INTERNATL MICROBIOL (1998) 1:311–318

REVIEW ARTICLE

The foot-and-mouth disease RNA virus as a model in experimental phylogenetics

Mario A. Fares[1]
Eladio Barrio[1]
Nelsson Becerra[2]
Cristina Escarmís[3]
Esteban Domingo[3]
Andrés Moya[1]

[1] Department of Genetics, and Cavanilles Institute of Biodiversity and Evolutionary Biology, University of Valencia, Spain
[2] Department of Ecology and Evolutionary Biology, University of California, Irvine, USA
[3] Center of Molecular Biology Severo Ochoa, CSIC, Autonomous University of Madrid, Spain

Received 22 May 1998
Accepted 30 June 1998

Correspondence to:
Andrés Moya. Department of Genetics.
University of Valencia. Dr. Moliner, 50.
46100 Burjassot. Valencia. Spain.
Tel.: +34-963983180. Fax: +34-963983029.
E-mail: andres.moya@uv.es

Summary Phylogenetic reconstruction methods are subject to two types of limitations: our knowledge about the true history of organisms and the gross simplification implied in the numerical simulation models of the relationships between them. In such a situation, experimental phylogenetics provides a way to assess the accuracy of the phylogenetic reconstruction methods. Nonetheless, this capacity is only feasible for organisms in which replication and mutation rates are high enough to provide valuable data. On the other hand, experimental phylogenetics also provides insights on the main evolutionary processes acting on viral variability under different population dynamics. Our study with the foot-and-mouth disease virus (FMDV) strongly suggests that the phylogenetic reconstruction methods can infer erroneous phylogenies due to nucleotide convergences between isolates belonging to different experimental lineages. We also point out that the diverse evolutionary mechanisms acting in different experimental dynamics generate alterations and change the frequencies of genetic variants, which can lead to the misinterpretation of the real evolutionary history.

Key words Foot-and-mouth disease virus (FMDV) · RNA virus evolution · Distance-based methods · Molecular phylogenies · Convergent evolution

Introduction

The important increase of computational power to implement phylogenetic algorithms and the great deal of data specially provided by molecular investigations have promoted an extraordinary development of explicit methods for phylogenetic inference [27]. These methods have been applied to a broad range of studies, from population genetics to the evolution of macromolecules, and particularly, to the reconstruction of the evolutionary history of organisms [20]. However, these phylogenetic reconstruction methods are limited by the incomplete knowledge of the evolutionary history of organisms under study, and therefore, it is necessary to ask how reliable these methods are for the inference of real processes which have occurred during the history of life.

There are two ways to answer this question: by the assessment of the accuracy of phylogenetic reconstruction via direct observation of the evolutionary history of organisms (experimental phylogenetics) and by the computational simulation of phylogenies. Because evolutionary history usually cannot be observed directly, the assessment of phylogenetic methods has mainly relied on simulation of phylogenies. Numerical simulations assume a particular model of evolution and then generate characters according to the model and to a given phylogeny in order to test the different phylogenetic methods. However, the assumption of an evolutionary model includes gross simplifications of a complex biological process, which constitutes the major limitation of this approach [17].

The analysis of known phylogenies cannot be performed with complex DNA-based organisms that undergo little genetic differentiation. In contrast, the high number of generations per

year of viruses, the limited sequence space of the viral genomes, their extremely large population sizes and their higher mutational rates permit the manipulation of viral lineages in the laboratory through thousands of generations per year, thus making experimental phylogenetic studies feasible [16]. As such, experimental phylogenetics has been possible by either the reconstruction of known phylogenies of strains whose history has been recorded, or by the manipulation, under controlled experimental conditions, of viral lineages [3, 4, 16].

RNA viruses are characterized by an extremely high mutation rate, which in addition to the other characteristics common to DNA viruses (huge population numbers, short life cycles, and low complexity), make them very suitable models for the study of experimental evolution [25], including of course, experimental phylogenetics.

In the present study, our main goal was not only to test the accuracy of the phylogenetic inference methods to reconstruct the known experimental evolutionary history of a RNA virus, the foot-and-mouth disease virus (FMDV), but also to characterize the evolutionary processes acting on viral variability under different population dynamics.

Work performed

Cells, viruses and infections The initial FMDV clone (C-S8c1) was subjected to different experimental procedures that could be classified in massive and plaque-to-plaque transfers (Fig. 1). Massive passages were carried out in either persistent or cytolytic infections of BHK-21 cells, and in the

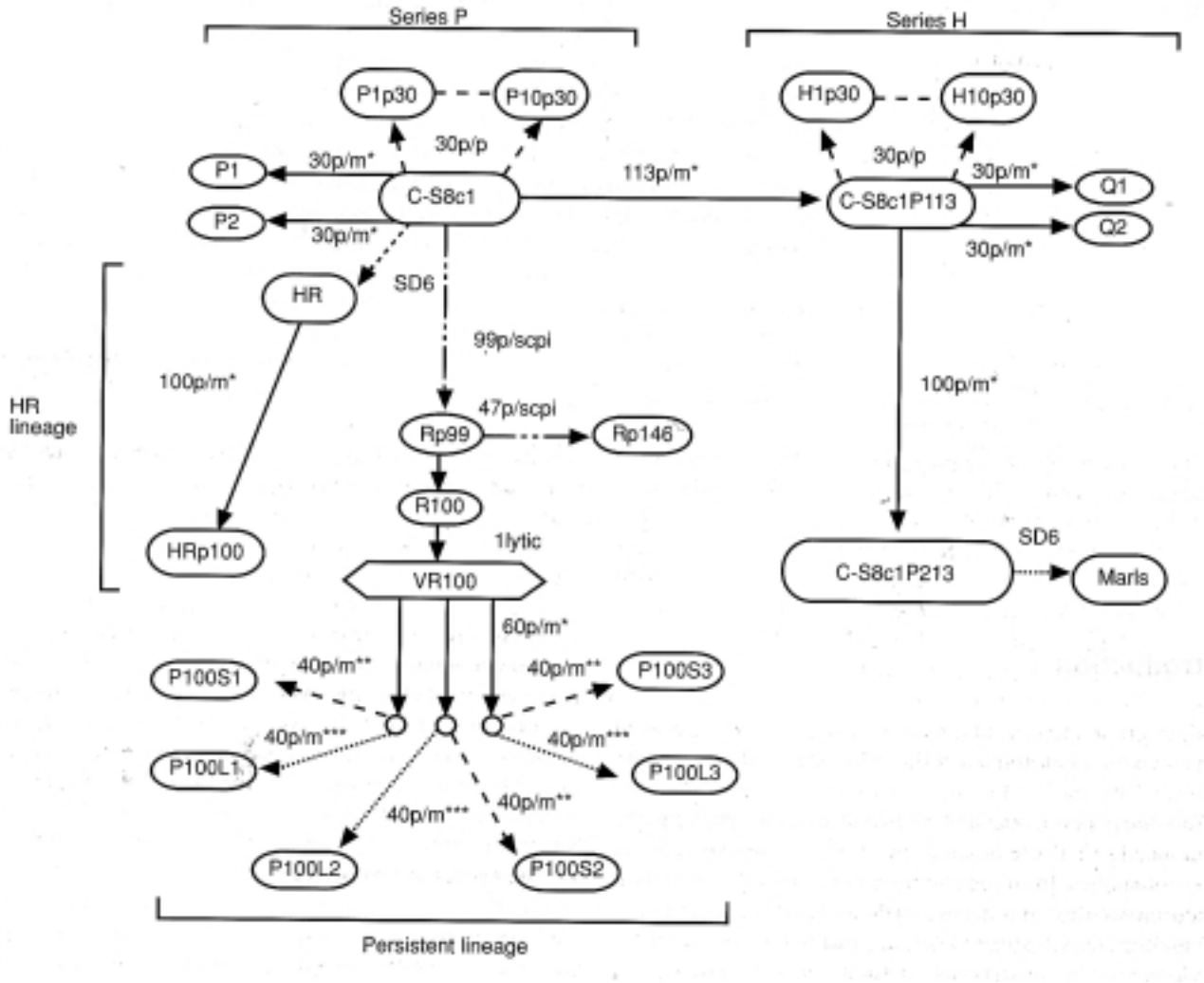

Fig. 1 Scheme of the experimental evolutionary history of the foot-and-mouth disease virus. p/p, p/m, p/scpi and 1p/lytic indicate plaque-to-plaque transfers, massive passages, passages of the supernatant of persistently infected cells, and 1 cytolytic infection, respectively. Massive passages were performed by the infection of 5×10^6 BHK-21 cells with a multiplicity of infection of 1 particle per cell (m*) or of 0.04 particles per cell (m**) or by infection of 2×10^8 cells with a multiplicity of infection of 1 particle per cell (m***). SD6 indicates the treatment with antibody SD6 against antigenic site A within protein VP1

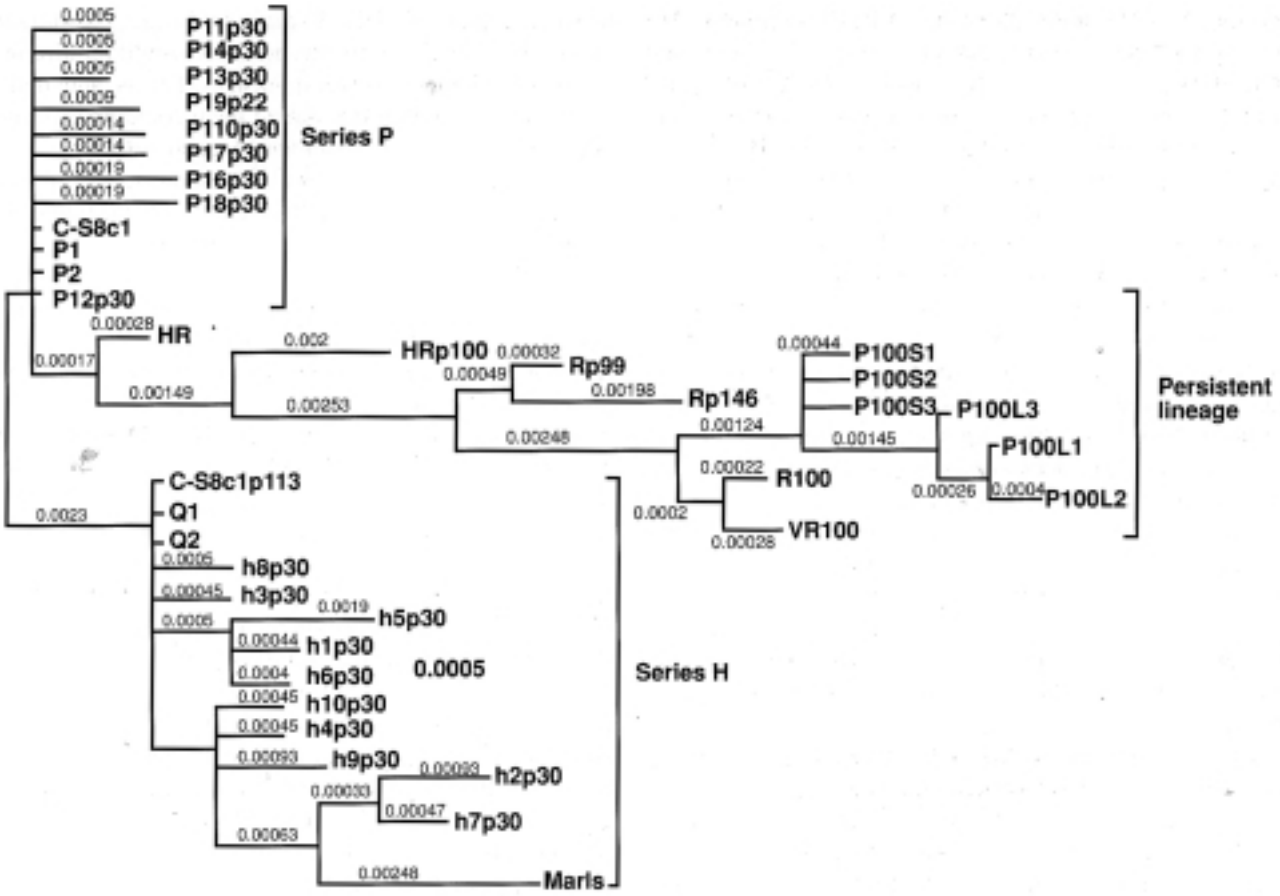

Fig. 2 Phylogenetic tree of the foot-and-mouth disease virus capside protein sequences inferred by the Fitch and Margoliash's method [13]. Numbers on the branches indicate the distances given in nucleotide substitutions per site. Distances were estimated according to the Jukes and Cantor's method [18]

presence or absence of monoclonal antibody SD6. The original clone C-S8c1 and its derivative C-S8c1p113 (obtained after 113 massive passages) were subjected to plaque-to-plaque transfers giving rise to the clones of the series P and H, respectively. Marls was derived from the C-S8c1p213 clone, isolated upon 100 passages of C-S8c1p113, and further passed in the presence of antibody SD6. P100 clones were isolated upon 60 cytolytic infections of VR100 followed by 40 massive passages with different infection multiplicity, giving place to isolate groups P100L and P100S. Similarly, HR and Rp99 were massively passed to obtain isolates HRp100 and Rp146, respectively.

cDNA synthesis, PCR amplifications and nucleotide sequencing Viral RNA extraction, reverse transcription for the synthesis of cDNA, PCR amplification with specific primers of the capside protein region and direct nucleotide sequencing of the isolates and clones to obtain consensus sequences of the viral populations were performed as described elsewhere [9].

Phylogenetic inference The 38 sequences of the region coding for the capside proteins VP1 to VP4 were aligned using the CLUSTAL X program [29]. For phylogenetic inference, we used three different methods based on optimality criteria: (i) Fitch and Margoliash's [13] distance-based method (FM), (ii) maximum-parsimony (MP) analysis [14], and (iii) maximum-likelihood (ML) procedures [10]. FM trees were obtained from Jukes and Cantor's [18] distances, using the computer programs DNADIST and FITCH from the PHYLIP package v 3.5 for Windows [12]. Alternative trees were obtained with the MacClade program v. 3 [21], and compared by the test for maximum parsimony proposed by Templeton [28] and developed by Felsenstein [11], and by the test of maximum likelihood originated by Kishino and Hasegawa [19]. These tests were performed with the programs DNAPARS and DNAML, respectively, also from the PHYLIP package.

Phylogenetic analysis

The FMDV genome consists of a positive single-strand RNA molecule of 8500 nucleotides, which encodes for a single polyprotein processed by viral proteases into several mature proteins (for more details see [1]). The region analyzed in the present study encompassed a 2142 nucleotide segment coding for the four capside proteins VP4 (partial sequence from

positions 1 to 213 of the alignment), VP2 (from position 214 to 857), VP3 (858 to 1519) and VP1 (1520 to 2142). A total of 31 of the 38 sequences analyzed were different and yielded 2142 nucleotide positions in the alignment because there were no insertions/deletions. In the alignment, 78 variable sites (3.6%) were observed, of which 34 were phylogenetically informative (1.6%). Of the variable sites, 22 corresponded to first codon positions (13 of them being informative), 18 to second codon positions (11 informative), and 38 to third codon positions (10 informative). Of the inferred amino acid positions, 35 were variable (out of 714), of which 21 were informative.

To perform phylogenetic analysis based on distances, the Jukes and Cantor's [18] method of distance estimation was applied to correct for superimposed nucleotide substitutions. These distances (data not shown) were used to obtain the Fitch-Margoliash (FM) tree depicted in Fig. 2. Equally maximum parsimonious (MP) and maximum likelihood (ML) trees, shown in Figs. 3 and 4, respectively, were derived from the sequence alignment. With the maximum-parsimony criterion, four trees requiring the same mutational steps were obtained, one of them presenting the same topology as the ML tree. These four trees only differed in the alternative positions of two ambiguous mutations.

The five topologies obtained with the three optimality criteria methods of phylogenetic inference (FM, MP and ML) corresponded to a similar phylogenetic reconstruction (Figs. 2, 3 and 4). In fact, two methods of testing whether alternative topologies are significantly better or worse, one based on maximum-likelihood [19] and the other based on maximum parsimony [12, 28], indicated that the five topologies were not significantly different (results not shown). However, the same tests suggested that the real tree was significantly worse than the five phylogenetic reconstructions for explaining the evolution of the sequences under study.

Although the reconstructed trees were not completely congruent with the "history" of the isolates generated under controlled experimental conditions, some of the discrepancies can be easily explained. Only one of 20 sequences from the series H clones, each one generated after 30 plaque to plaque transfers of the isolate C-S8c1p113, directly derived from the consensus sequence of this isolate (Figs. 2, 3 and 4). However, as the sequences under study corresponded to consensus sequences (obtained from direct sequencing of the RT-PCR amplification products) of the viral population presented in the different isolates, these phylogenetic results can be explained by the existence in the population of at least 4 or 5 viral genomes at high frequencies. Of these, the most frequent corresponded to the consensus sequence of the isolate C-S8c1p113, and the other three or four corresponded to the ancestral sequences (nodes) connecting the clones of the series H in the alternative reconstructed trees. These ancestral sequences were probably generated during massive passages and fixed during plaque-to-plaque transfers. Of course, this explanation is correct only if several parallel changes (shared between sequences) occur in different lineages leading to H clones, but the alternative hypothesis would require many convergent changes in several lineages. This is quite unlikely despite the possibility that one of them could have occurred according to the most parsimonious reconstruction.

This situation is in contrast to the sequences of the series P clones, also obtained after 30 plaque-to-plaque transfers but forming the original clone C-S8c1 (Fig. 1). In this case, none of the clones shared an ancestor different from the C-S8c1 consensus sequence, which was totally congruent with the star radiation of the clones in the experimental protocol. As such, this was indicative of a much lower variability in the viral population from the original clone C-S8c1 than in the viral population from isolate C-S8c1p113. Therefore, the increase of variability in the C-S8c1p113 population would have occurred during the 113 massive passages.

Polymorphism in the viral population can also explain the relationships among sequences from isolates Rp99, Rp146, R100, VR100, and the three series L-S (Figs. 1 to 4). Only the existence of several variants in the viral populations present in the isolates allows the interpretation of discrepancies between the real tree and the phylogenetic reconstructions, but in the present case this polymorphism was present in viral populations from persistently infected host cells. Thus, the fact that sequences P100S1, S2, S3 and P100L1, L2, L3 shared more than one common ancestor indicate the existence of polymorphism in the viral population R100 or VR100 (obtained after 100 passages in persistently infected cells and after an extra cytolytic infection, respectively) from which the three parallel series L-S were initiated. Also, at least three and two variants should be present in the viral population from isolates Rp99 and VR100, respectively, to explain the results of the phylogenetic analysis.

Analysis of nucleotide substitutions

It can be observed that the four most parsimonious trees differed in the position of two ambiguous nucleotide sites (Fig. 3). These two sites were ambiguous because, according to the trees, they experienced either two parallel changes or a substitution and its corresponding reversion during the evolution of the experimental viral populations. However, these were not the only convergences (we use the term "convergence" to include both parallel and reverse changes according to Doolittle [8]) which occurred during this evolutionary process. Thus, another striking incongruence between the real experimental tree and the phylogenetic inferences was the position of the consensus sequence HRp100, which, although derived from the antibody-treated isolate HR, appeared in the parsimonious trees clustered with the "persistence" lineage. This HRp100 sequence only shared one single nucleotide substitution with HR, causing the replacement of His to Arg in the antigenic region A of the VP1 capside protein selected by the antibody SD6. However,

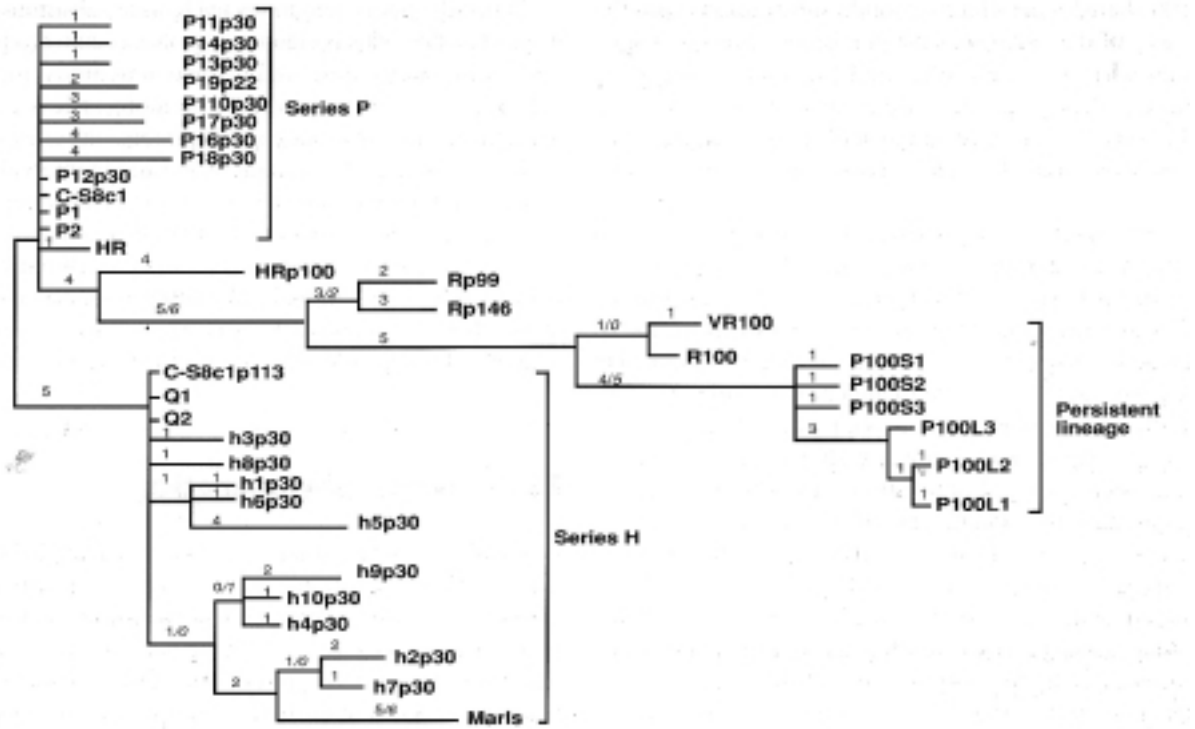

Fig. 3 Maximum parsimony tree of the foot-and-mouth disease virus capside protein sequences. Numbers on the branches correspond to absolute nucleotide substitutions between isolates. Numbers in italics indicate alternative substitutions due to ambiguous changes. The combination of these two ambiguous changes generates the four equally parsimonious trees

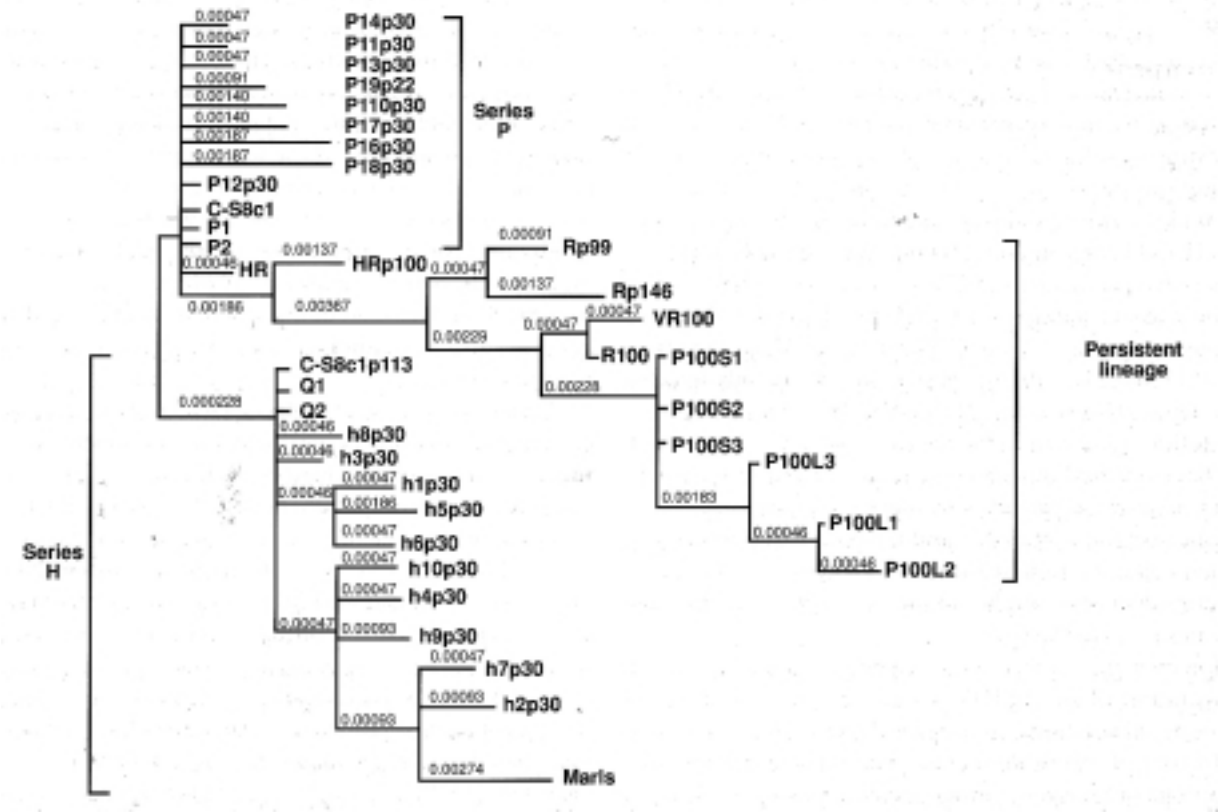

Fig. 4 Maximum likelihood tree of the foot-and-mouth disease virus capside protein sequences. Numbers on the branches are the maximum likelihood distance estimates given in nucleotide substitutions per site

HRp100 shared 4 parallel nucleotide substitutions with the sequences of the isolates of the persistence lineage, which explains why they were clustered together in the trees. Moreover, these changes were also nonsynonymous substitutions causing three amino acid replacements (two of them occurred in the first and second positions of the same codon).

The only alternative hypothesis to the convergent evolution would be the pre-existence of the postulated common ancestor of HRp100-persistence lineage together with C-S8c1 in the initial viral population. Then, when an aliquot of the initial population was taken, two rare variants having the replacement His-Arg were selected by the treatment with antibody SD6, one of them proceeding from the viral population of isolate C-S8c1 (corresponding to sequence HR) and the other from the postulated ancestral population "HRp100-persistence" (corresponding to sequence Hrp100). The first was more frequent just after SD6 treatment and the second became more frequent after massive passages. On the other hand, the postulated ancestor also increased its frequency in the persistence lineage. However, this hypothesis requires the presence of a rich polymorphism in the initial C-S8c1 viral population (4 replacement substitutions between both variants), which as already mentioned, was not observed, according to the comparison of the polymorphisms in series P and H (see above). Moreover, other 11 convergences (parallel or reverse changes) are also required to explain the evolution of the persistence lineage, and this abundance of convergences supports the parallel substitution hypothesis for explaining the position of HRp100 in the inferred trees.

A detailed analysis of the nucleotide substitutions which occurred during the experimental evolution of FMDV yielded interesting results. The average numbers of nucleotide substitutions per passage were 0.05 and 0.07 (both ranging from 0 to 0.13) in the thirty plaque-to-plaque transfers of series H and P, respectively; 0.06 (from 0.04 to 0.07) in the 113 massive passages from C-S8c1 to C-S8c1p113; 0.08 in the 100 massive passages from HR to HRp100; 0.06 in the 100 massive passages from C-S8c1p113 to Marls; and 0.20 (from 0.15 to 0.24) in the 100 passages of persistently infected cells from C-S8c1 to R100-VR100 (including the substitutions observed in the ancestors of series L-S, which must have occured during persistence, as deduced above). That is, assuming equal absolute times for a massive passage, a plaque-to-plaque transfer and a passage in persistently infected cells, the rate of substitutions (per absolute time) was more than twice higher in the persistence lineage than in the massive passages.

Moreover, the number of convergences required to explain the evolution of the FMDV sequences was 0 (out of 19 nucleotide substitutions) in series P, 2 (out of 29) in series H, and 14 (out of 39) in the persistence lineage. Finally, the proportions of nonsynonymous substitutions were 36.8% in series P, 51.7% in series H, and 61.5% in the persistence lineage.

When the nonsynonymous nucleotide substitutions were mapped on the codon positions where they occurred, a particular distribution was observed. Nonsynonymous nucleotide substitutions were observed to be more or less randomly distributed along the four capside proteins in series P and H, with no more than 2–3 mutations located in intervals of less than 10 codon positions. However, the persistence lineage mutations mainly concentrated within two narrow regions, 9 (out of 24) nonsynonymous substitutions occurred within an interval of 8 codons located at the amino terminal end of VP3, and 8 other substitutions took place in a 9-codon length interval located within the antigenic site A of capside protein VP1.

Experimental phylogenetics

Our study has demonstrated that experimental phylogenetics is a feasible way to test not only the accuracy of phylogenetic methods but also to simulate and understand evolutionary processes that organisms undergo under different population dynamics. Thus, in the present study, we have demonstrated how the presence of different variants in a viral population subjected to different population dynamics, in which diverse evolutionary mechanism may be acting, can explain changes in their frequencies.

Our results suggest that FMDV (C-S8c1) obtains genetic variation during massive passages and generates a large amount of mutants that become more frequent in the viral population. This observation could be supported by the concept of metapopulation dynamics which maintains that a local infection leads to spatially structured populations that are aggregates of smaller subpopulations whose local population dynamics are influenced by interpopulation migration [15]. Thus, during massive passages, two types of viral populations can be distinguished: one composed of viral particles with higher competition capability to obtain resources, which become dominant, and other subpopulations with different genetic variants maintained in low proportions within the metapopulation.

Underlying the use of molecular data is the assumption that convergent evolution is statistically unimportant, that is, that identical independent changes in different lineages are not common enough to obscure the true historical signal [8, 26]. However, our study shows the relevance of convergence during the evolution of FMDV. Convergent evolution has been previously suggested for FMDV and in other picornaviruses [2, 23, 24]. Independent lineages could be associated with a restrictive tolerance for accepting nucleotide substitutions as a result of functional and structural constraints [7, 30]. The analysis of phenotypic characteristics of FMDV demonstrated that phenotypic traits, acquired by R100 in the course of its persistence in BHK-21 cells, reverted completely or partially during the cytolytic infection to VR100 [5]. Several studies

also demonstrated that during the cytolytic passage from R100 to VR100, the new viral population become hypervirulent and that this hypervirulence was maintained and further accentuated, possibly due to a rapid completion of the infectious cycle and an increased rate of cell death [6, 22].

As observed, the vast majority of the convergent evolution occurred in the persistence lineage during both the persistence and the subsequent cytolytic infections. Convergent evolution may indicate the beneficial effects of some substitutions, although mutation pressure and functional constraints alone are unlikely to cause a meaningful increase in substitution frequency, and consequently, the plausible alternative would be selection. The accumulation of substitutions in a small number of codons also supports the role of selection as an explaination of the evolutionary process of this lineage. Further analyses are needed to confirm or discard positive selection acting on FMDV.

As a final conclusion, experimental phylogenetics may provide valuable insights into the relationships between different virus strains and into the main evolutionary processes acting on them. These kinds of studies could have enormous importance from an epidemiological standpoint for determining the origin of epidemic events and the mechanisms acting on viral variability. Our study suggests that the real evolutionary history could be misinterpreted as a result of the action of evolutionary forces generating and modulating genetic variation. This variance could be accentuated during interhost transmission since an intrahost environment is less heterogeneous, temporally and spatially, than the interhost one. Therefore, in virusal infections, the high degree of mutation and generation of further genetic variants should be taken into account since this high mutation rates could generate convergence of genotypic or phenotypic traits and, consequently, obscure the relationships between different strains.

Acknowledgments This work was supported by grants PM97-0060-C02-01 to ED and PM97-0060-C02-02 to AM. Computer analyses were carried out using the "Servei de Bioinformàtica" (University of Valencia). MAF was recipient of a fellowship from the Generalitat of Valencia. NB was supported by a fellowship from the Program "Minority International Research Training" of the University of California, Irvine, USA.

References

1. Belsham GJ (1993) Distinctive feature of foot-and-mouth disease virus, a member of picornavirus family; aspects of virus protein synthesis, protein processing and structure. Prog Biophys Mol Biol 60:241–260
2. Borrego B, Novella I, Giralt E, Andreu D, Domingo E (1993) Distinct repertoire of antigenic variants of foot-and-mouth disease virus in the presence or absence of immune selection. J Virol 67:6071–6079
3. Bull JJ, Cunningham CW, Molineux IJ, Badgett MR, Hillis DM (1993) Experimental molecular evolution of bacteriophage T7. Evolution 47:993–1007
4. Bull JJ, Badget MR, Wichman HA, Huelsenbeck JP, Hillis DM, Gulati A, Ho C, Molineux IJ (1997) Exceptional convergent evolution in a virus. Genetics 147:1497–1507
5. de la Torre JC, Dávila M, Sobrino F, Ortín J, Domingo E (1985) Establishment of cell lines persistently infected with foot-and-mouth disease virus. Virology 145:24–35
6. de la Torre JC, Martínez-Salas E, Díez J, Villaverde A, Gebauer F, Rocha E, Dávila M, Domingo E (1988) Coevolution of cells and viruses in a persistent infection of foot-and-mouth disease virus in cell culture. J Virol 62:2050–2058
7. Domingo E, Escarmís C, Sevilla N, Moya A, Elena SF, Querol J, Novella IS, Holland J (1996) Basic concepts in RNA virus evolution. FASEB J 10:1–6
8. Doolittle RF (1994) Convergent evolution: the need to be explicit. Trends Biochem Sci 19:15–18
9. Escarmís C, Dávila M, Charpentier N, Bracho A, Moya A, Domingo E (1996) Genetic lesions associated with Muller's ratchet in an RNA virus. J Mol Biol 264:255–267
10. Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376
11. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
12. Felsenstein J (1993) PHYLIP (Phylogenetic Inference Package) version 3.5c. Department of Genetics, University of Washington, Seattle
13. Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. A method based on mutation distances as estimated from cytochrome c sequences is of general applicability. Science 155:279–284
14. Fitch WM (1971) Towards defining the course of evolution: minimum change for a specific tree topology. Syst Zool 20:406–416
15. Hanski IA, Simberloff D (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. In: Hanski IA, Gilpin ME (eds) Metapopulation Biology: Ecology, Genetics and Evolution. San Diego: Academic Press, pp 5–26
16. Hillis DM, Bull JJ, White ME, Badgett MR, Molineux IJ (1992) Experimental phylogenetics: generation of a known phylogeny. Science 255:589–592
17. Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenetics. Science 264:671–677
18. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian Protein Metabolism. New York: Academic Press, pp 121–123
19. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topology from DNA sequence data, and the branching order in hominoid. J Mol Evol 29:170–179
20. Li WH (1997) Molecular Evolution. Sunderland, MA: Sinauer
21. Maddison WP, Maddison DR (1992) MacClade: Analysis of Phylogeny and Character Evolution. Sunderland, MA: Sinauer
22. Martín-Hernández AM, Carrillo EC, Sevilla N, Domingo E (1994) Rapid cell variation can determine the establishment of a persistent viral infection. Proc Natl Acad Sci USA 91:3705–3755
23. Mason MW, Rieder E, Baxt B (1994) RGD sequence of foot-and-mouth disease virus is essential for infecting cells via the natural receptor but can be bypassed by antibody-dependent enhancement pathway. Proc Natl Acad Sci USA 91:1932–1936
24. Mateu MG, Hernández J, Martínez MA, Feigelstoock D, Lea S, Pérez JJ, Giralt E, Stuart D, Palma EL, Domingo E (1994) Antigenic heterogeneity of a foot-and-mouth disease virus serotype in the field is mediated by very limited sequence variation at several antigenic sites. J Virol 68:1407–1417
25. Moya A (1996) RNA viruses as model systems for testing the population genetics view of evolution. In: Tojo J (ed) Proceedings of the First International Conference on Philosophy of Science. Vigo: Servicio de Publicaciones de la Universidad de Vigo, pp 207–218
26. Sanderson MJ, Donoghue MJ (1996) The relationship between homoplasy and confidence in a phylogenetic tree. In: Sanderson MJ, Hufford L (eds) Homoplasy: The Recurrence of Similarity in Evolution. New York: Academic Press, pp 67–89

27. Swofford DL, Olsen GJ, Waddel PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK (eds) Molecular Systematics (2nd edn). Sunderland, MA: Sinauer, pp 407–514
28. Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. Evolution 37:221–244
29. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680
30. Wimmer E, Helen CVT, Cao X (1993) Genetics of poliovirus. Annu Rev Genet 27:353–436