

# Variation in the VP1 gene of *Foot-and-mouth disease virus* serotype A associated with epidemiological characteristics of outbreaks in the 2001 epizootic in Argentina

Andres M. Perez,<sup>1</sup> Guido König, Ernesto Späth, Mark C. Thurmond

**Abstract.** A mixed binomial Bayesian regression model was used to quantify the relation between nucleotide differences in the VP1 gene of *Foot-and-mouth disease virus* (FMDV) serotype A, and epidemiologic characteristics of the outbreaks from which the viruses were obtained between January and December 2001 in Argentina. An increase in the probability of different nucleotides between isolates was associated with a longer time between isolation dates, a greater distance between isolation locations, an increase in the difference between attack rates, and an increase in the difference in outbreak durations. The farther apart the outbreak herds were in the southerly and easterly directions, the greater the difference in nucleotide changes. The model accurately predicted genetic distances of isolates obtained in 2001 compared with a 2002 isolate ( $P < 0.01$ ), which suggested that the predictive modeling approach applied in the present study may be useful in understanding the epidemiology of evolution of FMDV and in forensic analysis of disease epidemics.

**Key words:** Argentina; *Foot-and-mouth disease virus*; molecular epidemiology; risk factors.

## Introduction

*Foot-and-mouth disease virus* (FMDV; family *Picornaviridae*, genus *Aphthovirus*) is a highly contagious virus that affects a wide range of cloven-hoofed mammals throughout the world. The disease impairs meat and milk productivity, and limits the use of affected animals for draft purposes. Countries with FMDV are not allowed to export animals or animal products to FMDV-free countries, which contributes further to the costs of the disease. Consequently, FMDV is considered one of the most important diseases of livestock, because it imposes such profound and sustained economic and social hardships on the people and the governments of countries with FMDV.

*Foot-and-mouth disease virus* has 7 known immunologically distinct serotypes and can experience a high rate of mutation,<sup>20</sup> in which the average rate of nucleotide changes is believed to range from  $<4 \times$

$10^{-4}$  to  $4 \times 10^{-2}$  per locus per replication.<sup>2</sup> Much of this variation ( $10^{-2}$ – $10^{-3}$  nucleotide changes per locus per replication) occurs in the VP1 gene, which has 639 nucleotides that encode for the capsid protein believed to play an important role in cell infection and host immune response.<sup>2,14,18</sup>

Based on the quasi-species theory,<sup>1,2</sup> the virus population would be expected to evolve over the course of the disease and that manifestations of infection, such as morbidity, mortality, and rate of transmission, would differ among the affected herds. Several genes, including the VP1 gene, encode for proteins that affect virulence, replication, entrance of the virus into host cells, and the severity of disease.<sup>6</sup> Thus, outbreaks of FMDV caused by strains more distantly related to each other would be expected to have a wider variation in morbidity, mortality, and intraherd transmission compared with outbreaks caused by strains more closely related to each other.

There is little understanding of how genetic variation of FMDV might be influenced by host and environmental factors or how the genetic variation can be expected to manifest in the host population over the course of the disease in natural conditions. The objective of the present study was to identify relationships between the variation in number of differing nucleotides within the VP1 gene of FMDV, the nature and severity of the natural disease in cattle populations, and features of the host population that were associated with the viruses.

---

From the Center for Animal Disease Modeling and Surveillance, Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, CA (Perez, Thurmond), CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Buenos Aires, Argentina (Perez, König), the Instituto de Biotecnología, INTA (Instituto de Tecnología Agropecuaria), Castelar, Argentina (König), and the Grupo de Sanidad Animal, INTA, Balcarce, Argentina (Späth).

<sup>1</sup>Corresponding Author: Andres M. Perez, Center for Animal Disease Modeling and Surveillance, Department of Medicine and Epidemiology, School of Veterinary Medicine, One Shields Avenue, University of California, Davis, CA 95616. amperez@ucdavis.edu

**Materials and methods**

**Data collection**

From 2000 to 2002, Argentina experienced a major FMDV epidemic. Most of the outbreaks were caused by 2 subtypes of serotype A virus; 1 subtype was isolated only in 2000 (A2000) and the other was isolated mainly in 2001 (A2001). The A2001 isolates were classified into 2 subgroups specified as A2001A and A2001B.<sup>10,15,16,23</sup>

A2001 FMDV isolates and epidemiologic data were obtained from 23 herds affected by FMDV in 2001 (Fig. 1). The procedure used to isolate and sequence the FMDV from the clinical samples was described elsewhere.<sup>10</sup> The geographic direction of the epidemic, as indicated by the direction of the progression of the 2,519 reported FMDV outbreaks,<sup>15</sup> was estimated by using the Oden test,<sup>7</sup> which computed the average direction of the epidemic (estimated as the average angle of the vectors that connect each outbreak with all the consecutive outbreaks in the epidemic).

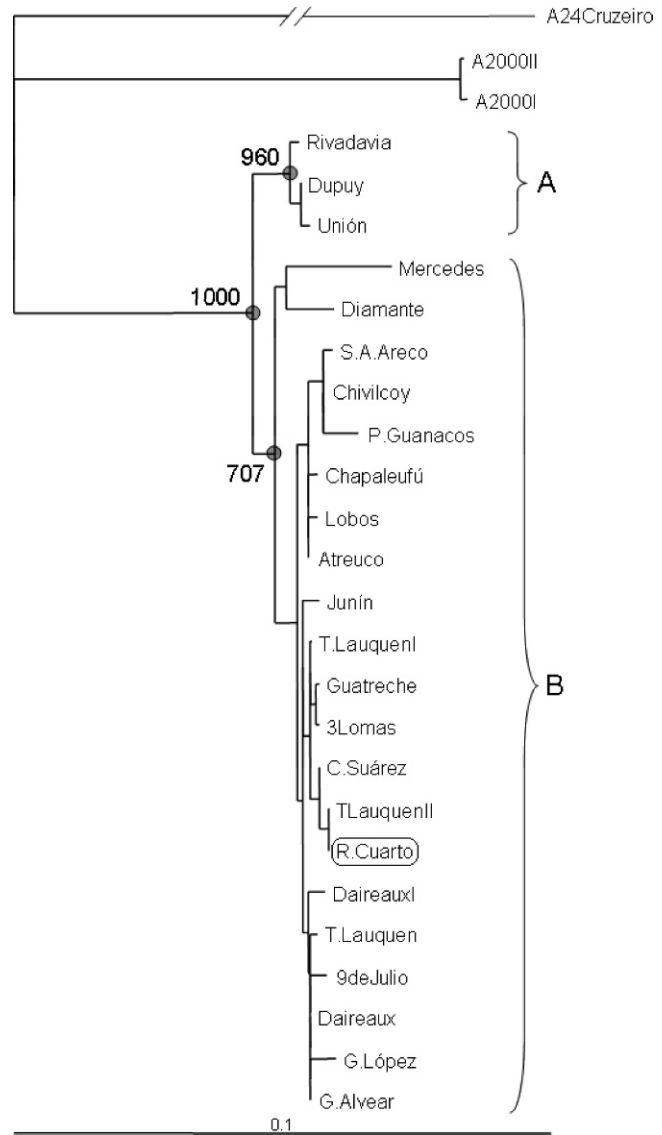
Hypothesized epidemiologic features of the disease, as manifested by each herd, were examined to assess whether nucleotide changes between any 2 isolates (i and j) from 2 herds were associated with differences in disease manifestation for the herds. The variables were

- $H_{i,j}$  = the absolute difference between the number of animals in the 2 herds.
- $DU_{i,j}$  = the absolute difference between the durations (days) of the outbreaks in the 2 herds. Duration was estimated as the difference between the date the last clinical case was observed in the herd and the estimated date the virus was first introduced into the herd, which was estimated from information obtained in the epidemiologic investigations of the outbreaks. Information included ages of the oldest lesions and histories of animal movement into the herds.
- $AR_{i,j}$  = absolute difference in the attack rates for the 2 herds and was estimated as the absolute value of the difference in the proportion of animals that had clinical signs of FMDV for each of the 2 herds, or  $AR_{i,j} = |AR_i - AR_j|$ .
- $T_{i,j}$  = absolute difference in time (days) between the estimated onset of the outbreaks in the herds; the date an outbreak began was estimated as described above for  $DU_{i,j}$ , or  $T_{i,j} = |T_i - T_j|$ .
- $KM_{i,j}$  = spatial distance (km) between the 2 herds; distance was measured by using the Haversine formula, which gives the great-circle distances between 2 points on a sphere of radius  $R$ , based on the longitude (long) and latitude (lat) of the points, so that

$$KM_{i,j} = 2 \times R \times \arcsin \sqrt{\sin^2 \frac{|\text{lat}_i - \text{lat}_j|}{2} + \cos(\text{lat}_i) \times \cos(\text{lat}_j) \sin^2 \frac{|\text{long}_i - \text{long}_j|}{2}}$$

where  $R = 6,367$  km (i.e., the radius of the Earth), and lat and long are measured in radians.

- $S_{i,j}$  = distance (km) along the north-south axis between the 2 herds. This variable and the  $E_{i,j}$  variable below were included as measures of spatially



**Figure 1.** Neighbor-joining tree outgroup-rooted on A24 Cruzeiro and A2000 isolates showing relationships among A2001 *Foot-and-mouth disease virus* isolates obtained in Argentina in 2001 (n = 23) and 2002 (n = 1, marked with a circle) based on complete VP1 nucleotide sequence. Subgroups (A2001A and A2001B), relevant nodes supported by bootstrap method (1,000 resampling), and relative genetic distance (bar at the bottom) are indicated. GenBank accession numbers from **top** to **bottom** are K03340, AM179989, AM17998, AM180000, AM180002, AM180003, AM180011, AM180022, AM180005, AM180023, AM180009, AM180001, AM180006, AM180007, AM180013, AM180015, AM180016, AM180014, AM180012, AM180024, AM180017, AM180019, AM180008, AM180020, AM180010, AM180004, and AM180021.

directional changes over the course of the epidemic, which would not be accounted for simply by the straight-line distance between outbreaks or by the time difference between outbreaks. A positive sign indicated that the distance moved in a southerly direction between outbreaks and was associated with

a higher probability of nucleotide changes compared with distance moved in a northerly direction.

- $E_{i,j}$  = distance (km) along the east-west axis between the 2 herds. A positive sign indicated that the distance moved in an easterly direction and was associated with a higher probability of nucleotide changes, compared with distance moved in a westerly direction.

### Sequence analysis

VP1 sequences of the A2001 FMDV isolates were aligned by using DNASIS, version 2.5,<sup>a</sup> and ClustalX Multiple Sequence Alignment Program, version 1.81.<sup>b,21</sup> The final alignment was made manually. A phylogenetic tree was constructed by using the PHYLIP (Phylogeny Inference Package), version 3.5,<sup>c</sup> and the Kimura 2-parameters model,<sup>8</sup> and was analyzed by using the neighbor joining method.<sup>17</sup> A Parsimony tree was constructed with 100 random addition sequences tree bisection reconstruction with NONA, version 2.0.<sup>d,4</sup> The support of internal tree nodes was assessed by using a 1,000 replication bootstrap analysis.

### Model to predict genetic differences

A model was developed to examine the extent to which VP1 genomic differences between isolates obtained from herds were associated with variation in the 7 hypothesized factors described above. The number of different nucleotides (dn) in the VP1 gene was calculated for each pair of isolates ( $i,j$ ) from all possible pairs of the 23 herds, by using MEGA (Molecular Evolutionary Genetics Analysis) software, version 2.1.<sup>e,11</sup> Associations between  $dn_{i,j}$  and the 7 hypothesized factors were assessed by using a Bayesian regression model, where  $dn_{i,j}$  was assumed to follow a binomial ( $n,p$ ) process:  $n$  was the number of nucleotides compared in each pair of isolates ( $n = 639$ , the length of VP1 gene), and  $p$  was the probability that a nucleotide differs between the 2 VP1 sequences, which when presented as a percentage is referred to as genetic distance. Each factor was incorporated in the model as an independent fixed effect, and 1 nonstructured random effect ( $U$ ) was included to account for variation in factors not explicitly considered in the model. The model was formally expressed as  $dn \sim \text{Binomial}(n,p)$ ,  $\ln[p/(1-p)] = \alpha + \beta f + U$ , where  $\alpha$  denoted the model intercept and  $\beta$  denoted the regression coefficients for each of the 7 hypothesized factors ( $f$ ).

Because no information exists about any association between the hypothesized factors and genetic distance, regression coefficients for fixed and random effects were modeled by using a noninformative prior distribution of the form  $N(0,100)$  and  $N[0, \delta \sim \text{gamma}(0.5, 0.0005)]$ , respectively. Models were run by using WinBUGS (Bayesian inference by using Gibbs sampling) software,<sup>f,12</sup> with 20,000 iterations after 1,000 iterations discarded. Values of the variables were standardized by dividing the difference between the value of the variable for a pair of outbreaks and the overall mean value of the variable by the standard deviation of the variable. Variable transformations that were considered included log, square root, power, and exponential, and all possible 2-way interactions were examined. Depending on the number of variables in the

model, a large value of the deviance information criterion (DIC) was considered evidence that the model did not fit the data as well as a model with a small DIC value.<sup>19</sup> The strength of an association between factors  $f$  and  $p$  was measured by the odds ratio (OR), which was estimated as the exponential of the posterior distribution of the regression coefficients.

### Model validation

The model that best fits the data, as indicated by the DIC, was used to estimate a predicted genetic distance (PD) between each of the 23 isolates obtained in 2001 and an isolate obtained in January 2002.<sup>10</sup> Accuracy of the model was evaluated in part by the correlation coefficient (R), estimated with the Spearman correlation test, between the values of PD and the corresponding observed genetic distance (OD). In addition, outliers in the correlation were identified with the Grubb test,<sup>5</sup> in which an outlier was interpreted to indicate failure to PD. To assess whether there was a tendency of the model to systematically underestimate or overestimate the genetic distance, a paired  $t$ -test was used to test the mean differences between the OD and PD. A  $P$  value of  $\geq 0.05$  was interpreted to indicate no over- or underestimation of genetic distance.

The error in the PDs, compared with ODs, was estimated by the mean absolute prediction error (MAPE) as

$$P_x = \left[ \sum (|Pdn_{i,j} - Odn_{i,j}| \leq x) \right] / 23,$$

where  $Pdn_{i,j}$  and  $Odn_{i,j}$  are the respective predicted and observed number of nucleotides that differed between each isolate  $i$  obtained in 2001 and the isolate  $j$  obtained in 2002,  $x$  is a threshold value that varied from 1 to 10 for the error in the predicted number of nucleotides that differed between 2 isolates, and  $P_x$  is an estimate of the proportion of isolates  $i$  with an absolute difference between predicted and observed values equal to or less than the threshold value.

### Model application

Results of the model were used to assess whether the 2002 isolate evolved from the 2001A or the 2001B lineage. The deviance ( $D_i$ ) between the OD of the isolates obtained in 2001 and the 2002 isolate, and the PD from the model, was estimated as  $OD/PD_i$  for each of the  $i$  isolates obtained in 2001. By using a Mann-Whitney test, the mean value of  $D_i$  was compared between the 2 lineages (A2001A and A2001B). A significant difference ( $< 0.05$ ) was interpreted to indicate the 2002 isolate was more closely related to the lineage with the smaller mean value of deviance. If there was a significant difference in the mean  $D_i$  values, the Spearman correlation test was used to estimate the strength of correlation between PD and OD separately for each lineage. A Moran I was computed to assess the extent at which  $D_i$  was spatially clustered. Failure to find a Moran I that was significantly ( $P < 0.05$ ) different than 1 was interpreted to indicate that  $D_i$  was not spatially clustered. Results of the model were validated by using results of the phylogenetic tree analysis.

**Table 1.** Association between epidemiologic characteristics of *Foot-and-mouth disease virus* (FMDV) outbreaks and the number of different nucleotides and rate of change in the VP1 gene of 23 FMDV serotype A isolates obtained during the FMDV 2001 epidemic in Argentina.\*

| Epidemiologic factor                    | Median (minimum/maximum)        | Value of 1 standardized unit | Odds ratio (95% PI)† | Predicted rate of nucleotide differences (95% PI) |
|---|---------------------------------|------------------------------|----------------------|---|
| Distance in southerly direction         | 19 (-1257/1183) km              | 229 km                       | 1.08 (1.00–1.16)     | 4 (3–4) per 100 km                                |
| Distance in easterly direction          | 22 (-524/592) km                | 398 km                       | 1.10 (1.03–1.20)     | 2 (2–2) per 100 km                                |
| Distance                                | 290 (0/1257) km                 | 609 km                       | 1.19 (1.08–1.32)     | 1 (1–2) per 100 km                                |
| Time                                    | 28 (0/239) days                 | 124 days                     | 1.14 (1.02–1.26)     | 24 (21–30) per year                               |
| Difference in attack rates              | 0.16 (0/0.67) cases/no. at risk | 0.3 cases/no. at risk        | 1.10 (0.99–1.22)     | 2 (2–3) per 10% difference                        |
| Difference in durations of outbreaks    | 8 (0/32) days                   | 16.0 days                    | 1.03 (0.96–1.13)     | 2 (2–3) per day of difference                     |
| Interaction of attack rate and duration | NA                              | NA                           | 1.07 (1.00–1.14)     | NA  |

\* PI = probability interval; NA = not applicable.

† Increase in odds of all nucleotides of the VP1 genes of 2 isolates being different when the difference in the value of the epidemiological factor increases by 1 standardized unit.

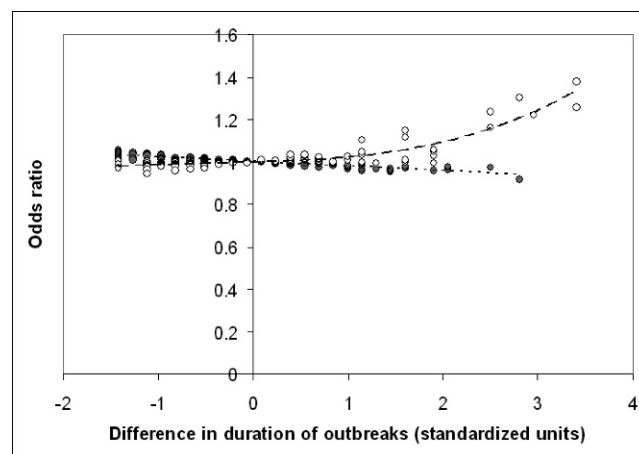
## Results

The mean and median number of nucleotides that differed among the 253 pairwise comparisons possible for the 23 isolates, which included both the A2001A and A2001B groups (Fig. 1), were 8.6 and 6.0, respectively. The minimum and maximum numbers of differing nucleotides were 0 and 24, respectively. The average direction of spread of the epidemic was 116 degrees southeasterly (Oden test,  $P < 0.01$ ).

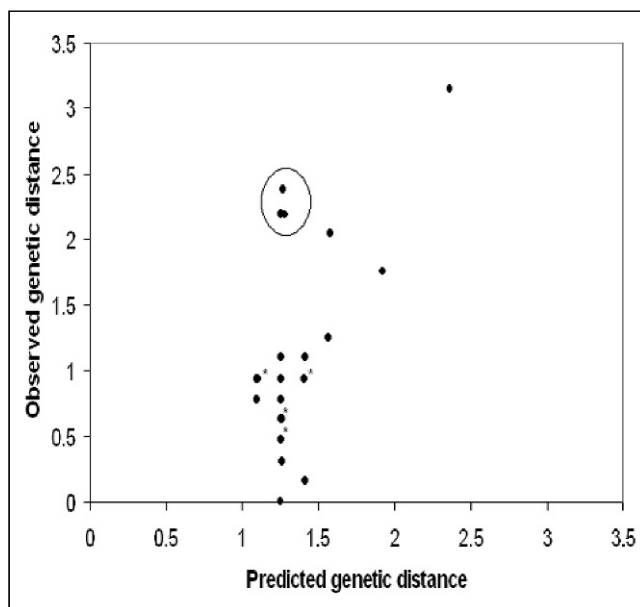
The model that best fit the data (lowest DIC value = 1,371.5) included all of the hypothesized epidemiologic factors except herd size and an interaction between the difference in the attack rates for each herd from which the isolates were obtained, and the difference in durations of the outbreaks (Table 1). As indicated by the OR values  $>1.0$ , the farther apart the herd outbreaks were in time (24 nucleotide changes per year) and in spatial distance (1 nucleotide change per 100 km), the greater were the number of nucleotides that differed in the VP1 gene. The number of nucleotide differences between any 2 isolates was more likely to be large if the latter of the 2 outbreaks was located farther to the south (4 nucleotide changes per 100 km) or to the east (2 nucleotide changes per 100 km) of the earlier outbreak, than if located elsewhere. The interaction factor that involved the difference in attack rates and the difference in disease durations indicated that the association between the genetic distance ( $p$ ) and the difference in attack rates (2 nucleotide changes per 10% difference in attack rates) depended on the difference in outbreak duration (2 nucleotide changes per day of difference) between 2 herds, and vice versa. Specifically, the odds of there being a difference in the nucleotide sequences of the VP1 genes increased when both the difference in the attack rates and the difference in outbreak durations were high. However, if differences were low, or if only

the difference in the durations was high, the odds of there being a difference in the nucleotide sequences were low or nil. The interaction is seen visually in Figure 2 as the divergence in OR lines as the difference in duration of the outbreaks increases.

Based on results of the validation procedures, the model appeared to fit the data well. The ODs (percentage of nucleotides that differ) between the VP1 genes of each isolate obtained in 2001 and that of the isolate obtained in 2002 were significantly correlated ( $R = 0.6$ ;  $P < 0.01$ ), with the corresponding genetic distances predicted by the model (Fig. 3). No outliers were detected in the series of paired predicted and observed values (Grubb test,  $P < 0.05$ ). There were no significant differences (paired  $t$ -test,  $P = 0.11$ ) between mean PD (1.3%) and mean OD (1.1%), which



**Figure 2.** Interaction of the difference in attack rates and the difference in outbreak durations in an association with differences in the number of nucleotides of 23 isolates of *Foot-and-mouth disease virus* in Argentina, 2001. Dots indicate greater than (white dots) and less than (gray dots) the median standardized difference in attack rates, respectively.



**Figure 3.** Observed and predicted genetic distance, expressed as a percentage, between 23 A2001 *Foot-and-mouth disease virus* isolates obtained in Argentina in 2001 and an isolate obtained in 2002. Dots indicate pairs, and the asterisk (\*) indicates more than 1 pair. The circle indicates the 3 pairs corresponding to the A2001 subgroup A.

indicated that the error was randomly distributed below and above the predictions, as would be expected if the model did not systematically over- or underestimate the ODs. The absolute difference between the observed and predicted number of nucleotide changes was  $\leq 3$  nucleotides for more than 11 (50%) of the isolates and  $\leq 8$  nucleotides for all 23 isolates of A2001, compared with the 2002 isolate (MAPE = 3.8).

Compared with the 2002 isolate, the ratio of observed-to-predicted genetic distances was significantly higher for the A2001A lineage (mean  $D_i = 1.78$ ) than for the A2001B lineage (mean  $D_i = 0.66$ ; Mann-Whitney test,  $P < 0.01$ ). Similarly, compared with the 2002 isolate, correlations between OD and PD for viruses in lineages A and B were  $R = -0.1$  ( $P = 0.93$ ) and  $R = 0.8$  ( $P < 0.01$ ), respectively, which suggested that the 2002 isolate was more closely related to lineage B than to lineage A viruses. No evidence of spatial clustering of  $D$  was found (Moran  $I = -0.38$ ,  $P = 0.26$ ), which suggested that the error in the predictions was spatially heterogeneous. Results of the phylogenetic analysis for validation suggested that the 2002 isolate was a member of lineage B (Fig. 1), which supported the model results.

### Discussion

Results of the present study of changes in the VP1 gene of A2001 FMDV isolated in Argentina suggest

that an increase in the variation between any 2 isolates, as measured by the number of different nucleotides between isolates, was associated with epidemiologic features of the disease in a given outbreak, namely the reduced temporal and spatial proximity of outbreaks to each other and the increased severity of disease, as estimated by the outbreak duration and attack rate. Although the annual rate of change in the VP1 gene estimated in the present study by the pairwise comparisons (Table 1) is higher than some previous estimates for the VP1 genes of types O and A,<sup>3,9,13,22</sup> it is substantially lower than the rate of 47 nucleotides per year estimated for persistent FMDV infection in cattle.<sup>2,3</sup> Interestingly, even after accounting for the effect of the difference in time between outbreaks, an effect of spatial proximity was found in which a greater distance between outbreaks was associated with greater nucleotide differences. Proximity may be a reflection of herd type or management practices, because herds farther away from each other might tend to be more dissimilar with respect to management factors that presumably could influence genetic diversity, such as timing or the extent of FMDV vaccination for a herd. The use of geospatial directional variables (Table 1), which take into account the sequential occurrence of the outbreaks and the spatial distance in a southerly direction and in an easterly direction, identified an additional spatial-temporal component of nucleotide change. The direction that was associated with the greatest nucleotide differences, moving from north to south and west to east, was the same as the direction of the spread of the epidemic (southeasterly 116 degrees). These findings indicate that, in addition to the “crude” spatial and temporal differences between any 2 outbreaks, there was an additional effect of directional spatial distance between outbreaks when they were placed in sequential order, which could be a proxy for the direction of the spread of the epidemic.

The interaction found between the elements of duration of an outbreak and the attack rate on the genetic distance offers insight into possible dynamics of viral evolution within a herd. The findings in the present study indicate that, when the pairs of outbreaks had similar attack rates or durations of disease, the odds of a nucleotide change (nucleotide substitution) in the VP1 gene did not change. However, when both the difference in the attack rates and the difference in durations between 2 pairs of outbreaks increased, the odds of nucleotide substitution in the region coding for the VP1 protein increased (Fig. 2). If the attack rate or duration of the outbreak was similar for a pair of outbreaks and all other factors were the same, then the number of virus

replications and the odds of mutation would be expected to be similar, because the same proportion of animals are infected (attack rate) and infection is present for the same period of time. However, higher attack rates together with a longer duration of an outbreak presumably could result in even more virus replication, which would be expected to result in more mutants within the virus population compared with the virus population emerging from outbreaks with lower differences in attack rates or durations of disease. Thus, one would have expected that, with a constant duration of the outbreak and a constant attack rate, an increase in the difference in herd size would have resulted in a greater number of FMDV replications and mutations. Differences in herd size between pairs of outbreaks, however, did not improve the model fitness, as indicated by the higher value of DIC computed when herd size was included into the model. A possible reason for the lack of significant contribution of herd size to the model fitness is that the number of samples or the magnitude of the difference in herd sizes was not sufficient to provide enough power to detect an association with herd size. In addition, the inclusion of other variables in the model may have accounted for the variation associated with differences in herd size, either as proxies or as confounders.

The model developed in the present study could be applied to estimate and understand degrees of relatedness between strains, which could provide information complementing that available from traditional phylogenetic trees. The model's predicted genetic distances between the 2002 isolate and the A2001B isolates were highly correlated with the ODs (Spearman  $R = 0.8$ ), whereas respective correlation for the A2001A isolates was very low (Spearman  $R = -0.1$ ). Despite the lesser confidence in the estimate of the correlation coefficient for the A2001A isolates, which was based on 3 isolates, compared with 20 A2001B isolates, visual inspection of the correlation cloud (Fig. 3), and the large difference between the estimates of correlation for A2001A (Spearman  $R = 0.8$ ) and A2001B (Spearman  $R = -0.1$ ) isolates suggest that the 2002 isolate was much more likely to be related to the A2001B lineage than to the A2001A lineage. These results were supported by the position of the 2002 isolate in the phylogenetic tree (Fig. 1), which was created independently of the model. One could argue that testing the model against an A2001A isolate would have been necessary to demonstrate that the model performed well for both lineages. In that case, the correlation between OD and PD would have been expected to be higher for A2001A than for A2001B isolates. However, only 1 FMDV outbreak occurred in Argentina in 2002 and, for that reason, no more samples were available to test the

performance of the model. Thus, how well the model might perform for the A2001A lineage was not able to be assessed.

An advantage with the modeling approach presented in the current study is that in the face of an outbreak, and only by using epidemiologic data obtained in the field, an epidemiologic model could be used to predict the genetic distance of the strain that caused the outbreak, compared with candidate strains (in this case, A2001A and A2001B strains). Thus, without specific data about the nucleotides of the outbreak strain, some inferences can be made about how closely related a new outbreak strain is to previously collected strains. The modeling approach presented will certainly not replace the use of traditional phylogenetic models. The model complements traditional phylogenetic modeling by providing insight into the epidemiology and clinical outcome of outbreaks associated with variation in the number of different nucleotides of prevailing epidemic strains of FMDV. However, in the face of multiple outbreaks, where a lack of resources often preclude rapid nucleotide analysis of isolates, the model could be used to identify which outbreaks might not be caused by closely related strains and to help establish priorities for sampling and nucleotide analysis.

Utility of such models in predicting genomic variation will depend on the quality of epidemiologic data obtained in the face of an outbreak. For example, in the present study, the effect of vaccination could not be accounted for because such data were not available. As noted above, however, it is believed that the time between outbreaks and proximity of outbreaks to each other could in part be proxies for vaccination; herds affected toward the end of the epidemic were more likely to have been vaccinated for FMDV than herds affected early in the epidemic.<sup>15,16</sup> Thus, herds with outbreaks that were closer to each other in time and place were probably more likely to have had a similar vaccination status. Similarly, distance and herd size could be in part proxies for management system and herd density, because one would expect that herds that have a similar size and that are closer to each other will be more likely to have similar management practices and animal densities than herds that are located far away from each other. It should be noted that the associations found between genetic distance and the hypothesized factors should not be interpreted to indicate causality in either direction. In the present retrospective study, there was no way to determine which came first, the change in attack rate or disease duration or the change in genetic distance.

In conclusion, the model described in the present study was able to predict genetic distance of a FMDV

strain isolated in 2002 by using sequence data for strains isolated in 2001 and epidemiologic data of the outbreaks. Further development and application of the methods presented in the present study could be useful as a complement to traditional phylogenetic analysis in studies of factors that drive genome changes during FMDV epidemics and in forensic analysis to predict the origin of FMDV isolated during the course of an epidemic.

### Acknowledgements

The authors thank Drs. Bernardo Cosentino and Gastón Funes (Department of Epidemiologia, SENASA, Buenos Aires, Argentina) for providing the epidemiologic data of the outbreaks and for assistance in interpreting the information. This project was funded in part by the United States Armed Forces Medical Intelligence Center, the Secretaría de Ciencia y Técnica, Buenos Aires, Argentina (grant PICT 08-08796), and the Instituto Nacional de Tecnología Agropecuaria, Buenos Aires, Argentina (project 52.2012).

### Sources and manufacturers

- a. MiraiBio Hitachi, Tokyo, Japan.
- b. French Society of Bioinformatics, Strasbourg, France. Available at: <http://bips.u-strasbg.fr/en/Documentation/ClustalX/>.
- c. University of Washington, Seattle, WA. Available at: <http://evolution.genetics.washington.edu/phylip.html>.
- d. INSUE Fundación e Instituto Miguel Lillo, San Miguel de Tucuman, Argentina. Available at: [http://www.cladistics.com/about\\_nona.htm](http://www.cladistics.com/about_nona.htm).
- e. Tokyo Metropolitan University, Tokyo, Japan; Stanford University, Palo Alto, CA; Pennsylvania State University, Pennsylvania, PA; Arizona State University, Tempe, AZ. Available at: <http://www.megasoftware.net/index.html>.
- f. MRC Biostatistics Unit, Cambridge, UK. Available at: <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>.

### References

1. Domingo E, Escarmis C, Baranowski E, et al.: 2003, Evolution of foot-and-mouth disease virus. *Virus Res* 91:47–63.
2. Domingo E, Ruiz-Jarabo CE, Arias A, et al.: 2004, Quasispecies dynamics and evolution of Foot-and-mouth disease virus. *In: Foot and mouth disease: current perspectives*, ed. Sobrino F, Domingo E, pp. 261–304. Horizon Bioscience, Norfolk, United Kingdom.
3. Gebauer F, de la Torre JC, Gomes I, et al.: 1998, Rapid selection of genetic and antigenic variants of foot-and-mouth disease virus during persistence in cattle. *J Virol* 62:2041–2049.
4. Goloboff PA: 1999, Analyzing large data sets in reasonable times: solutions for composite optima. *Cladistics* 15:415–428.
5. Grubbs FE: 1969, Procedure for detecting outlying observations in samples. *Technometrics* 11:1–21.
6. Grubman MJ, Baxt B: 2004, Foot-and-mouth disease. *Clin Microbiol Rev* 17:465–493.
7. Jacquez GM, Waller LA: 1999, The effect of uncertain locations on disease cluster statistics. *In: Quantifying spatial uncertainty in natural resources: theory and applications for GIS and remote sensing*, ed. Mowrer HT, Congalton RG, pp. 53–64. Sleeping Bear Press, Chelsea, MI.
8. Kimura H: 1980, A simple method for estimating evolutionary rate base of substitution through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.
9. Knowles NJ, Samuel AR: 2003, Molecular epidemiology of foot-and-mouth disease virus. *Virus Res* 91:65–80.
10. Konig GA, Palma EL, Maradei E, Piccone ME: 2007, Molecular epidemiology of foot-and-mouth disease virus types A and O isolated in Argentina during the 2000–2002 epizootic. *Vet Microbiol* 124:1–15.
11. Kumar S, Tamura K, Jakobsen IB, Nei M: 2001, MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 12:1244–1245.
12. Lunn DJ, Thomas A, Best N, Spiegelhalter D: 2000, WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10:325–337.
13. Martinez MA, Dopazo J, Hernandez J, et al.: 1992, Evolution of the 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.
14. Mohapatra JK, Sanyal A, Hemadri D, et al.: 2002, Sequence and phylogenetic analysis of the L and VP1 genes of foot-and-mouth disease virus serotype Asia1. *Virus Res* 87:107–118.
15. Perez A, Ward M, Carpenter T: 2004, Epidemiological investigations of the 2001 foot-and-mouth disease epidemic in Argentina. *Vet Rec* 154:777–782.
16. Perez AM, Ward MP, Carpenter TE: 2004, Control of a foot-and-mouth disease epidemic in Argentina. *Prev Vet Med* 65:217–226.
17. Saitou N, Nei M: 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425.
18. Samuel AR, Knowles NJ: 2001, Foot-and-mouth disease type O viruses exhibit genetically and geographically distinct evolutionary lineages (topotypes). *J Gen Virol* 82:609–621.
19. Spiegelhalter DJ, Best N, Carlin BP, van der Linde A: 2002, Bayesian measures of model complexity and fit (with discussion). *J R Stat Soc Ser B Stat Methodol* 64:583–639.
20. Steinhauer DA, Domingo E, Holland JJ: 1992, Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* 122:281–288.
21. Thompson JD, Gibson TJ, Plewniak F, et al.: 1997, The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882.
22. Villaverde A, Martinez MA, Sobrino F, et al.: 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.
23. Ward MP, Perez AM: 2004, Herd demographics correlated with the spatial distribution of a foot-and-mouth disease epidemic in Buenos Aires province, Argentina. *Prev Vet Med* 65:227–237.