Evidence for Positive Selection in Foot-and-Mouth Disease Virus Capsid Genes From Field Isolates

Daniel T. Haydon,* Armanda D. Bastos,† Nick J. Knowles‡ and Alan R. Samuel‡

**Centre for Tropical Veterinary Medicine, University of Edinburgh, Roslin, Midlothian, EH25 9RG Scotland, †Exotic Diseases Division, Onderstepoort Veterinary Institute, Onderstepoort 0110, South Africa and ‡Institute for Animal Health, Pirbright Laboratory, Woking GU24 0NF, England*

Manuscript received June 5, 2000
Accepted for publication September 13, 2000

ABSTRACT

The nature of selection on capsid genes of foot-and-mouth disease virus (FMDV) was characterized by examining the ratio of nonsynonymous to synonymous substitutions in 11 data sets of sequences obtained from six different serotypes of FMDV. Using a method of analysis that assigns each codon position to one of a number of estimated values of nonsynonymous to synonymous ratio, significant evidence of positive selection was identified in 5 data sets, operating at 1–7% of codon positions. Evidence of positive selection was identified in complete capsid sequences of serotypes A and C and in VP1 sequences of serotypes SAT 1 and 2. Sequences of serotype SAT-2 recovered from a persistently infected African buffalo also revealed evidence for positive selection. Locations of codons under positive selection coincide closely with those of antigenic sites previously identified with the use of monoclonal antibody escape mutants. The vast majority of codons are under mild to strong purifying selection. However, these results suggest that arising antigenic variants benefit from a selective advantage in their interaction with the immune system, either during the course of an infection or in transmission to individuals with previous exposure to antigen. Analysis of amino acid usage at sites under positive selection indicates that this selective advantage can be conferred by amino acid substitutions that share physicochemically similar properties.

FOOT-AND-MOUTH disease virus (FMDV) is one of two species in the genus Aphthovirus in the family *Picornaviridae* and is the causative agent of a highly infectious and economically devastating disease of cloven-hoofed livestock. FMDV is a positive-stranded RNA virus of ~8.4 kb that occurs as seven distinct serotypes (A, C, O, Asia 1, and SAT 1–3), representatives of which are widely distributed throughout South America, Africa, the Middle East, and Asia. FMDV displays considerable antigenic variability in the field (Mateu *et al.* 1988; Martinez *et al.* 1992; Curry *et al.* 1996), which necessitates a continual need to update vaccine strains (Domingo and Holland 1992; Feigelstock *et al.* 1996). Animals inoculated with inactivated vaccine are protected for up to 6 months against viral strains that are closely related to vaccine strains, but protection is of shorter duration against more antigenically distant viruses (Woolhouse *et al.* 1996). Infection is cleared, usually within 2–3 weeks, by antibody opsonization and macrophage clearance (McCullough *et al.* 1992; Brown 1995).

FMDV capsids are composed of 60 copies of each of four proteins, VP1–4, assembled into a nonenveloped icosahedral structure ~300 Å in diameter (Acharya *et al.* 1989). Only VP1–3 are exposed on the capsid surface. Proteins VP1–3 are structurally similar, taking the form of eight-stranded antiparallel β-barrels. The eight β-strands are conventionally labeled B–I, and loops that join adjacent β-strands are referred to by the two strands they bridge. X-ray crystallography has shown the FMDV particle structure (Acharya *et al.* 1989; Lea *et al.* 1994) to be similar to those determined for other picornaviruses.

Monoclonal antibodies (mAbs) have been used to study antigenic sites of many serotypes of FMDV involved in virus neutralization. Epitopes map principally to loops linking β-strands and that project outward from the virus surface. Independent antigenic sites were identified, involving the three surface-exposed capsid proteins of 0_1Kaufbeuren strain of FMDV (Kitson *et al.* 1990; Crowther *et al.* 1993a). Site 1, located in the center of the protomer, includes the G–H loop (residues 574–604; for numbering, see Figure 1) and C-terminal region (residue 657) of VP1. Site 2, at the threefold axis, involves the B–C (residues 70–77) and E–F loops (residue 130–137) of VP2. Site 3, involving the B–C loop (residues 482–487) of VP1, is located near the fivefold axis of the virus particle. Site 4, also close to the threefold axis, consists of residues 366–367 of VP3.

In studies of FMDV serotype A, epitopes were also identified on the VP1 G–H loop, together with an additional site within the VP1 H–I loop (residue 618), adjacent to the B–C loop of VP1 (Thomas *et al.* 1988; Baxt

Corresponding author: Daniel T. Haydon, Ctr. for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, Scotland. E-mail: daniel.haydon@ed.ac.uk

et al. 1989; Lea *et al.* 1994). Mapping of mAb escape mutants of serotypes SAT 2 (Crowther *et al.* 1993b) and Asia 1 (Butchaiah and Morgan 1997; Marquardt *et al.* 2000) has also characterized epitopes located within the VP1 G–H loop, and in the case of Asia 1, in the B–C loop of VP2, the B–B knob of VP3, and N terminus of VP2. The antigenicity of type C viruses appears broadly similar to that of type O (Mateu *et al.* 1988, 1990; Lea *et al.* 1994). Mapping antigenic sites of serotypes SAT 1 and SAT 3 has not yet been accomplished with similar precision.

Significant antigenic diversification has been shown to arise even as a result of single point mutations (Mateu *et al* 1990; Martinez *et al.* 1991) and thus, in light of the high mutation rates characteristic of RNA viruses, it is likely that even brief FMDV epidemics might result in generation of substantial antigenic variability. However, the adaptive significance of this variation remains unclear (Haydon *et al.* 1998). Antigenic variation might be of adaptive value for two reasons: first, antigenic variation generated over the course of a viremia might act to prolong or intensify a single infection, thereby resulting in greater transmission potential from infected animals; second, sufficiently distinct strains might be capable of more rapid reinfection of hosts with some previous experience of related antigen, thereby effectively increasing the susceptible host population size.

Due to the immunogenic nature of virus capsids it is inevitable that point mutations to genes encoding the capsid will result in antigenic variation, and this has been repeatedly demonstrated (even in the absence of antibody; Diez *et al.* 1989), but whether or not such variation is subject to positive selection remains unknown. A common test for positive selection is to test whether the ratio of nonsynonymous substitutions per nonsynonymous site (d_n) to synonymous substitutions per synonymous site (d_s) observed in pairwise comparisons of aligned sequences significantly exceeds unity (Sharp 1997). Average d_n/d_s ratios in sliding windows of specified length can be calculated for all possible pairs of sequences of a data set. The problem with this indicator is that purifying selection will inevitably be encountered at some codon positions, while if positive selection acts at all, it is likely to do so at only a few sites. Thus average estimates of d_n/d_s ratios over extended gene fragments may reflect a mixture of positive and purifying selection forces within the window, and yet still fall beneath the unit threshold. A further weakness of these tests is that difficulties are encountered in accounting for unobserved multiple synonymous changes, which might lead to overestimation of d_n/d_s ratios. However, such methods have identified positive selection in human immunodeficiency virus (HIV) *env* genes (Seibert *et al.* 1995), the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum* (Hughes 1992), and the antigen-binding cleft of class 1 major-histocompatibility-complex molecules (Hughes *et al.* 1990).

Weaknesses of pairwise comparison methods are partly overcome by examining d_n/d_s ratios over a phylogeny and assigning values of d_n/d_s to individual codons. Various methods have recently been developed to do this (Fitch *et al.* 1991; Nielsen and Yang 1998; Bush *et al.* 1999a,b; Suzuki and Gojobori 1999; Yang *et al.* 2000). Here we apply the methods of Yang *et al.* (2000) to 11 different sets of FMDV capsid genes and find statistically significant evidence for positive selection acting on between 1 and 7% of codons in 5 of these data sets.

MATERIALS AND METHODS

The methods of Nielsen and Yang (1998) and Yang *et al.* (2000) use a likelihood-based approach to identify selection, which has a number of advantages over pairwise methods: (1) Sequence comparisons are made on a phylogenetic rather than pairwise basis, and a codon substitution matrix (with fitted transition:transversion ratio) is specified to account for changes at sites along branches; (2) it has a formal statistical foundation, using maximum likelihood to fit various (nested) models in which each codon position in the sequence is assigned according to an empirical Bayesian analysis to categories that evolve according to a different d_n/d_s ratio. Statistical significance of model support is established using likelihood-ratio tests. Yang *et al.* (2000) propose a model (denoted M7 in Yang *et al.* 2000) with 10 categories of d_n/d_s ratio, the values of which are determined by a fitted β-distribution constrained to span the interval 0–1. Individual codons are assigned to categories and posterior probabilities are calculated to indicate the confidence with which each assignment is made. The likelihood of this "null" model is then compared with that of a similar model that possesses one additional category whose d_n/d_s ratio may be any value (M8 in Yang *et al.* 2000). Because the null model is nested within this more complex model the respective likelihoods may be compared using a likelihood-ratio test with 2 d.f.—representing the difference in the number of parameters by which the two models differ. If the more complex model has significantly higher likelihood and the additional category is chosen to have a d_n/d_s ratio >1, then positive selection is inferred.

Additional modifications described in Yang *et al.* (2000) permit models with k categories each with different d_n/d_s ratios to be fitted to sequence data and a supplied phylogenetic tree. The models, denoted M3(k), assume that each codon evolves according to one of the k different d_n/d_s ratios, values of which are once again obtained through maximum-likelihood fitting of the model to data. Since models with fewer categories of d_n/d_s ratios are nested within those with more, their likelihoods can again be compared with a likelihood-ratio test. Successive models were fitted to the data, starting with $k = 1$, and successively increasing k by 1. After the first occasion on which the likelihood of the model with k categories was not significantly improved by fitting $k + 1$ categories, the model with k categories was adopted as the model best describing the data [see Nielsen and Yang (1998) and Yang *et al.* (2000) for a complete description of these models, their assumptions, and statistical methods for model selection].

For each data set the phylogenetic trees required for these analyses were constructed using distance methods implemented in the Fitch program of the PHYLIP package (Felsentein 1993).

Physicochemical properties of amino acids at individual codon positions within a data set that were identified as being under positive selection were examined using randomization routines. We asked if average absolute differences in a physical chemical property between c amino acids found at particular sites under positive selection were significantly different from those expected between c randomly selected amino acids. To establish the expected distribution of differences, we examined where the observed average difference occurred in a ranking of average differences of 2000 randomly selected sets of c amino acids. Physical chemical properties were polar requirement, hydrophilicity, molecular weight (Zamyatnin 1972), and volume (Creighton 1983).

We applied these methods to 11 different sets of FMDV sequences.

D1: 28 complete capsid sequences of serotype O;
D2: 12 complete capsid sequences of serotype A;
D3: 10 complete capsid sequences of serotype C;
D4: 43 VP1 sequences of serotype O (a subset of those reported in Haydon *et al.* 1998);
D5: 17 VP1 sequences of serotype C (13 from the Philippines and 4 closely related type C's from South America);
D6: 10 partial VP1 sequences (amino acids 86–221) of serotype SAT-1 isolated from South Africa between 1990–1998;
D7: 32 partial VP1 sequences (amino acids 85–216) of serotype SAT-2 isolated from impala and African buffalo in the Kruger National Park (KNP), South Africa, between 1988–1996 (Bastos *et al.* 2000);
D8: 10 partial VP1 sequences (amino acids 85–216) of serotype SAT-2 isolated from a single persistently infected African buffalo over 12 months (Vosloo *et al.* 1996);
D9: 12 partial VP1 sequences (amino acids 88–217) of serotype SAT-3 isolated from buffalo in the KNP, South Africa, between 1990–1997;
D10: 46 partial VP1 sequences (amino acids 157–202) of serotype O from isolates from Saudi Arabia between 1983–1995 (Samuel *et al.* 1997);
D11: 14 partial VP1 sequences (amino acids 157–202) of serotype O from isolates from North Africa isolated between 1989–1991 (Samuel *et al.* 1999).

RESULTS

Evidence for positive selection was found in 5 of 11 data sets (D2, type A capsids; D3, type C capsids; D6, SAT-1 VP1 genes; D7, SAT-2 VP1 genes from impala and buffalo; and D8, SAT-2 VP1 genes from a persistently infected buffalo). Nonsignificant indications of positive selection were found in two further data sets (D5, type C VP1 genes from the Philippines; and D9, SAT-3 VP1 genes). No indication of positive selection was found in data from Saudi Arabia (D10), North Africa (D11), or type O capsids (D1) or VP1 sequences (D4)—see Table 1 for estimated parameter values. Where positive selection was indicated, d_n/d_s ratios were generally <3, the one exception being the persistent SAT-2 virus isolates where a d_n/d_s ratio of 25 was identified.

Sites under positive selection might be identified according to variously stringent criteria. The least stringent is all sites identified as belonging to categories with d_n/d_s ratios >1 by models that have the greatest likelihood (*i.e.*, most parameters). Using this criterion, of 662 codon positions examined, some evidence for positive selection was identified at 51 different positions in nine different data sets (18 of these sites were coidentified in more than one data set). Of the 51 identified

TABLE 1

Parameters and likelihood values corresponding to results from models M7 and M8

Data set	No. codons	Serotype/gene	n[a]	$\bar{d}$[b]	M7: LL[c]	M8 p_ω[d]	M8 ω[e]	M8 LL	M8 Significance[f]
D1	736	O capsids	28	0.159	−16202.6	0.021	1.00	−16180.2	≤0.0001
D2	651	A capsids	12	0.168	−8759.3	0.025	1.34	−8755.0	0.0136
D3	732	C capsids	10	0.102	−6619.3	0.0125	2.99	−6610.7	0.0002
D4	210	O VP1	43	0.138	−5584.9	Same model		−5584.9	—
D5	211	C VP1	17	0.108	−2074.0	0.100	1.64	−2072.6	0.2466
D6	143	SAT-1 VP1	10	0.370	−1581.4	0.044	3.64	−1570.6	≤0.0001
D7	138	SAT-2 VP1	32	0.176	−2638.2	0.070	1.38	−2631.3	0.0010
D8	144	SAT-2 VP1	10	0.012	−711.5	0.029	24.83	−705.5	0.0025
D9	139	SAT-3 VP1	12	0.124	−1364.5	0.019	2.69	−1362.3	0.1108
D10	55	O VP1	46	0.055	−1015.2	0.45	0.13	−1015.02	1.0
D11	55	O VP1	14	0.026	−379.2	Same model		−379.2	—

Underlined numbers indicate models associated with significantly increased level of support compared to the null model.

[a] Number of sequences in the data set.
[b] Average pairwise difference assuming Kimura two-parameter model.
[c] The log likelihood of the fitted model.
[d] The proportion of codons assigned to the additional category of M8.
[e] The value of d_n/d_s assigned to the additional category in M8.
[f] P values corresponding to the likelihood-ratio tests comparing model M7 with M8, assuming twice the log-likelihood difference to be chi-square distributed with 2 d.f.

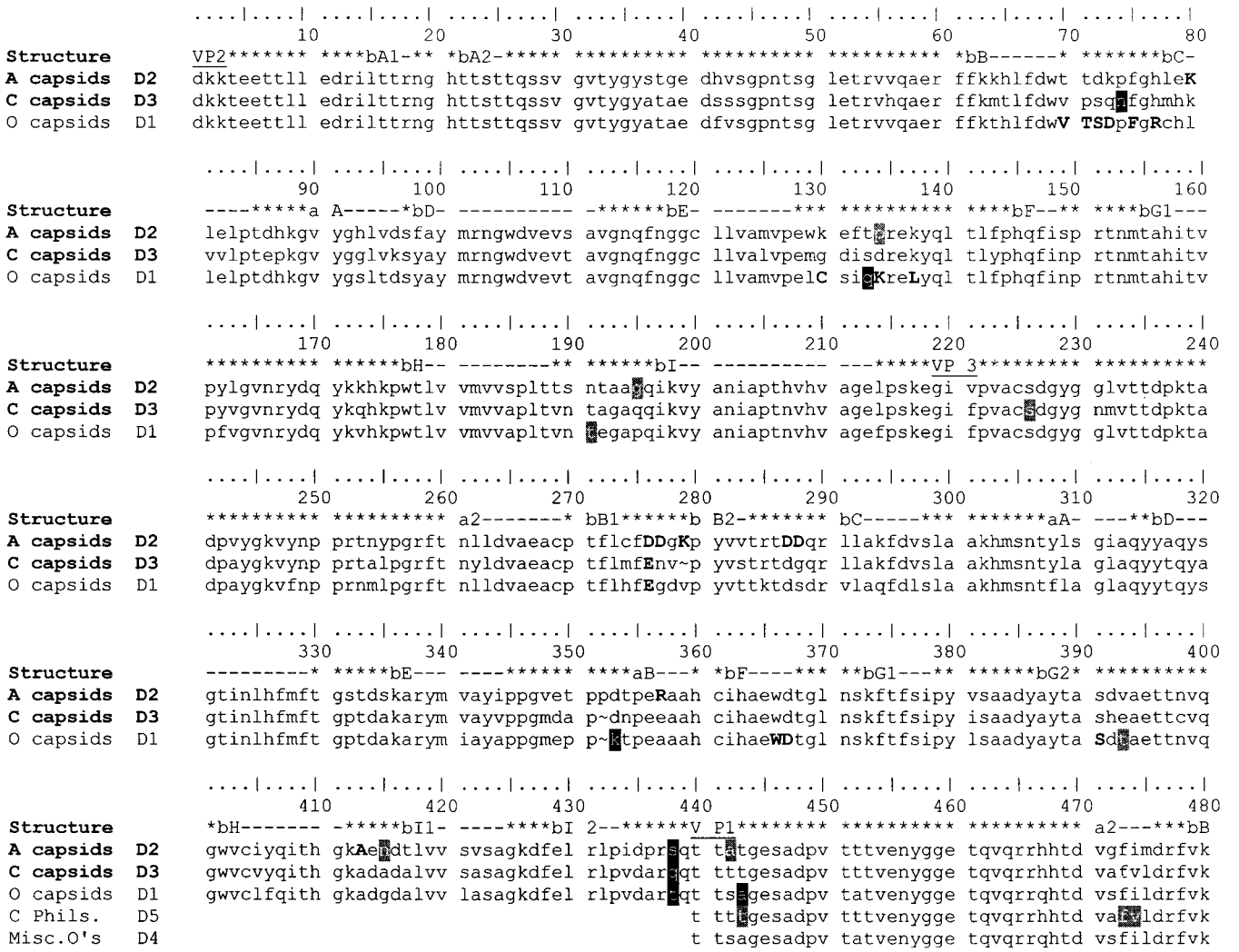

FIGURE 1.—Alignment of amino acid consensus sequences of the 11 data sets (obtained by eye). The structural guide is from serotype O BFS 1860/UK/67 (accession no. J02185) and will apply only approximately to other serotypes (b-, β-strands; a-, α-helices; and *, other secondary structures). Row titles in boldface indicate that significant evidence for positive selection was identified (see Table 1). Amino acids in uppercase boldface type are those previously associated with monoclonal antibody escape mutants. Amino acids shown in black and gray boxes were those placed in categories with d_n/d_s ratios >1. Black box indicates that the site was assigned to a positively selected category with posterior probability >0.95, gray box indicates that the site was assigned to a positively selected category with posterior probability >0.5 but <0.95. Black and gray boxed amino acids in rows with titles that are not boldface are residues indicated to be under positive selection by models with largest likelihoods (although not significantly larger than an alternative simpler model). Results for D1 [M8] and D4 [M3(3)] are included on this figure as values of d_n/d_s between 1.00 and 1.005 were identified. Note that site positions indicated by this alignment will not necessarily coincide with other published numbering schemes for these genes. Monoclonal antibody escape mutant data are from XIE *et al.* (1987); PFAFF *et al.* (1988); STAVE *et al.* (1988); THOMAS *et al.* (1988); BAXT *et al.* (1989); PARRY *et al.* (1989); KITSON *et al.* (1990); FEIGELSTOCK *et al.*(1992); MARTINEZ *et al.* (1992); CROWTHER *et al.* (1993a,b); and DAVIDSON *et al.* (1995).

sites, 5 were in β-strands or α-helices, 41 sites were located in the VP1 gene (16 in the G–H loop, 4 in the B–C loop, 3 in the E–F loop, and 4 in the H–I loop); 5 were in VP2 (1 in the B–C loop and 2 in the E–F loop), and 5 were in VP3 (see Figure 1). Half of the identified sites were located in previously recognized antigenic sites—over seven times a naive random expectation.

The most stringent identification criterion is those sites identified as belonging to categories with d_n/d_s ratios >1 with posterior probabilities exceeding some threshold (*e.g.*, 0.95), by models with likelihoods significantly greater (at 0.05 level) than the null model. Using this criterion, evidence for positive selection was identified at 17 different positions in five data sets (two of which were coidentified in more than one data set). Of these 17 sites, only 2 were in β-strands; 15 sites were located on the VP1 gene; and 1 was in each of VP2 and VP3 (see Figure 1). Two-thirds of these sites were located in previously recognized antigenic sites—almost 10 times a naive random expectation.

The M3(k) models indicated that no data sets justified establishment of more than three different d_n/d_s categories (four categories were fitted to all data sets and none

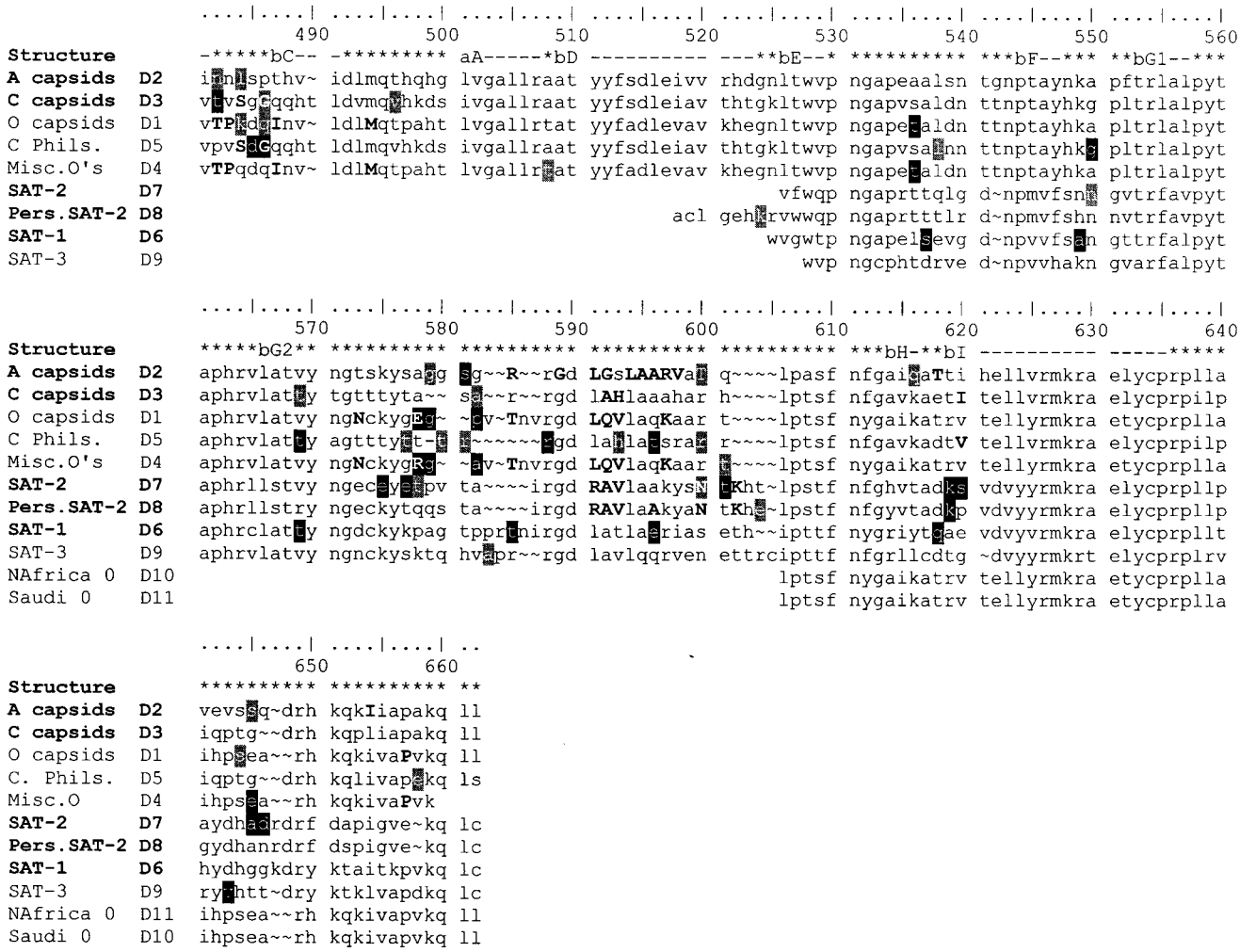

FIGURE 1.—*Continued.*

provided a significant improvement over three). Only the shortest sequences from North Africa and Saudi Arabia were adequately described by a single category, and the remainder required either two or three categories (Table 2). Taken together these analyses suggest that while the vast majority of all codons in all data sets are subject to mild or strong purifying selection, positive selection may act on between 1 and 7% of codons.

Examination of 4 physicochemical properties of the different amino acids observed at each of 17 sites involved $17 \times 4 = 68$ tests so $\sim$4 results significant at the 0.05 level were expected by chance alone. In fact, 23 tests proved to be significant at the 0.05 level, all of which suggested physicochemical properties were being conserved at particular sites (see Table 3 for details). Of the 4 physicochemical properties hydrophilicity was conserved at 12 of 17 sites.

DISCUSSION

This analysis reveals statistically significant evidence that positive selection can act on between 1 and 7% of codon positions in FMDV capsid genes. The concentration of positively selected sites in loop regions of VP1 is consistent with findings of studies that have used monoclonal antibodies to map epitopes (see Figure 1) and serum immunoglobulin fractionation (MATEU *et al.* 1995). Loop regions in VP2 and VP3 genes contain far fewer positively selected sites. The few sites identified by these methods as being under positive selection that are outside of regions previously implicated in antigenic determination might serve as a guide for future studies attempting more complete characterization of the antigenicity of FMDV.

Because the power of this analysis is largely unknown, there is no reason to suppose that all sites on which positive selection might act have been identified. The ability of this test to identify positive selection may depend on the number and relative divergence of the sequences, clarity and structure of phylogenetic signal, and adequacy of the underlying model of nucleotide evolution (in this case the model of GOLDMAN and YANG 1994) to describe the evolution of sequences. Furthermore, this model, like almost all others, assumes that

TABLE 2

Parameters and likelihood values corresponding to results from model M3

Data set	Serotype/gene	M3(1)		M3(2)						M3(3)							
		ω_1	LL	p_1	ω_1	p_2	ω_2	LL	Significance	p_1	ω_1	p_2	ω_2	p_3	ω_3	LL	Significance
D1	O capsids	0.06	−16503.3	0.88	0.02	0.12	0.38	−16207.2	≤0.0001	0.17	0.21	0.81	0.01	0.02	0.97	−16179.8	≤0.0001
D2	A capsids	0.08	−8934.5	0.87	0.02	0.13	0.59	−8760.6	≤0.0001	0.80	0.01	0.16	0.30	0.04	1.15	−8755.3	0.0050
D3	C capsids	0.09	−6694.6	0.92	0.03	0.08	0.92	−6616.9	≤0.0001	0.77	0.01	0.22	0.30	0.01	2.80	−6610.5	0.0017
D4	O VP1	0.13	−5695.8	0.8	0.05	0.20	0.50	−5592.0	≤0.0001	0.64	0.03	0.31	0.25	0.05	1.00	−5580.3	≤0.0001
D5	C VP1	0.26	−2097.1	0.83	0.09	0.17	1.33	−2072.8	≤0.0001	0.46	0.00	0.44	0.28	0.10	1.68	−2072.5	0.7410
D6	SAT-1 VP1	0.10	−1643.6	0.93	0.04	0.07	1.93	−1574.8	≤0.0001	0.81	0.01	0.15	0.28	0.04	3.65	−1570.7	0.0166
D7	SAT-2 VP1	0.13	−2750.0	0.86	0.03	0.14	0.87	−2635.3	≤0.0001	0.83	0.02	0.11	0.43	0.06	1.51	−2629.1	0.0020
D8	SAT-2 VP1	0.71	−715.0	0.96	0.24	0.04	17.96	−705.8	≤0.0001	0.82	0.00	0.16	2.59	0.02	34.08	−705.3	0.7410
D9	SAT-3 VP1	0.12	−1385.7	0.88	0.04	0.12	0.87	−1364.1	≤0.0001	0.76	0.01	0.22	0.36	0.02	2.71	−1362.1	0.1353
D10	O VP1	0.07	−1016.5	0.51	0.01	0.49	0.13	−1015.0	0.223	0.22	0.01	0.29	0.01	0.49	0.13	−1015.0	1.0
D11	O VP1	0.06	−379.2	0.50	0.06	0.50	0.06	−379.2	1.0	0.33	0.06	0.33	0.06	0.33	0.06	−379.2	1.0

For each model the proportion of codon sites in the ith category ($i = 1, 2, 3$) is reported (p_i), together with the values of d_n/d_s assigned to those categories (ω_i) and the log likelihood of the fitted model (LL). For M2 and M3 we report the significance of the improved likelihood deduced from the likelihood-ratio tests comparing model M3(k) with M3($k-1$), assuming twice the log-likelihood difference to be chi-square distributed with 2 d.f. Underlined numbers indicate models associated with a significantly increased level of support compared to the next simplest model.

codon positions evolve independently of each other and at fixed d_n/d_s ratios in all branches, while at the same time assuming that recombination does not occur. Therefore, while absence of positive selection in data sets from North Africa and Saudi Arabia is understandable because sequencing did not span gene regions thought to encode epitopes, we do not think that our failure to detect significant levels of positive selection in SAT-3, C and O type VP1, and complete O capsid sequences is necessarily indicative of its absence.

This general form of analysis has been used previously to identify the action of positive selection in HIV *env*, *gag*, *pol*, *vif*, and *nef* genes (Nielsen and Yang 1998; Zanotto *et al.* 1999; Yang *et al.* 2000) and also in hominoid mitochondrial protein genes, vertebrate β-globin genes, and the HA1 domain of the HA gene of influenza A (Yang *et al.* 2000). The method has failed to detect positive selection in flavivirus E-glycoprotein genes, Japanese encephalitis *env* genes, and tick-borne flavivirus NS-5 genes (Yang *et al.* 2000). This method has also been used successfully to recover d_n/d_s ratios of sequences generated by computer simulation at known d_n/d_s ratios (D. Haydon, unpublished results). However, it should be stressed that while no analyses that have previously used this method have detected a great sensitivity of results to reconstructed tree topology, a rigorous understanding of likelihoods of false-positive or false-negative results remains to be developed. The relationship of these various models to each other is interesting in this respect: M3 models represent a generalization of the earlier "positive selection" model of Nielsen and Yang (1998) in which d_n/d_s ratios for two of three categories are fixed at values of 0 and 1. The M7 and M8 models represent a more flexible analysis still. It is noteworthy that the positive selection model of Nielsen and Yang (1998; denoted M2 in Yang *et al.* 2000) failed to identify positive selection in any of these data sets, whereas use of nested M3 models would lead to a type 1 error with respect to type C viruses from the Philippines (see Table 2) when compared to analyses based on the M7 and M8 models.

Unrecognized recombination events occurring within a bifurcating phylogeny may lead to an overestimate of the number of substitution events over the tree, but there is no reason to suppose that the estimated overall number of nonsynonymous substitutions will be more biased by this process than the estimated overall number of synonymous substitutions. However, since this analysis has been specifically constructed to identify codon positions with higher d_n/d_s ratios it appears possible that recombination could potentially lead to some false-positive results, and that this issue requires consideration when recombination rates may be very high. While recombination in FMDV is thought to occur frequently, the short duration of infection and rarity of dual infections suggest that the genetically meaningful recombination rate is likely to be small in most of these data sets.

TABLE 3

Amino acid usage at sites identified to be under positive selection

Site under positive selection	Amino acids used at site within data set	Significance
74 (VP2 BC loop)[C]	E K N S	
438 (VP3 C terminus)[CA]	A L PQRSTV	p h
482 (VP1 BC loop)[C]	H K NPQ T	pv h
537 (VP1 EF loop)[SAT1]	A D G N ST	v h
549 (VP1 FG loop)[SAT1]	A H K T	
569 (VP1 GH loop)[SAT1]	A L P ST	p m
575 (VP1 GH loop)[SAT2]	E K STV	h
577 (VP1 GH loop)[SAT2]	DE K N QRST	vmh
581 (VP1 GH loop)[A]	A EG M P ST	h
585 (VP1 GH loop)[SAT1]	DEG T	h
596 (VP1 GH loop)[SAT1]	A E H N Q	h
601 (VP1 GH loop)[SAT2]	A G IK R T	h
618 (VP1 HI loop)[SAT1]	DE H K Q	vmh
619 (VP1 I strand)[SAT2]	A E K N ST	h
620 (VP1 I strand)[SAT2]	A Q S	
645 (VP1 C terminus)[SAT2]	A G NP V	m
646 (VP1 C terminus)[SAT2]	D G N S	h

Only sites identified by models (M8) indicating positive selection that provide significant increases in likelihood relative to the null model (M7, see Table 1) are included, and of these, only sites with posterior probabilities of belonging to the positively selected category exceeding 0.95. Superscripts in the first column indicate with which serotype the selected sites were associated. Lowercase p, v, m, and h indicate significant conservation (at the 0.05 level) of polar requirement, amino acid volume, molecular weight, and/or hydrophilicity, respectively. Uppercase letters would indicate amino acid sets with significantly different physicochemical properties (no such sets were encountered).

The exception might be the sequences from extended infections of buffalo, where meaningful recombination rates could possibly be higher.

With these cautionary remarks in mind, the identification of positive selection in the SAT-2 sequences from the Kruger National Park (D7) isolated from impala and buffalo is particularly interesting since the phylogeny of these isolates shows repeated interspecific transmission between these two host species (Bastos *et al.* 2000). The highest d_n/d_s ratio is seen in SAT-2 virus isolated from a single persistently infected individual buffalo (D8). While this very high ratio should be regarded skeptically, based as it is on a low number of sequences, the presence of positive selection in the data is interesting as it leads to the suggestion that the generation of antigenic variation might be a mechanism by which virus is able to persist in the face of an immune response. Cross-neutralization tests using serum from this persistently infected buffalo indicated that substantial antigenic change had accumulated within the virus isolated over the course of this infection (Vosloo *et al.* 1996). How virus evades immune clearance in persistently infected animals remains unclear (Salt 1993), and previous work (Salt *et al.* 1996) has indicated no role for novel antigenic variation, but this is a mechanism suggested by these results.

Finally, this analysis shows that positive selection can result in a pattern of amino acid substitution that is physicochemically conservative. Conservatism may be evident in volume, polarity, hydrophilicity, or molecular weight of amino acids, depending presumably on the exact nature of the structural context of residue positions. However, that the conservation of hydrophilicity appears to be the principal requirement is not unexpected given that these hypervariable amino acid positions are located on the outer capsid surface. This analysis is consistent with that of Haydon *et al.* (1998) in suggesting that structural constraints imposed by capsid architecture are sufficient to preclude more radical amino acid substitutions. This is in contrast to the findings of Hughes *et al.* (1990) who found that positive selection promoted charge profile diversity in the binding cleft of class 1 MHC molecules.

Positive selection on variants could arise during the course of infection as a result of a specific immune response and be transmitted to immunologically naive individuals. Alternatively, it is possible that the bulk of virus replication and transmission occurs in a short window prior to activation of specific immune responses, and that arising variants are of no selective advantage over the course of a single infection. However, these variants may be endowed with selective advantage at the transmission stage at which point they may be able to infect partially immune individuals with some experience of FMDV antigen, either through previous infection or vaccination. Examination of multiple virus sequences

obtained from single individual (nonpersistent) infections could be used to distinguish between these two hypotheses.

We thank Simon Frost, Simon Wain-Hobson, Mark Woolhouse, and Ziheng Yang for many helpful discussions regarding the use of these techniques; two anonymous referees for helpful comments on the manuscript; Susan Lea for providing the structural guide to the O serotype; and David Ansell for technical assistance. D.T.H. is supported by a grant from Wellcome Trust.

LITERATURE CITED

Acharya, R., E. Fry, D. Stuart, G. Fox, D. Rowlands *et al.*, 1989 The three dimensional structure of foot and mouth disease virus at 2.9 A resolution. Nature **337:** 709–716.

Bastos, A. D. S., C. I. Boshoff, D. F. Keet, R. G. Bengis and G. R. Thomson, 2000 Natural transmission of foot-and-mouth disease virus between African buffalo (*Syncerus caffer*) and impala (*Aepyceros melampus*) in the Kruger National Park, South Africa. Epidemiol. Infect. **124:** 591–598.

Baxt, B., V. Vakharia, D. M. Moore, A. J. Franke and D. O. Morgan, 1989 Analysis of neutralising antigenic sites on the surface of type A12 foot-and-mouth disease virus. J. Virol. **63:** 2143–2151.

Brown, F., 1995 Antibody recognition and neutralization of foot and mouth disease virus. Semin. Virol. **6:** 243–248.

Bush, R. M., W. M. Fitch, C. A. Bender and N. J. Cox, 1999a Positive selection on the H3 hemagglutinin gene of human influenza virus A. Mol. Biol. Evol. **16:** 1457–1465.

Bush, R. M., C. A. Bender, K. Subbarao, N. J. Cox and W. M. Fitch, 1999b Predicting the evolution of human influenza A. Science **286:** 1921–1925.

Butchaiah, G., and D. O. Morgan, 1997 Neutralization antigenic sites on type Asia-1 foot-and-mouth disease virus defined by monoclonal antibody-resistant variants. Virus Res. **52:** 183–194.

Creighton, T. E., 1983 *Proteins: Structures and Molecular Properties*. W. H. Freeman, New York.

Crowther, J. R., S. Farias, W. C. Carpenter and A. R. Samuel, 1993a Identification of a fifth neutralizable site on type O foot and mouth disease virus following characterization of single and quintuple monoclonal antibody escape mutants. J. Gen. Virol. **74:** 1547–1553.

Crowther, J. R., C. A. Rowe and R. Butcher, 1993b Characterization of monoclonal-antibodies against a type SAT-2 foot-and-mouth disease virus. Epidemiol. Infect. **111:** 391–406.

Curry, S., E. Fry, W. Blakemore, R. Abu-Ghazaleh, T. Jackson *et al.*, 1996 Perturbations in the surface of A22 Iraq foot-and-mouth disease virus accompanying coupled changes in host cell specificity and antigenicity. Structure **4:** 135–145.

Davidson, F. L., J. R. Crowther, J. Nqindi, N. J. Knowles, S. J. Thevasagayan *et al.*, 1995 Antigenic analysis of SAT 2 serotype foot-and-mouth disease virus isolates from Zimbabwe using monoclonal antibodies. Epidemiol. Infect. **115:** 193–205.

Diez, J., M. G. Mateu and E. Domingo, 1989 Selection of antigenic variants of foot-and-mouth disease virus in the absence of antibodies as revealed by an *in situ* assay. J. Gen. Virol. **70:** 3281–3289.

Domingo, E., and J. J. Holland, 1992 Complications of RNA heterogeneity for the engineering of virus vaccines and antiviral agents, pp. 13–32 in *Genetic Engineering, Principles and Methods*, Vol. 14, edited by J. K. Setlow. Plenum Press, New York.

Feigelstock, D. A., M. G. Mateu, M. E. Piccone, F. De Simone, E. Brocchi *et al.*, 1992 Extensive antigenic diversification of foot-and-mouth disease virus by amino acid substitutions outside of the major antigenic site. J. Gen. Virol. **73:** 3307–3311.

Feigelstock, D. A., M. G. Mateu, D. A. Valero, D. Andrew, E. Domingo *et al.*, 1996 Emerging foot-and-mouth disease virus variants with antigenically critical amino acid substitutions predicted by model studies using reference viruses. Vaccine **14:** 97–102.

Felsentein, J., 1993 PHYLIP (Phylogeny inference package), version 3.5c. Department of Genetics, University of Washington, Seattle.

Fitch, W. M., J. M. E. Leiter, X. Li and P. Palese, 1991 Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. USA **88:** 4270–4274.

Goldman, N., and Z. H. Yang, 1994 Codon-based model of nucleotide substitution for protein-coding dna-sequences. Mol. Biol. Evol. **11:** 725–736.

Haydon, D. T., S. Lea, E. Fry, N. J. Knowles, A. R. Samuel *et al.*, 1998 Characterizing sequence variation in the VP1 capsid proteins of foot and mouth disease virus (serotype O) with respect to virion structure. J. Mol. Evol. **46:** 465–475.

Hughes, A. L., 1992 Positive selection and interallelic recombination at the Merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol. Biol. Evol. **9:** 381–393.

Hughes, A. L., T. Ota and M. Nei, 1990 Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class 1 major-histocompatibility-complex molecules. Mol. Biol. Evol. **7:** 515–524.

Kitson, J. D. A., D. McCahon and G. J. Belsham, 1990 Sequence analysis of monoclonal antibody resistant mutants of type O foot and mouth disease virus: evidence for the involvement of the three surface exposed capsid proteins in four antigenic sites. Virology **179:** 26–34.

Lea, S., J. Hernandez, W. Blakemore, E. Brocchi, S. Curry *et al.*, 1994 The structure and antigenicity of a type C foot and mouth disease virus. Structure **2:** 123–139.

Marquardt, O., M. M. Rahman and B. Freiberg, 2000 Genetic and antigenic variance of foot-and-mouth disease virus type Asia1. Arch. Virol. **145:** 149–157.

Martinez, M. A., J. Hernandez, M. E. Piccone, E. L. Palma, E. Domingo *et al.*, 1991 Two mechanisms of antigenic diversification of foot-and-mouth disease virus. Virology **184:** 695–706.

Martinez, M. A., J. Dopazo, J. Hernandez, M. G. Mateu, F. Sobrino *et al.*, 1992 Evolution of capsid protein genes of foot-and-mouth disease virus: antigenic variation without accumulation of amino acid substitutions over six decades. J. Virol. **66:** 3557–3565.

Mateu, M. G., D. A. Silva, E. Rocha, D. L. De Brum, A. Alonso *et al.*, 1988 Extensive antigenic heterogeneity of foot-and-mouth disease virus of serotype C. Virology **167:** 113–124.

Mateu, M. G., M. A. Martinez, L. Capucci, D. Andrew, E. Giralt *et al.*, 1990 A single amino acid substitution affects multiple epitopes in the major antigenic site of foot-and-mouth disease virus of serotype C. J. Gen. Virol. **71:** 629–637.

Mateu, M. G., J. A. Camarero, E. Giralt, D. Andreu and E. Domingo, 1995 Direct evaluation of the immunodominance of a major antigenic site of foot-and-mouth-disease virus in a natural host. Virology **206:** 298–306.

McCullough, K. C., F. D. Simone, E. Brocchi, L. Capucci, J. R. Crowther *et al.*, 1992 Protective immune response against foot and mouth disease. J. Virol. **66:** 1835–1840.

Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:** 929–936.

Parry, N. R., P. V. Barnett, E. J. Ouldridge, D. J. Rowlands and F. Brown, 1989 Neutralizing epitopes of type O foot-and-mouth disease virus. II. Mapping three conformational sites with synthetic peptide reagents. J. Gen. Virol. **79:** 1493–1503.

Pfaff, E., H.-J. Thiel, E. Beck, K. Strohmaier and H. Schaller, 1988 Analysis of neutralizing epitopes on foot and mouth disease virus. J. Gen. Virol. **62:** 2033–2040.

Salt, J. S., 1993 The carrier state in foot-and-mouth-disease—an immunological review. Br. Vet. J. **149:** 207–223.

Salt, J. S., A. R. Samuel and R. P. Kitching, 1996 Antigenic analysis of type O foot-and-mouth disease virus in the persistently infected bovine. Arch. Virol. **141:** 1407–1421.

Samuel, A. R., N. J. Knowles, R. P. Kitching and S. M. Hafez, 1997 Molecular analysis of foot-and-mouth disease type O viruses isolated in Saudi Arabia between 1983 and 1995. Epidemiol. Infect. **119:** 381–389.

Samuel, A. R., N. J. Knowles and D. K. J. Mackay, 1999 Genetic analysis of type O viruses responsible for epidemics of foot-and-mouth disease in North Africa. Epidemiol. Infect. **122:** 529–538.

Seibert, S. A., C. Y. Howell, M. K. Hughes and A. L. Hughes, 1995 Natural selection on the gag, pol, and env genes of the human immunodeficiency virus (HIV-1). Mol. Biol. Evol. **12:** 803–813.

Sharp, P. M., 1997 In search of molecular Darwinism. Nature **385:** 111–112.

Stave, J. W., J. L. Cane, D. O. Morgan and V. N. Vakharia, 1988

Neutralisation sites of type O1 foot-and-mouth disease virus defined by monoclonal antibodies and neutralisation escape variants. Virology **162:** 21–29.

Suzuki, Y., and T. Gojobori, 1999 A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **16:** 1315–1328.

Thomas, A. A. M., R. J. Woortmeijer, W. Puijk and S. J. Barteling, 1988 Antigenic sites on foot-and-mouth disease virus type A10. J. Virol. **62:** 2782–2789.

Vosloo, W., A. D. Bastos, E. Kirkbride, J. J. Esterhuysen, D. J. Van Rensburg *et al.*, 1996 Persistent infection of African buffalo (*Syncerus caffer*) with SAT-type foot-and-mouth-disease viruses: rate of fixation of mutations, antigenic change and interspecies transmission. J. Gen.Virol. **77:** 1457–1467.

Woolhouse, M. E. J., D. T. Haydon, A. Pearson and R. P. Kitching, 1996 Failure of vaccination to prevent outbreaks of foot and mouth disease. Epidemiol. Infect. **116:** 363–371.

Xie, Q-C., D. McCahon, J. R. Crowther, G. L. Belsham and K. C. McCullough, 1987 Neutralization of foot and mouth disease virus can be mediated through any of at least three separate antigenic sites. J. Gen. Virol. **68:** 1637–1647.

Yang, Z., R. Nielsen, N. Goldman and A.-M. Krabbe Pederson, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Zamyatnin, A. A., 1972 Protein volume in solution. Prog. Biophys. Mol. Biol. **24:** 109–123.

Zanotto, P. M., E. G. Kallas, R. F. Souza and E. C. Holmes, 1999 Genealogical evidence for positive selection in the nef gene of HIV-1. Genetics **153:** 1077–1089.

Communicating editor: S. Yokoyama