNA viruses was due to episodic positive Darwinian selection, which would have helped the new variant to escape the immunogenic pressure from the hosts.

**Key words:** Food-and-mouth disease virus — Serotypes — Molecular clock — Periodic selection — Maximum-likelihood method

## Introduction

The absence of proofreading in the RNA replicase (Domingo and Holland 1988) causes the evolution of RNA viral genes to be much faster than that of nuclear genes of eukaryotic organisms. This makes RNA viruses interesting materials for the study of molecular evolution. The latter has been studied during the last decade under two different antagonistic views, neutral and positive Darwinian evolution. As mentioned by Gojobori et al. (1990), viral genes are particularly well suited to examine the molecular clock hypothesis and, consequently, the most widely accepted theoretical justification of it, i.e., the neutral theory. Under this theory, an exact molecular evolutionary clock would be expected if the mutation rates for neutral alleles per gene of a given molecule were exactly equal for all organisms at all times. So, selection is viewed as a disturbing factor of the previous process, because the alteration of the selective constraints on each molecule could change the observed fixation rate of mutants.

The selectionist theory claims a relevant role for positive Darwinian evolution at the molecular level (Fitch et al. 1991 and references therein), especially in the case of molecular evolution of some RNA viral genes (Dopazo et al. 1988; Eigen and Biebricher 1988). The evolution of a gene subject to positive selection would not show clocklike behavior, as the different selection intensities would be reflected in different speeds of fixation of mutations along the branches of the corresponding phylogenetic tree. As stated by Fitch et al. (1991), the neutral theory is an important contribution to the understanding of evolution at the molecular level, but it is time now to turn our attention to finding out under what circumstances positive selection is acting.

Several methods have been proposed for comparing evolution rates in an evolutionary tree (reviewed by Gillespie 1986). However, most of the methods described in that paper allow only for the testing of a "global" clock for the whole phylogeny or the singling out of a deviating branch. Hasegawa and his coworkers (Hasegawa et al. 1985; Kishino

*Offprint requests to:* A. Moya

and Hasegawa 1990) have developed a maximum-likelihood method that allows not only the direction of different rates along branches but also the dating of splits when rates vary. As this method is based on an explicit evolutionary model, it is advantageous over other statistical methods that merely compare final trees (Felsentein 1981, 1984; Kimura 1983).

In the present paper we have tried to combine the statistical analysis of molecular evolutionary rates in the phylogeny of the VP1 gene of foot-and-mouth disease virus (FMDV) with relevant biological phenomena in the evolutionary history of this virus, such as the appearance of new serological types and subtypes.

FMDV is an aphtovirus of the Picornaviridae family (Palmenberg 1988). Its genome consists of a single-stranded RNA molecule that acts as an mRNA in vivo and in vitro, and it is translated to give a polyprotein that is proteolytically processed to yield the mature structural and nonstructural viral proteins.

One of the apparent consequences of genetic variability of FMDV is its antigenic diversity. The many isolates of FMDV analyzed to date have been classified into seven serological types (A, C, O, SAT-1, SAT-2, SAT-3, and Asia 1) and many subtypes. Although the molecular bases for type and subtype classification of FMDV are not clear, at least some of the relevant immunogenic properties of the virus can be assigned to capsid protein VP1 (see Domingo et al. 1990 for a recent review).

In the present work we check the constancy of evolutionary rates in different branches corresponding to the phylogenetic tree of the VP1 protein gene of five FMDV isolates, representing three different serotypes and two subtypes of one of them. We are specially interested in answering the following null hypothesis: is the appearance of a new type or subtype accompanied by a change in its evolutionary rate?

## Methods

We have used the maximum likelihood approach developed by Hasegawa et al. (1985). The convention of classifying nucleotide substitutions into two categories, transitions (U to C and A to G) and transversions (U or C to A or G) is followed, with the transition rate exceeding the transversion rate in most cases. It is assumed that each site evolves according to a Markov process in which a nucleotide $x$ is replaced by another nucleotide $y$ in an infinitely short interval of time, $dt$, with probability given by

$$P_{xy}(dt) = \begin{cases} \alpha\pi_y dt & \text{for transition} \\ \beta\pi_y dt & \text{for transversion} \end{cases} \quad (1)$$

where $\pi_y$ is the proportion of nucleotide $y$ in the sequences. The parameters $\alpha$ and $\beta$ can vary between different lineages. It is assumed that all the nucleotide sites evolve independently and are homogeneously variable with substitution probability given by Eq. (1). The expectations of the numbers of transversion and transition type differences between two sequences $i$ and $j$ with $r$ nucleotides are denoted by $\overline{V_{ij}(t)}$ and $\overline{S_{ij}(t)}$, respectively. If $i$ and $j$ diverged $t$ time units ago and $\alpha$ and $\beta$ do not differ between the two lineages, they can be derived under this model as follows

$$\overline{V_{ij}(t)} = 2fr\pi_Y\pi_R[1 - \exp(-2\beta t)] \quad (2)$$

$$\overline{S_{ij}(t)} = 2fr\{(\pi_T\pi_C + \pi_A\pi_G) + (\pi_T\pi_C\pi_R\pi_Y + \pi_A\pi_G\pi_Y\pi_R) \cdot\exp(-2\beta t) - (\pi_T\pi_C\pi_Y)\exp[-2t(\alpha\pi_Y + \beta\pi_R)] - (\pi_G\pi_A/\pi_R)\exp[-2t(\alpha\pi_R + \beta\pi_Y)]\} \quad (3)$$

where $f$ is the probability that a given site is variable and $\pi_Y = \pi_T + \pi_C$ and $\pi_R = \pi_A + \pi_G$.

When the evolutionary rates differ among different lineages, $\overline{V_{ij}(t)}$ and $\overline{S_{ij}(t)}$ can be given by a similar formulae. Consider an evolutionary tree, where $i$ and $j$ diverged $t$ time units ago, and evolution proceeded at first with transition and transversion rates $\alpha_1$ and $\beta_1$ along the two lineages. However, along the line leading to $i$, the rates changed to $\alpha_2$ and $\beta_2$, $t_1$ time units ago and again changed to $\alpha_3$ and $\beta_3$ at $t_2$ ($t > t_1 > t_2$). In this case, $2\alpha t$ and $2\beta t$ in Eq. (2) and (3) are replaced by $(2t - t_1)\alpha_1 + (t_1 - t_2)\alpha_2 + t_2\alpha_3$ and $(2t - t_1)\beta_1 + (t_1 - t_2)\beta_2 + t_2\beta_3$, respectively. A rooted tree with $n$ contemporary sequences [i.e., operational taxonomic units (OTUs)] has $n - 1$ branching points. Fixing a date as calibration point, there are $n - 2$ remaining branching points, the dates of which ($t_i$) are unknown parameters. The unknown parameters $\alpha_1$, $\beta_1$, and $t_i$ are estimated from the data by a maximum-likelihood method (Hasegawa et al. 1985).

The vector of parameters $\Theta = [t_k\ (k \neq c), f, \alpha, \beta]^T$ (superscript T denotes transposed vector) that minimizes the weighted sum of squares between the observed and the expected, under the previously described model, transition and transversion differences among sequences can be estimated by a generalized least-squares method as developed in Hasegawa et al. (1985) and Kishino and Hasegawa (1990).

The generalized least-squares method uses Newton's algorithm for finding the convergent values of the components of vector $\Theta$. It is necessary to provide a good initial estimate of $\Theta$ components ($\alpha_k$, $\beta_k$, and $t_k$). These estimates were obtained by trial and error from a plot of transition versus transversion as it will be illustrated below.

Based on the information theory, Akaike (1974) derived a criterion for the comparison of nonnested models, which is very useful when comparing different tree topologies or evolutionary models with varying rates. The AIC (Akaike information criterion) statistic for a given model is defined as

$$\text{AIC} = -2(\text{estimated log-likelihood}) + 2(\text{number of free parameters}) \quad (6)$$

and it provides the basis for model selection. The better the fit of the model to the data, the lower is the first term. On the other hand, the more complex the model, the higher is the second term. The model with minimum AIC is considered to be the most appropriate model (Akaike 1974; Hasegawa et al. 1990; Kishino and Hasegawa 1990). Using this criterion, it can be determined whether or not additional parameters should be introduced in a given model.

As the five sequences were not isolated at the same time, and given the high evolutionary rate for RNA viruses, it is necessary to correct the estimated parameters $\alpha$ and $\beta$. The correction performed was $\Theta' = \Theta t/(t - \delta t)$ where $t$ is the time of divergence between two sequences estimated by this method and $\delta t$ is the difference between the corresponding isolation dates. The same correction can be applied to the global evolutionary rate, $\mu$, being $\mu = 2[(\pi_T\pi_C + \pi_A\pi_G)\alpha + \pi_R\pi_Y\beta]$.

```
CS14    ACTACGGCCA CCGGTGAATC TGCTGACCCC GTCACCACTA CCGTTGAGAA CTACGGAGGA
CS18
C3P           A    T           C
OIsr        C  TT TG  G  C  G   A  G  T  T          C          A          T  C
SAT-3       A  TAG G  G     GGG A  C   GT           G A   CACC C  C T  T  T

CS14    GAGACTCAGG TCCAACGTCG CCACCACACC GACGTTGCCT TCGTTCTTGA CCGGTTTGTG
CS18
C3P       A  A  AA                      T
OIsr         A   A        GA A        A     G   T  CT A     A AT G    TA A
SAT-3     A  GTGAGT    T C A A    T  G        A     G AG    TC G  G     A A  CACA

CS14    AAGGTCACAG TGTCGGACAA CCAACACACA CTCGACGTGA TGCAGGCACA CAGAGATAAT
CS18                    T
C3P           CAT         GT          T  C   G     T          T      AG  C G
OIsr        G    C CAAAA    C  AATTA TGTG T A   C       A  CTC C  GCTC C C
SAT-3     C C   GGCA AAGTCA  G  GAGCAGA  C A TTC C       G CA GA   GGA C C CG

CS14    ATCGTGGGCG CGCTTCTTCG CGCGGCCACG TACTACTTTT CTGATTTGGA AATAGCAGTG
CS18                          A              T
C3P       T           A  C  C      A     A  T     C       C
OIsr    C G     A     A  C          T  T  C        T  CG     C       GG G  G
SAT-3   C G     A     AA CTCCG  TT C              T          G   CC      G T  CA C

CS14    ACCCACACTG GGAAGCTCAC ATGGGTGCCC AACGGTCACC AGTTCTGCAA ACAACACAAC
CS18
C3P                                                C C                G          G
OIsr     AA    GAG    A  C       C      C  G       G G   T AA A   GG            C
SAT-3   TTGGG    C  C TG GCGG    C                    G   CACA AC GG   GTGG GG

CS14    AACCCCACTG CCTACCATAA GGGCCCGGTG ACTCGACTGG CTCTCCCATA CACCGCGCCA
CS18      T                                                                 T
C3P                     C       A   C        G                             A
OIsr          A  A  AT     C      CA  AC T    C  G  T   A  G  C        G   A
SAT-3         AGT   TTC  TCA    A  AT     C GTC  CT C  GCT G  C        A

CS14    CACCGCGTGT TGGCTACGGC GTACACTGGT ACTACGACTA CACCGCCAGT GCATTGGCAC
CS18         T
C3P                   C  A          C       A   G         T             C A  G
OIsr                     T T T      G  A     G AGG    TGGTAA GT     GC     G
SAT-3       TG    T      A  T T T    TG CA     GTA G     GT AA C   CA C     G

CS14    GACGAGCATG CGCCGACATC GTTCAACTTT GGGCAGTTAA GGCGAAACAA ACTGAGTTGC
CS18
C3P      G  G   C           G                                A   G         A   C
OIsr     G     A C       T  C   C          AC      CA C    A     GC GG     C
SAT-3   A    GTT C   AA   C GGA  A           C    AAGA  GT T TTG  G G GG     TGC T

CS14    TCGTGCGCAT GAAGCGTGCT GAACTCTATT GTCCAGGCCG ATTCTTCCGA TTCAGCCAGG
CS18
C3P      T              A                    C   C    A    G C       G   C  A
OIsr     TTAC  G          A G  C     ACA  C                C T G  GG A    C   C    TA
SAT-3   A TAT              C       G  T  C    C    C    C GC  G GA G   AT CA   CA

CS14    TAGACACAAG CAACCGCTCG TCGCACCTGC AAAGCAACTG CTG
CS18          T                 G
C3P     C                   A  T  G   A   G  A
OIsr    C          A   GAAAA T   G     G T G   A   GACT T
SAT-3   C  GT       AC    A  G     AA      A C     GA   TGC
```

**Fig. 1.** Partial alignment (excluding gaps) for the VP1 sequences of the following strains: CS14 (isolated in December 1979; serotype C), CS18 (February 1980; serotype C), C3P (November 1978; serotype C), OIsr (April 1981; serotype O), and SAT-3 (January 1965; serotype SAT-3). Only those positions used for the present analyses have been included.

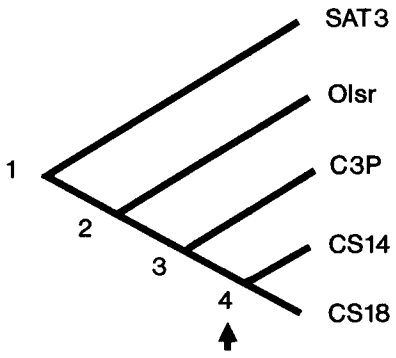


**Fig. 2.** Unrooted maximum-likelihood tree obtained for the five sequences. The arrow indicates the branching point taken as a reference, fixed time.

## Results

Five VP1 sequences have been used in this study. The strains, serotypes, and the corresponding isolation dates along with their alignment are shown in Fig. 1. The alignment was performed by means of the hierarchical clustering algorithm (Higgins and Sharp 1988, 1989). Gaps are not considered in the Markov process (Hasegawa et al. 1985), and consequently all the positions that involved a gap in at least one sequence were discarded for the analyses.

**Table 1.** Transition (below diagonal) and transversion (above diagonal) hemimatrices for the five sequences

| | CS14 | CS18 | C3P | OIsr | SAT-3 |
|---|---|---|---|---|---|
| CS14 | — | 0 | 13 | 106 | 140 |
| CS18 | 8 | — | 13 | 106 | 140 |
| C3P | 61 | 63 | — | 109 | 143 |
| OIsr | 89 | 93 | 80 | — | 136 |
| SAT-3 | 119 | 125 | 123 | 124 | — |

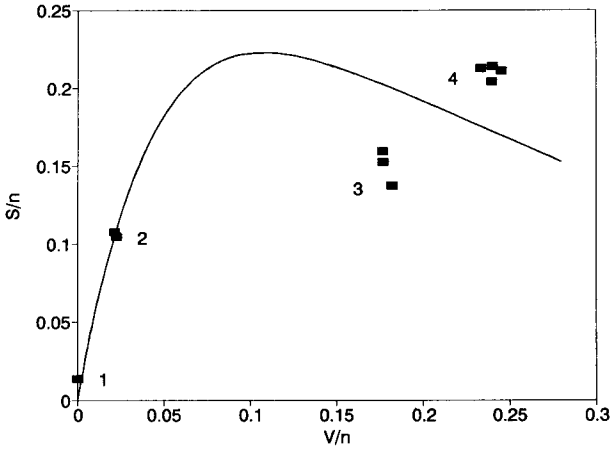


**Fig. 3.** Relationship between the number of transitions ($S/n$) and transversions ($V/n$) per nucleotide site for the five sequences used in the analysis. The numbering of groups corresponds to that of nodes in Fig. 2.

The maximum-likelihood method (Felsenstein 1981) was used to infer the phylogenetic tree (Fig. 2). The topologies obtained under the molecular clock assumption (using program DNAMLK in package PHYLIP version 3.4) and without this assumption (using DNAML in the same package) are the same; hence, the molecular clock assumption was tested by means of the log-likelihood ratio test proposed by Felsenstein. No significant departure from the molecular clock was obtained ($\chi^2 = 2.28$ with 3 df).

The observed number of transitions and transversions between each pair of sequences is shown in Table 1. From these data, we estimated the initial values for parameters $\alpha$, $\beta$, and $t_k$. The estimation is performed by trial and error using the expressions for $\overline{S_{ij}}$ and $\overline{V_{ij}}$ to fit the observed data (Fig. 3). It is unfortunate that Newton's method, used in the iterative estimation of the parameter's vector $\Theta$, does not converge unless good initial estimates are given. It is also important to realize that this method allows finding local optima, but not global optima, for the parameters, and therefore different initial estimates may lead to radically different outcomes. We have tried to cover as wide a range for $\alpha$ and $\beta$ initial estimates as possible.

Four different evolutionary models have been tested. We are interested in testing the null hypothesis that the appearance of a new serotype corresponds to a peak shift in the adaptive landscape of the virus, and that this shift is accompanied by a sudden increase in the evolutionary rate, thus allowing the fast establishment of a new serological type. The assumption of a molecular clock leads to the first model, with constant rates along all the branches. The second, third, and fourth models correspond to the different steps in this hypothesis: the appearance of a new serological type or subtype can be accompanied by a change in the evolutionary rate. So, model 2 corresponds to a set of parameters ($\alpha_1$, $\beta_1$) along the branches 1-SAT-3 and 1-2, and another set ($\alpha_2$, $\beta_2$) for the remaining branches. The acceptance of this model would imply that SAT-3 would have evolved at a different rate than the other sequences. Model 3 reflects the idea that each serotype (SAT, O, and C) evolved with different rates. Finally, in model 4 the possibility that subtype C3P evolved at different rates than subtype CS is tested.

The calibration date was taken as the estimated divergence point between CS14 and CS18. These are the most closely related pair of sequences, and to obtain good estimates of $\alpha$ and $\beta$ it is necessary

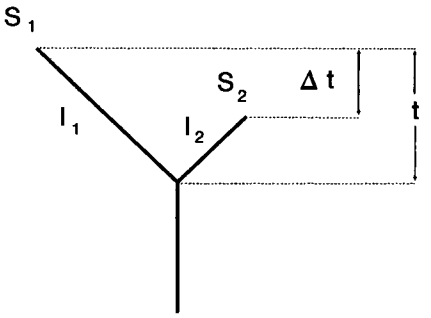


**Fig. 4.** Correction of estimated rates for branches corresponding to different isolation time for two OTUs.

**Table 2.** Branching dates (years* $10^{-2}$) and evolutionary rates estimated from VP1 sequences

| Parameter | Model | | | |
|---|---|---|---|---|
| | Molecular clock | Two rates | Three rates | Four rates |
| Rates in branches | | | | |
| 1-SAT-3 | $\alpha_1\beta_1$ | $\alpha_2\beta_2$ | $\alpha_2\beta_2$ | $\alpha_2\beta_2$ |
| 1-2 | $\alpha_1\beta_1$ | $\alpha_2\beta_2$ | $\alpha_2\beta_2$ | $\alpha_2\beta_2$ |
| 2-OIsr | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_3\beta_3$ | $\alpha_3\beta_3$ |
| 2-3 | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_3\beta_3$ | $\alpha_3\beta_3$ |
| 3-C3P | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_4\beta_4$ |
| 3-4 | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_4\beta_4$ |
| 4-CS14 | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ |
| 4-CS18 | $\alpha_1\beta_1$ | $\alpha_1\beta_2$ | $\alpha_1\beta_1$ | $\alpha_1\beta_1$ |
| Branching dates | | | | |
| $t_1$ | 0.6789 ± 0.2881 | 0.7692 ± 0.1216 | 0.8968 ± 0.0606 | 0.7319 ± 0.0604 |
| $t_2$ | 0.4067 ± 0.1796 | 0.4224 ± 0.2132 | 0.4636 ± 0.1326 | 0.2788 ± 0.0531 |
| $t_3$ | 0.0312 ± 0.0117 | 0.0301 ± 0.0120 | 0.0313 ± 0.0073 | 0.0112 ± 0.0009 |
| $t_4$ | 0.0025 | 0.0025 | 0.0025 | 0.0025 |
| Rates | | | | |
| $\alpha_1$ | 15.3452 ± 6.4042 | 15.9276 ± 7.0398 | 15.7779 ± 3.9099 | 19.4394 ± 6.3378 |
| $\alpha_2$ | — | 45.3319 ± 5.6005 | 9.4322 ± 4.1826 | 58.9858 ± 5.7161 |
| $\alpha_3$ | — | — | 24.8302 ± 8.9043 | 37.0904 ± 6.6165 |
| $\alpha_4$ | — | — | — | 43.9139 ± 5.0866 |
| $\beta_1$ | 1.0750 ± 0.4175 | 1.0130 ± 0.4479 | 1.0347 ± 0.2407 | 0.0000 ± 1.0668 |
| $\beta_2$ | — | 0.8923 ± 0.0840 | 0.5646 ± 0.2760 | 0.8134 ± 0.0791 |
| $\beta_3$ | — | — | 0.7681 ± 0.3386 | 1.4850 ± 0.2849 |
| $\beta_4$ | — | — | — | 2.8501 ± 0.8039 |
| $\mu_1$ | 4.2756 ± 1.7688 | 4.3865 ± 1.9388 | 4.3612 ± 1.0734 | 4.7354 ± 2.0773 |
| $\mu_2$ | — | 5.7148 ± 1.8955 | 1.7009 ± 0.9547 | 12.4533 ± 3.1015 |
| $\mu_3$ | — | — | 8.2037 ± 4.2435 | 3.8276 ± 1.7011 |
| $\mu_4$ | — | — | — | 0.4259 ± 0.0931 |
| SSR | 53.008 | 117.324 | 255.463 | 302.445 |
| df | 15 | 35 | 63 | 99 |
| $P$ | <0.00001 | <0.00001 | <0.00001 | <0.00001 |
| $\sigma$ | 1.6584 | 1.6713 | 1.8836 | 1.6582 |
| AIC(2) | 98.5557 | 105.6396 | 99.7504 | 113.6911 |

$\alpha$ and $\beta$ values are shown as obtained from the least-squares method, whereas $\mu$ values have been corrected to account for the differences in isolation dates. The proportion of variable sites ($f$) is assumed to be constant and equal to 0.5935

**Table 3.** Synonymous (below diagonal) and nonsynonymous (above) evolutionary rates (expected number of changes per nucleotide position) for VP1 sequences

| | CS14 | CS18 | C3P | OIsr | SAT-3 |
|---|---|---|---|---|---|
| CS14 | — | 0.0046 (0.0032) | 0.0987 (0.0164) | 0.2988 (0.0310) | 0.5420 (0.0504) |
| CS18 | 0.0436 (0.0181) | — | 0.0987 (0.0164) | 0.3062 (0.0316) | 0.5532 (0.0515) |
| CP3 | 0.2869 (0.0532) | 0.3069 (0.0559) | — | 0.2887 (0.0302) | 0.5957 (0.0555) |
| OIsr | 1.1355 (0.1805) | 1.1883 (0.1937) | 1.0397 (0.1574) | — | 0.5507 (0.0500) |
| SAT-3 | 1.3082 (0.2209) | 1.4603 (0.2716) | 1.2111 (0.2015) | 1.2742 (0.2292) | — |

Standard errors are shown in parentheses below the corresponding mean value

to compare sequences for which transitions and transversions increase linearly with time (Hasegawa et al. 1985). This branching date was estimated as follows. From the branching point, two lineages, $S_1$ and $S_2$, have accumulated $l_1$ and $l_2$ differences in the respective branches (Fig. 4). If both branches had the same length, i.e., had accumulated the same number of differences, $k$, the global rate, would equal $2l/t$, but if $l_1 \neq l_2$ and the isolation times are different by $\delta t$ units, then $k_1 = l_1/t$ and $k_2 = l_2/(t - \delta t)$. Solving for $t$ leads to $t = l_1\ \delta t/(l_1 - nl_2)$, where $n = k_1/k_2$. If $k_1 = k_2$ (molecular clock), then $n = 1$.

The smallest AIC value (see Table 2) was obtained for the molecular clock model. In this model, as in the other models tested, independence of evolution in neighbor sites cannot be assumed and AIC(2) values are the ones to be compared ($P$ values are around $10^{-20}$, except for model one, with $P \approx 10^{-5}$). As a consequence, standard error estimates for parameters in Table 2 must be multiplied by the $\sigma$ value corresponding to their model (Kishino and Hasegawa 1990). Under the constant rate assumed in the molecular clock model, the estimated branching dates for the four nodes in Fig. 2, correspond to 60, 40, 4.3, and 1.8 years, for nodes 1, 2, 3, and 4, respectively, before the isolation date of the most recent isolate (OIsr, April 1981).

The other models show larger values for AIC(2), and under Akaike's criterion such models must be discarded and constancy of evolutionary rates must be assumed. However, AIC for model 3 is very close to that of model 1, which could imply that discrimination between these two models is not easy. There is no pattern when global evolutionary rates are compared among models with increasing number of parameters. $\mu_1$ remains roughly constant (Table 2), but $\mu_2$ decreases and then increases from model 2 to model 3 and from this to model 4. This makes it impossible to assign a relative speed to each branch, thus undermining the original hypothesis.

Once constancy of evolutionary rates has been demonstrated, it is possible to calculate the corresponding rates for synonymous and nonsynonymous positions. These values, computed according to the method of Li et al. (1985), are shown in Table 3. As expected according to the neutral theory, rates for synonymous substitutions are always higher than for nonsynonymous substitutions.

## Discussion

The observed clock behavior of these sequences can be interpreted in two ways. First, we can disregard the null hypothesis (see above) and accept that the appearance of a new serotype is not accompanied by an increase in the evolutionary rates, hence the constancy of evolutionary rates observed in the lineages of three different serotypes. Second, the null hypothesis could be accepted if the change in evolutionary rates was merely a sudden burst of change, during a very short period of time, followed by a new deceleration when the new serotype becomes the dominant one in the new environment. We must take into account that RNA viruses have extremely small generation times and very high replication rates. Consequently, this second interpretation could only be tested using a quite different set of data, such as differences from one cell culture passage to another or such like. The time scale involved cannot be months but weeks or even days.

The three rates model shows an AIC value very close to that of the constant rate model. It would be tempting to accept this model as the best one, as it gives clear support to the null hypothesis. If that were the case, the appearance of a new serotype would be accompanied by a change in the evolutionary rate, especially in the transversion rate. It is interesting to note that the change in evolutionary rate does not imply necessarily an acceleration of it. In the three rates model, the branch that leads to serotype OIsr, after split number 2 in Fig. 2, shows a lower rate than the branches leading to serotype SAT-3 and to serotype C, after split number 3. This variation could be due to a change in the selection pressure on the molecule, by means of changes in the conformation of the epitope recognized by antibodies from the hosts.

Gillespie (1984) has developed a model where selection acts episodically. In other words, in some periods positive Darwinian selection might be present, and in others, mutations might accumulate according to a pattern of neutral evolution. Episodic selection or some type of periodic selection (Maruyama and Birky 1991) could have an important

role in the evolution of RNA viruses, irrespective of our capacity to detect it. Fitch et al. (1991) have been able to provide evidence of the presence of periodic positive Darwinian selection in the evolution of structural viral proteins of influenza A virus. Once the new and advantageous serological type appears, it spreads in the population until fixation, the fixation speed depending on how effective the new driver mutation is. The effect on the phylogeny will be different substitution rates, but once different driver mutations are fixed in their respective populations, the pattern of mutant accumulation will be neutral. Taking into account this model for the evolution of viruses, the phylogenetic reconstruction, depending on the history of new driver mutations, the fitness of the corresponding genotypes as well as the relative duration of selective versus neutral periods, will be more or less well explained in terms of the molecular clock. The detection of such a selection pattern in the phylogenetic reconstruction would depend on the magnitude of the change in the intensity of selection, on how long it takes the new variant to become predominant in the population, and on the time elapsed from the appearance of the new variant to the isolation date of the sequences used for the phylogenetic reconstruction. Hence the usefulness of viral genes to put into evidence this kind of phenomena is apparent, as their evolutionary processes occur in very short intervals of time.

The molecular clock model has to be adopted because it shows the lowest AIC. Villaverde et al. (1991) also found molecular clock behavior in their analysis of the $C_1$ serotype of this same viral species. Assuming rate constancy, it is possible to date the splitting episodes that gave rise to new serotype and subtypes. These estimates show that serotypes OIsr and C are quite recent, both having appeared along this century. Using these estimated branching dates and the isolation dates given in Fig. 1, it is possible to compute the average rate per site per year of synonymous and nonsynonymous substitutions. The estimated values are 0.0353 ($\pm$0.0138) and 0.0085 ($\pm$0.0022) substitutions per site per year for synonymous and nonsynonymous changes, respectively. The ratio of synonymous to nonsynonymous rates is 4.153, similar to the average value of 5.0 detected by Li et al. (1985) for mammalian genes. It is also evident that this viral gene evolves much faster than eukaryotic ones, by a factor of $10^6$, and this observation does not deserve further comment. It is remarkable, however, that for both synonymous and nonsynonymous substitutions between C subtypes rates are higher than between them and the other, older serotypes as well as within these (7.118 and 4.153 times, respectively). Although not statistically significant, this is also observed in the three rates model (Table 2). Again, this can be taken as an indication of a recent episode of positive selection not yet completely hidden by ulterior neutral evolution.

*Acknowledgments.* We are indebted to Dr. Masami Hasegawa for his valuable comments and help in the realization of this work and for providing us with a copy of his program to obtain maximum-likelihood parameter estimates. We thank Dr. Esteban Domingo for his valuable comments on and corrections to the manuscript. S.F.E. has been supported by a fellowship from Conselleria de Cultura, Educació i Ciència de la Comunitat Valenciana. This work has been supported by grant BIO89-0668-C03-03 from the Comisión Interministerial de Ciencia y Tecnología (Spain) to A.M.

## References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Contr AC-19:716–723

Domingo E, Holland J (1988) High error rates, population equilibrium and evolution of RNA replication systems. In: Domingo E, Holland J, Ahlquist P (eds) RNA genetics, vol III. CRC Press, Boca Raton FL, p 3

Domingo E, Mateu MG, Martínez MA, Dopazo J, Moya A, Sobrino F (1990) Genetic variability and antigenic diversity of foot-and-mouth disease virus. In: Kurstak E, Marusyk RG, Murphy FA, van Regenmortel MVH (eds) Applied virology research, vol II. Plenum, New York, p 233

Dopazo J, Sobrino F, Palma E, Domingo E, Moya A (1988) Gene encoding capsid proteins VP1 of foot-and-mouth disease virus: a quasispecies model of molecular evolution. Proc Natl Acad Sci USA 85:6811–6815

Eigen M, Biebricher CK (1988) Sequence space and quasispecies distribution. In: Domingo E, Holland J, Ahlquist P (eds) RNA genetics, vol III. CRC Press, Boca Raton FL, p 211

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1984) Distance methods for inferring phylogenies: a justification. Evolution 38:16–24

Fitch W, Leiter JME, Li X, Palese P (1991) Positive Darwinian evolution in human influenza A virus. Proc Natl Acad Sci USA 88:4270–4274

Gillespie JC (1984) The molecular clock may be an episodic clock. Proc Natl Acad Sci USA 81:8009–8013

Gillespie JC (1986) Rates of molecular evolution. Annu Rev Ecol Syst 17:637–665

Gojobori T, Moriyama EN, Kimura M (1990) Molecular clock of viral evolution and the neutral theory. Proc Natl Acad Sci USA 87:10015–10018

Hasegawa M, Kishino H, Yano T (1985) Dating of the human–ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hasegawa M, Kishino H, Hayasaka K, Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in lemur. J Mol Evol 31:113–121

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Higgins DG, Sharp PM (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5(2):151–153

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Kishino H, Hasegawa M (1990) Converting distances to time:

an application to human evolution. Methods Enzimol 183: 550–570

Li WH, Wu CI, Luo CC (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2(2):150–174

Maruyama T, Birky CW (1991) Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. Genetics 127:449–451

Palmenberg A (1988) Sequence alignment of picornaviral capsid proteins. In: Semler BL, Ehrenfeld E (eds) Molecular aspects of picornavirus infection and detection. American Society for Microbiology, Washington, p 215

Villaverde A, Martínez MA, Sobrino F, Dopazo J, Moya A, Domingo E (1991) Fixation of mutations at the VP1 gene of foot-and-mouth disease virus. Can quasispecies define a transient molecular clock? Gene 103:147–153

Received January 18, 1992/Revised March 12, 1992