Virology 382 (2008) 250–256

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

The tale of a modern animal plague: Tracing the evolutionary history and determining the time-scale for foot and mouth disease virus

Damien C. Tully, Mario A. Fares *

Smurfit Institute of Genetics, Trinity College Dublin, Ireland

ARTICLE INFO

Article history:
Received 29 July 2008
Returned to author for revision
17 August 2008
Accepted 7 September 2008
Available online 21 October 2008

Keywords:
Foot and mouth disease virus
FMDV evolution
Coalescent

ABSTRACT

Despite significant advances made in the understanding of its epidemiology, foot and mouth disease virus (FMDV) is among the most unexpected agricultural devastating plagues. While the disease manifests itself as seven immunologically distinct strains their origin, population dynamics, migration patterns and divergence times remain unknown. Herein we have assembled a comprehensive data set of gene sequences representing the global diversity of the disease and inferred the time-scale and evolutionary history for FMDV. Serotype-specific rates of evolution and divergence times were estimated using a Bayesian coalescent framework. We report that an ancient precursor FMDV gave rise to two major diversification events spanning a relatively short interval of time. This radiation event is estimated to have taken place towards the end of the 17th and the beginning of the 18th century giving us the present circulating Euro-Asiatic and South African viral strains. Furthermore our results hint that Europe acted as a possible hub for the disease from where it successfully dispersed elsewhere via exploration and trading routes.

Introduction

Past epidemics of bovine spongiform encephalopathy, recurrent outbreaks of avian influenza and the latest foot and mouth disease (FMD) outbreaks in the United Kingdom have all put an increasing focus and emphasis on the origin of these veterinary pathogens. In this regard, FMD is considered to be one of the worst animal plagues endemic in many parts of the world, responsible for severe economic devastation and as such deserves special attention. FMD is a vesicular disease of domesticated and wild cloven-hoofed animals caused by a small non-enveloped virus with a positive sense RNA genome. The virus is a member of the family *Picornaviridae*, which contains a number of important human pathogens, such as Poliovirus, Hepatitis A virus and the common cold Human rhinovirus A. This family is considered to be among the oldest and most diverse of known viruses. Such antigenic diversity in FMDV is reflected in the existence of seven immunologically distinct circulating serotypes known as O, A, C, Asia 1 and the South African Territories (SAT) 1, SAT 2 and SAT 3.

Due to its agricultural importance FMDV has been comprehensively dissected from the genetic point of view yet the rate of molecular evolution of this virus or the age of the sampled genetic diversity, reflected in the time to the most recent common ancestor (TMRCA), have been largely unexplored. Indeed the subject on the origin of FMDV and its respective serotypes prior to approximately 1900 remains completely unstudied. With this in mind, we have assembled a comprehensive dataset of gene sequences for all serotypes isolated at different times to infer the evolutionary time-scale and history of the disease.

The earliest commonly cited description of the disease dates as far back as 1514 to Northern Italy when an epizootic affecting only cattle was described (Fracastoro, 1546). Irrespective of the uncertainty associated with this early sighting it was not until the 17th and 18th centuries that we find trustworthy proof of its existence when FMDV affected cattle, sheep, pigs and goats from Western Europe. During the 19th century the disease had become widely diffused extending from the Caspian Sea to the Atlantic Ocean owing to the changing commercial relations between civilised countries. It was not until the early 1920s that the antigenic diversity of the disease was realised after observations by Vallée and Carré (1922). They showed the existence of two types designated by their areas of origin, O for the Department of Oise in France and A for Allemagne. Their work was later confirmed by Waldmann and Trautwein who discovered a third serotype termed C (Waldmann and Trautwein, 1926). In the 1940s, three additional serotypes were identified in Southern Africa and were designated accordingly as Southern African Territories types SAT1, SAT2 and SAT3 (Brooksby, 1958). The last and seventh serotype to be discovered was Asia 1, found in the early 1950s isolated from India in 1951 and 1952 (Dhanda et al., 1957) and Pakistan in 1954 (Brooksby and Rogers, 1957).

Although descriptions of different disorders have been recorded little or no reliance can be placed on this evidence. Moreover given that this was the very first animal virus to be discovered and is undoubtedly the most economically important veterinary pathogen, a hypothesis on the origin and subsequent diversification of the virus has not been subjected to rigorous examination using gene sequence data.

* Corresponding author.
E-mail addresses: dtully@tcd.ie (D.C. Tully), faresm@tcd.ie (M.A. Fares).

0042-6822/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2008.09.011

Thus, in order to fill these gaps in the evolutionary puzzle on FMDV and to gain a better understanding of the disease we inferred the evolutionary dynamics of the virus utilising a Bayesian MCMC framework. To achieve this, we assembled a dataset of 236 viral isolates representing all serotypes from a wide range of geographical localities and over an extensive time span. The gene analysed was VP1 because of the relatively large data sets available from previous studies (Tully and Fares, 2006). Furthermore, the VP1 gene forms part of the structural capsid of the virion and contains several of the major immunogenic sites important to effective antibody neutralization and subsequent viral clearance by the immune system. From this, we are able to infer for the very first time the time-scale and evolutionary history of FMDV and provide direct evidence of two major diversification events, within a relatively short time span, responsible for the present day circulating strains. Indeed, our estimates yield remarkable concordance with historical accounts of the disease and highlight the importance of gene sequence data into providing new insights on the origin and spread of infectious disease.

Results

The maximum clade credibility trees produced from MrBayes (not shown) and BEAST reveal that FMDV can be divided into two groups, with distinct viral lineages observed in Africa and in Europe/Asia.

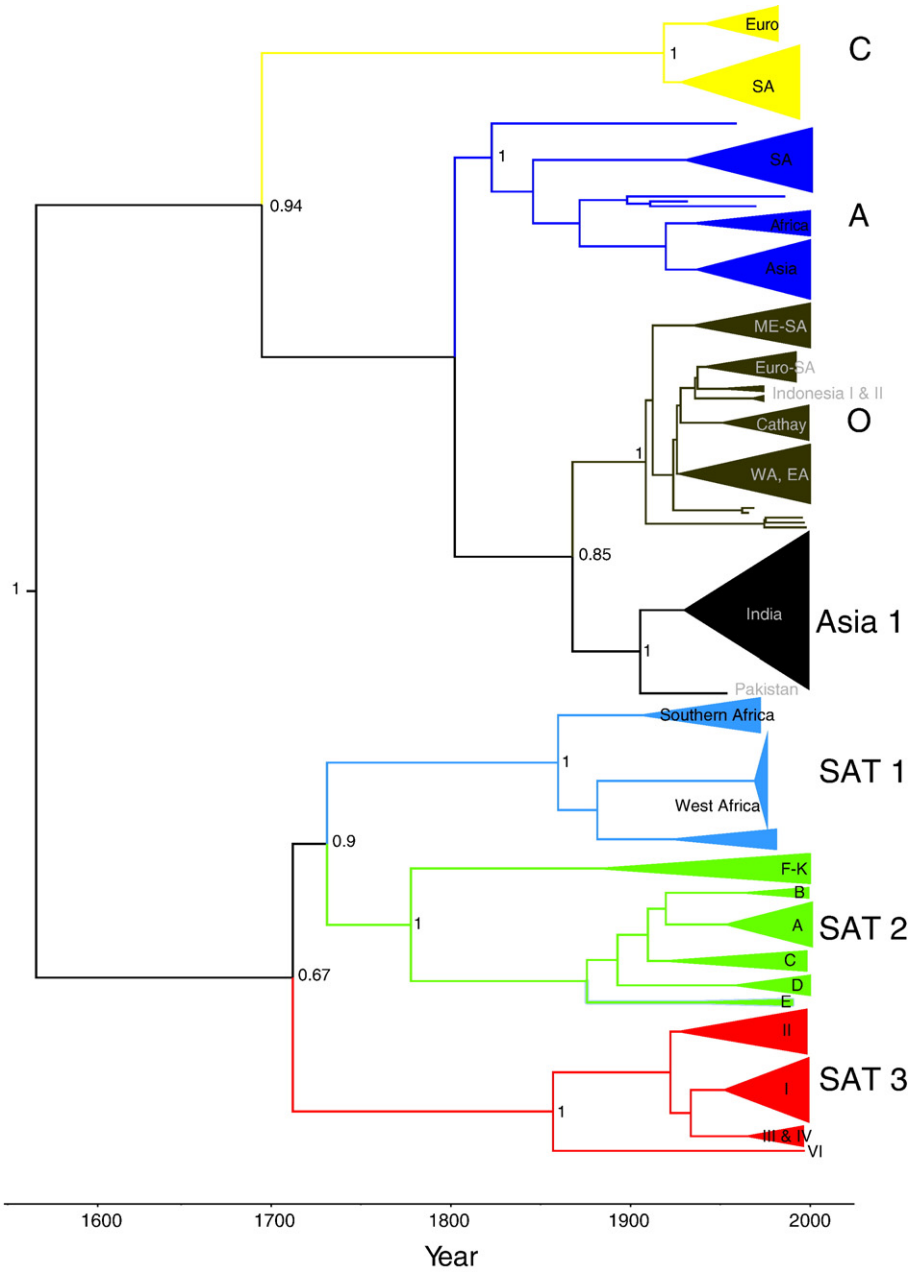

Fig. 1. Maximum clade credibility tree based on 236 FMDV gene sequences. The seven serotypes are schematically represented by different colour triangles and accordingly labelled with a timeline of divergence at the bottom. In addition, we show only at respective ancestral serotype nodes those posterior probabilities greater than 0.5. Further genetic diversity within serotypes is shown, in accordance with the FMD topotype concept. For consistency names are kept in agreement with previous designations. Serotype C, contains two distinct nodes comprised of viruses, of European and South American origin, termed Euro and SA, respectively. For type A there are three distinct clades comprising isolates from South America (SA), Africa and Asia. Type O also contains a number of distinct lineages named Europe–South America (Euro–SA), Middle East–South Asia (ME–SA), Cathay (an ancient name for China and east Tartary), west and East Africa (WA and EA). For SAT 2 viruses belonging within the South African region are designated A, C, D and K) while in West Africa two regionally distinct type are found (E and F) and the remaining viruses (B, G, H, J, K) belonging to East and Central Africa. Similarly, SAT 3 contains clades corresponding to geographically distinct regions with I–IV from Southern Africa while VI corresponds to East Africa.

Table 1
Bayesian estimates of divergence times and evolutionary parameters for FMDV serotypes

Data Set (n)	Date range of sequences	Demographic model	Substitution rates[a]	Age (Years)[a]	d_N/d_S	Recombination detected
A (37)[b]	1932–2001	Constant	4.26×10^{-3} (2.46×10^{-3} –6.16×10^{-3})	178 (78, 189)	0.155	Yes
O (47)[c]	1958–2000	Exponential	3.14×10^{-3} (1.84×10^{-3}–4.34×10^{-3})	92 (72, 173)	0.124	No
C (24)	1953–1994	Logistic	1.63×10^{-3} (6.17×10^{-4}–2.50×10^{-3})	82 (61, 145)	0.167	No
Asia 1 (34)	1954–1999	Constant	6.32×10^{-3} (3.39×10^{-3}–9.47×10^{-3})	96 (54, 161)	0.178	Yes
SAT 1 (32)	1962–1981	Constant	6.50×10^{-3} (2.03×10^{-3}–2.84×10^{-3})	141 (73, 255)	0.143	No
SAT 2 (32)	1948–2000	Logistic	1.07×10^{-3} (4.90×10^{-6}–1.14×10^{-3})	223 (87, 253)	0.105	Yes
SAT 3 (30)	1965–1999	Constant	2.58×10^{-3} (1.17×10^{-4}–5.59×10^{-3})	144 (76, 316)	0.156	No
All data (236)	1932–2001	Constant	2.48×10^{-3} (1.69×10^{-3}–3.31×10^{-3})	432 (218, 1250)	NA	NA

NA, not applicable.
[a] For all parameter estimates the 95% HPD values are given in parenthesis.
[b] Significant evidence for positive selection at codon 171under the single likelihood ancestor counting method.
[c] Significant evidence for positive selection at codons 48 and 139 under the single likelihood ancestor counting method.

Serotypes A, C, O and Asia 1 all cluster together and are supported by high posterior clade probabilities ($p>0.94$). Sequences from Africa also formed a distinct monophyletic cluster that diverged from the Euro-Asiatic serotypes, and further sub-divided into three sub-clades (SAT 1, 2 and 3) (Fig. 1). To evaluate the time-scale of the divergence of serotypes we examined evolutionary rates and dates of divergence using a Bayesian coalescent approach.

Estimation of nucleotide substitution rates

Broadly we observed equivalent rates of nucleotide substitution in all serotypes and across all 236 FMDV sequences. The mean nucleotide substitution rate was 2.48×10^{-3} substitutions per site, per year ranging from 1.07×10^{-3} (SAT 2) to 6.50×10^{-3} (SAT 1) substitution per site, per year (Table 1). To determine if these rates were affected by episodes of positive selection acting on specific codons or lineages we calculated d_N/d_S (also known as ω). Using the single likelihood ancestor counting (SLAC) method, available at the Datamonkey facility (Pond and Frost, 2005). The analysis revealed little evidence of positive selection, and an abundance of negatively selected sites. Overall, across all serotypes, we only found evidence of positive selection at a significance confidence level of $p<0.1$for three amino acid sites: one site for serotype A and two for serotype O. This number reduced to two sites when using the more stringent significance value of $p<0.05$.

All ω values were less than 0.2 indicating that most of the molecule has been undergoing strong purifying selection. Our results also show little if any recombination detected in the datasets (Table 1). Despite the occurrence of recombination in some of our alignments similar rates were observed in those serotypes where recombination was absent. Indeed, the mean substitution rate in SAT 1 was higher than any other serotype where recombination was present. In summary this supports previous studies that have demonstrated that recombination is largely constrained to non-structural genes with very few phylogenetic incongruities observed in the structural or capsid proteins (Carrillo et al., 2005; Jackson et al., 2007; van Rensburg et al., 2002). Thus, we reveal that neither evolutionary process has significantly biased our estimates thereby not violating the assumptions of coalescent analysis.

Divergence times of FMDV serotypes

From the Bayesian coalescent analysis, the deepest node on the FMDV phylogeny corresponded to a time of origin with a mean age of 432 years (95% HPD of 218–1250 years, Table 1). Our analyses resulted in the description of two primary clades that diverged from a common ancestral FMDV. The first clade comprises four sub-clades encompassing isolates from serotypes A, O, C and Asia 1 entitled the Euro-Asiatic clade. The estimated mean divergence time of the Euro-Asiatic clade was at approximately 306 years ago (95% HPD of 121 and

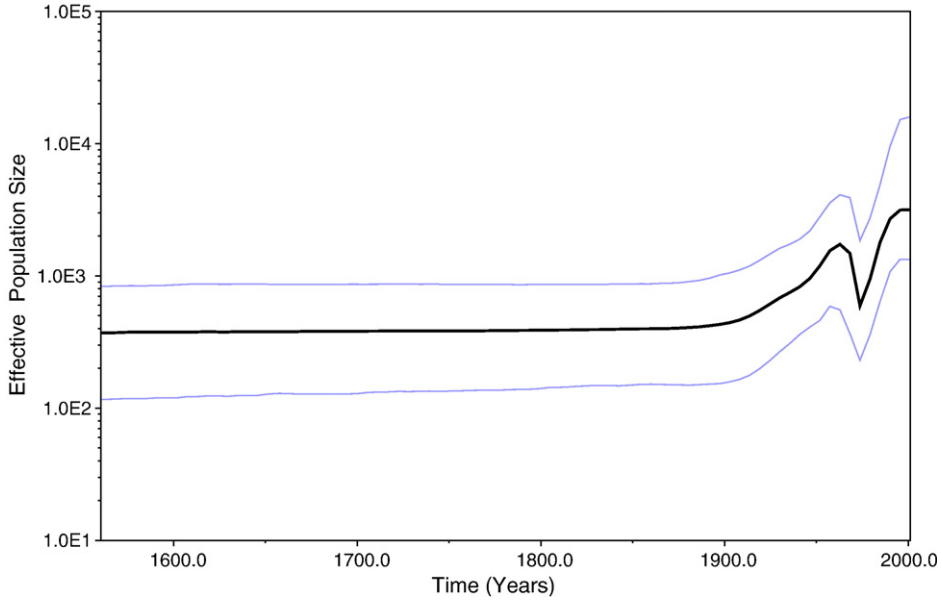

Fig. 2. Bayesian skyline plot estimated from the entire dataset of FMDV isolates sampled between 1932 and 2001. The bold line represents the median estimate of the effective number of infections through time. The blue lines indicate the upper and lower bounds of the 95% highest posterior density.

477 years, respectively) (See Fig. 1 and Table 1). The oldest estimated serotype within this primary clade is that of type A. The common ancestor of serotype A isolates existed around 1823 (95% HPD intervals of 1811–1923).

The second primary clade comprises those viruses indigenous to the African sub-continent known as the SAT type viruses. Similarly, with the Euro-Asiatic clade, the estimated mean divergence time of the African clade was approximately 289 years ago (95% HPD of 142–596 years).

Demographic history of FMDV

We found the best-fit model of population growth for each serotype under a relaxed molecular clock model allowing for rate variation implemented in BEAST v 1.4.7 (Table 1). We investigated four population growth models, constant, exponential, logistic and expansion growth. We calculated approximate marginal likelihoods of the four different demographic models. For serotypes A, Asia 1, SAT1 and SAT3 the Bayes Factor (BF) favoured a model assuming a constant population size. Although for serotype A the exponential population growth model had a slightly higher likelihood than the constant population size but given the standard errors associated with each model there was insufficient grounds to reject the latter model.

In contrast, serotypes C and SAT 2 show a significant logistic trend, in which an initially rapid growth phase is followed by a slowdown in the growth rate towards the present. While the BF was significant (log10 BF=5.04) for serotype O when an exponential model of population growth was assumed. The reconstruction of the demographic history of the most widely distributed serotype, type O revealed that the virus experienced a sharp increase after the 1920s until it reached its peak in the 1960s from where it started to decline gradually. In fact, this decline didn't last long (10–15 years) and the trend towards the present shows a modest expansion.

The Bayesian skyline plot analysis of FMDV isolates (Fig. 2) indicates that the onset of growth from the early 1900s and coincides with the appearance of a number of major clades in the phylogenetic tree. This graphical representation suggests that the virus grew rapidly until the 1970s, when it experienced a rapid sharp drop in population size followed by a sudden period of exponential population growth. However, comparisons of population dynamics models did not classify complex population models as significantly better to the data than a constant size population model (log10 BF<3.0).

Discussion

To our knowledge, this is the first study dating the divergence times of FMDV from the *Aphthovirus* genus within the *Picornaviridae* family using modern phylogenetic techniques. The use of such techniques has proved in the past to be extremely valuable when applied to a broad range of viruses. For instance, in reconstructing the emergence of HIV/AIDS in the Americas (Gilbert et al., 2007), in estimating the origin of Yellow Fever Virus into the Americas (Bryant et al., 2007), the origin of smallpox (Li et al., 2007), investigating the spread of Hepatitis C virus (Nakano et al., 2006; Njouom et al., 2006; Pybus et al., 2003; Verbeeck et al., 2006) and in deciphering the epidemiological dynamics of Influenza A virus (Rambaut et al., 2008). By performing a Bayesian coalescent analysis using serially sampled gene sequence data, we are able to provide important insights into the evolutionary and epidemiological characteristics of this veterinary pathogen.

Our coalescent analyses demonstrate that an ancient precursor of the virus gave rise to two early diversification events taken place within a relatively short span of time (approx 20 years). This deep bifurcation gave rise to the three South African Serotypes (SAT 1, SAT 2 and SAT 3) as well as to a lineage, which was the precursor of the present day Euro-Asiatic serotypes (A, C, O and Asia 1). This major event may reflect adaptive radiation, driven by ecological opportunity, where fast early diversification allows access to a greater diversity of resources (hosts). Therefore different clades could reflect independent adaptive radiations into different broad niches. Remarkably, the observation of separate diversification events for FMDV has been previously suggested albeit briefly using a simplistic phylogenetic approach (Dopazo et al., 1995). Although, until now this hypothesis has never been fully subjected to rigorous examination or even attempted to be dated using modern phylogenetic techniques.

The precursor of the current circulating strains of the different Euro-Asiatic serotypes underwent further diversification with three of the serotypes occurring more recently in the early 1900s. Serotype A has a much older ancestral origin dating back to approximately 1822 (or 178 years from the youngest tip of FMDV isolated in 2001). This older date of origin is of great significance as it suggests that the virus witnessed in the Americas in the late 19th century were type A because it is well conceived that European immigrants brought the disease into the continent.

Previous hypothesis have asserted imported cattle from European immigrants in the late 1860s or early 1870s as the alleged route of introduction of the disease into South America. Our results indicate that there is a strong genetic relationship between European strains and certain South American strains clearly vindicating this hypothesis. The strong European–South American relationship between type A viruses is well studied with a designated genetic lineage known as the Euro–SA topotype (Knowles and Samuel, 2003).

An almost simultaneous diversification event occurred shortly after the initial introduction of the disease into Africa giving rise to the formation of three clades appropriately named as the South African Territories (SAT 1, SAT 2 and SAT 3) where they have been maintained in sub-Saharan Africa ever since. The timing of such an event was estimated at approximately 289 years ago placing it at the start of the 18th century. At this point our estimates show us that the remaining clades (SAT 1 and SAT 2) had already diverged. Although both of these clades shared a similar time of divergence they do not appear to undergo rapid expansion with a constant population size best fitting the data. This constant population size would be consistent with a long-term persistent infection, a theme that is well studied within the SAT serotypes owing to the carrier state of African buffalo (*Syncerus caffer*) populations where FMDV can last for a number of years. The African Rinderpest pandemic that occurred at the end of the 19th century would have severely compromised FMDV diversity and during this time no reports of FMD in Southern Africa were reported. However, we believe that the SAT strains were maintained in isolated African buffalo herds but subsequent expansion did not occur till the early 1900s. Conversely, in the case of SAT 2 a logistic model of population growth rate best fits the data with coalescent times implying an earlier time of divergence around about 1778. Interestingly, SAT 2 is the serotype most often associated with outbreaks of foot-and-mouth disease (FMD) in livestock in southern and western Africa and is the only SAT type to have been recorded outside the African continent in the last decade with incursions made into the Middle East. Estimation of nucleotide substitution rates show that this serotype has the lowest substitution rate of $1.07\times10^{-3}$with a very wide range of 95% HPD values of between 4.90×10^{-6} and 1.14×10^{-3} reflected in its low ω value. Despite this wide range of estimates similar rates of nucleotide substitutions were observed from isolates in sub-Saharan Africa. All the substitution rates found in this study are consistent with previous estimates reported for FMDV and within the range reported for 50 RNA animal viruses (Jenkins et al., 2002) and of 49 other species from another study that included more slowly or rapidly evolving viruses (Hanada et al., 2004).

A very elegant older study examining the genetic diversification of serotype C over a 6-decade period attempted to date the origin of the ancestor of this serotype only to be met with considerable uncertainty (Martinez et al., 1992). But after omitting European viruses isolated

after a certain date a reasonable correlation was established with an extrapolated date of origin of 1897 with 95% confidence intervals of 1876 to 1912 (Martinez et al., 1992). Our estimates indicate a similar time of origin of 1918 with 95% HPD of 1856–1940, placing our mean date within 21 years of the previous study. Not only are our time estimates similar but also the rates of nucleotide substitution coincide, placing our mean estimates at 1.63×10^{-3} in comparison to 1.43×10^{-3} found in the former study.

As we show in our study, FMDV diverged from a recent common ancestor with two primary clades evolving into separate evolutionary paths and subsequently evolved in geographically discrete animal populations. Our coalescent analyses indicate that the divergence occurred within a short time span. But based on historical records and our dates of origin we speculate that the ancient precursor of the virus belonged to Europe. In fact, based on historical events we speculate that European exploration may have played a key role in the emergence of FMDV, not only to Africa and Asia but also in the New World (Fig. 3). It would also explain that the early observations of the disease in Northern Italy in the 16th century were indeed accurate reports of the virus.

Given the abundance of heterochronous sequences (i.e. sequences of viral genes isolated at different times) and the apparent infrequency of recombination makes this gene a feasible and appropriate candidate to reconstruct the evolutionary pathways of FMDV. Although selection has been previously reported within isolates (Fares et al., 2001; Haydon et al., 2001; Tully and Fares, 2006), this careful constructed dataset contains very little if any such selective pressure. In fact there is evidence of relatively strong purifying selection. Moreover, selection is not thought to significantly bias coalescent estimates (Grenfell et al., 2004; Hue et al., 2005; Lemey et al., 2003). However, recombination may severely bias estimates but the similar high substitution rates observed in those serotypes where recombination is thought to be absent in comparison to those serotypes where it is reported assures us that no differences are observed. This is validated by previous studies that have shown that this protein is not subjected to the same intensity of recombination as non-structural proteins are. Nevertheless, our ability to disentangle the reason why this pathogen emerged is hindered by ecological and genetic explanations that have been associated with anthropogenic factors such as changes in farming practices. Undoubtedly the increased level of trade between industrialised countries and other livestock movement patterns has had an effect in the spread of the disease. The emergence of the disease on the African continent was an important turning point in the history of FMDV where it is endemic with disease eradication being a grim prospect. Although the initial lack of restrictions on cattle traffic meant the disease was carried in all

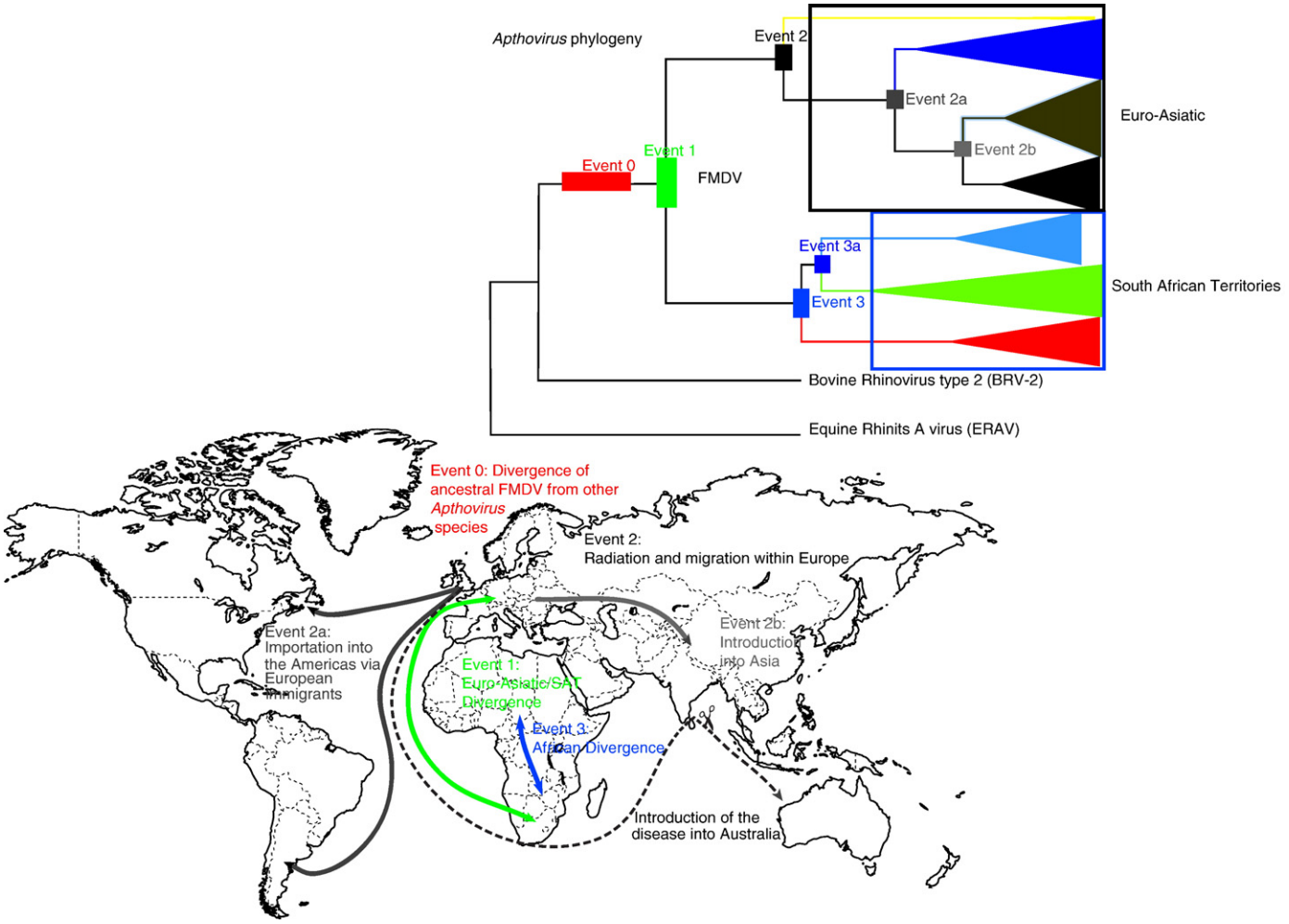

Fig. 3. Hypothesis on the spread of FMDV from ancient times. Event 0, the clustering of FMDV with other *Apthovirus* species suggests an ancient diversification event from a more recent common ancestor with BRV-2 than ERAV. Event 1, an ancestral precursor FMDV diverged into two primary clades designated the Euro-Asiatic serotypes and the South African Territories where they evolved independently. The dates and long relation of the disease in Europe suggests that the ancestral FMDV probably originated in Europe and later spread to Africa via infected animals or alternatively the divergence could have taken place in Africa. Event 2, FMDV diversification and migration throughout the Old World. Event 2a, European exploration brought FMDV into the Americas where it then spread. Event 2b, represents the diversification and migration of FMDV throughout Asia. Event 3, shows the diversification of the SATs on the African continent, followed by later introduction of additional sub-clades (Event 3a).

directions particularly by means of the African buffalo. The continual survival of this virus is likely driven by both ecological and evolutionary factors. This study adds pieces to solving the evolutionary origin of FMDV and of animal RNA viruses in general.

The implications of the results presented here will ultimately aid in our understanding of the genetic diversity between serotypes. For instance no previous study has convincingly accounted for the occurrence of lineage specific variation in rates of nucleotide substitution. The rate variation that exists between lineages within the phylogeny is suggestive of clade-specific epidemiology. In addition the reconstruction of the population history of FMDV serotypes provides indication of changing dynamics over the history of the lineage. This may give an indication of what serotypes are most likely to contribute to future epizootics. Appreciation of the underlying mutational dynamics and the ability of viruses to generate antigenic variation is undoubtedly important for epidemiological dynamics and vaccine development.

Materials and methods

Viral sequence data

We retrieved all viral sequences with known sampling dates from GenBank (Accession numbers in addition to dates of isolation of all viruses are provided in Table S1 of supplementary information). We constructed a protein sequence alignment using the program MUSCLE (Edgar, 2004) and then we built a corresponding protein-coding nucleotide alignment based on the concatenation of nucleotide triplets. We used the VP1 gene from viruses covering a sampling period of 69 years from 1932 to 2001 (Table S1). In order to maximise the evolutionary information contained among serotypes we used a sub-sampling procedure, with the view that the more diverse the viruses being compared the better we can account for the different evolutionary scenarios. This was necessary due to computational constraints associated with coalescent analysis that preclude us from using large number of sequences. The original compiled dataset from which we performed our sub-sampling comprised a total of 660 FMDV sequences. The final dataset sampled consisted of 236 sequences comprising all 7 serotypes of FMDV (n): SAT 1 (32), SAT 2 (32), SAT 3 (30), A (37), O (47), C (24) and Asia 1 (34) (See S1 for accession numbers). To perform unbiased (equilibrated sampling from serotypes) sampling from the set of 660 FMDV sequences, we used a "greedy" algorithm that maximises the evolutionary divergence between sequences in a given phylogeny (Pardi and Goldman, 2005). We chose not to randomly sub-sample sequences for each serotype, as this would increase the likelihood of selecting phylogenetically close sequences thereby decreasing the phylogenetic scope contained within the serotype. In order to ensure that this sub-sampling introduced no bias into the analysis and to make robust assumptions all coalescent estimates were performed on each complete dataset independently.

Sequence analysis

To determine the gene and site-specific selection pressures acting on each of the seven datasets of FMDV, we estimated the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions per site (d_N/d_S) using a maximum likelihood procedure available at the Datamonkey facility (Pond and Frost, 2005): the single likelihood ancestor counting (SLAC) method incorporating the GTR model of nucleotide substitution, with phylogenetic trees inferred using the neighbour-joining tree method.

In order to evaluate the extent of recombination we identified putative recombinant viruses in each of the datasets using the RDP3 package (http://darwin.uvigo.es/rdp.html) (Martin et al., 2005b) with the default thresholds. This package contains six recombination detection programs: RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), MaxChi (Smith, 1992), Chimeara (Posada and Crandall, 2001), Bootscan (Martin et al., 2005a) and SiScan (Gibbs et al., 2000). To exclude the possibility of false positive recombination detection, we considered only putative recombinant regions detected by at least three different programs.

Bayesian MCMC evolutionary analyses

We estimated the rates of nucleotide substitution, the age of the most recent common ancestor (MRCA) and changing profiles in demographic histories for each geographic serotype using a model that allows for rate variation among lineages under a relaxed (uncorrelated exponential) molecular clock (Drummond et al., 2006) as implemented in the Bayesian Markov chain Monte Carlo (MCMC) method available in BEAST version 1.4.6 (Drummond and Rambaut, 2007). We investigated four population models: constant population size, exponential population growth, logistic growth and expansion growth. In addition, we used the piecewise Bayesian skyline plot (BSP) to depict changes in genetic diversity over time. The BSP is a non-parametric method used for estimating past population dynamics through time without dependence on a pre-specified parametric model. In all cases we used sampling dates for each isolate as calibration points.

We based all our estimates on the GTR+I+Γ_4 model, with the frequency of each substitution type, proportion of invariant sites (I), and the gamma distribution of among-site rate variation with four rate categories (Γ4) estimated from the empirical data (parameter values available from the authors on request). In all cases statistically uncertainty in parameter values is reflected in the 95% highest probability density (HPD) values. Each MCMC analysis was ran for sufficient time to ensure convergence of all parameters (ESSs>100) with a discarded burn-in of 10% (as assessed using the Tracer program (version 1.4) (http://beast.bio.ed.ac.uk/). We calculated the Bayes factor (BF) to compare the performance of any two Bayesian models for the same dataset. This factor is the ratio of the marginal likelihoods with respect to the prior. This is a simple method that computes the BF via importance sampling using the harmonic mean of the sampled likelihoods (with the posterior as the importance distribution) (Suchard et al., 2001). A BF of >20, or a ln BF of >2.99, is defined as strong support for the favoured model. Although estimating marginal likelihoods is difficult so the estimates of BF outputted by Tracer v1.4 must be stated with caution. Consequently, we inherently took into account the error associated with calculating the BF using the harmonic mean approach.

Acknowledgments

We thank Dr Valentin Ruano Rubio for his assistance with the implementation of the greedy algorithm and fellow scientists for making their software freely available. This work was supported by Science Foundation Ireland.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2008.09.011.

References

Brooksby, J.B., 1958. The virus of foot-and-mouth disease. Adv. Virus Res. 5, 1–37.
Brooksby, J.B., Rogers, J., 1957. Methods used in typing the virus of foot-and-mouth disease at Pirbright, 1950–55. Methods of typing and cultivation of foot-and-mouth disease virus, European Productivity Agency of the Organisation for European Economic Co-operation, Project No208. Paris, pp. 31–34.
Bryant, J.E., Holmes, E.C., Barrett, A.D., 2007. Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. PLoS Pathog. 3, e75.

Carrillo, C., Tulman, E.R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G.F., Rock, D.L., 2005. Comparative genomics of foot-and-mouth disease virus. J. Virol. 79, 6487–6504.

Dhanda, M.R., Gopalakrishnan, V.R., Dhillon, H.S., 1957. Note on the occurrence of atypical strains of foot-and-mouth diseases virus in India. Ind. J. Vet. Sci. 27, 79–84.

Dopazo, J., M.J.R., Rodríguez, A., Sáiz, J.C., Sobrino, F., 1995. *Aphthovirus* evolution. Molecular Basis of Virus Evolution. Cambridge University Press, pp. 310–320.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic. Acids Res. 32, 1792–1797.

Fares, M.A., Moya, A., Escarmis, C., Baranowski, E., Domingo, E., Barrio, E., 2001. Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. Mol. Biol. Evol. 18, 10–21.

Fracastoro, G., (1546). *De Contagione et Contagiosis Morbis et eorum curatione*.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Gilbert, M.T., Rambaut, A., Wlasiuk, G., Spira, T.J., Pitchenik, A.E., Worobey, M., 2007. The emergence of HIV/AIDS in the Americas and beyond. Proc. Natl. Acad. Sci. U. S. A.

Grenfell, B.T., Pybus, O.G., Gog, J.R., Wood, J.L., Daly, J.M., Mumford, J.A., Holmes, E.C., 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303, 327–332.

Hanada, K., Suzuki, Y., Gojobori, T., 2004. A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes. Mol. Biol. Evol. 21, 1074–1080.

Haydon, D.T., Bastos, A.D., Knowles, N.J., Samuel, A.R., 2001. Evidence for positive selection in foot-and-mouth disease virus capsid genes from field isolates. Genetics 157, 7–15.

Hue, S., Pillay, D., Clewley, J.P., Pybus, O.G., 2005. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc. Natl. Acad. Sci. U. S. A. 102, 4425–4429.

Jackson, A.L., O'Neill, H., Maree, F., Blignaut, B., Carrillo, C., Rodriguez, L., Haydon, D.T., 2007. Mosaic structure of foot-and-mouth disease virus genomes. J. Gen. Virol. 88, 487–492.

Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. 54, 156–165.

Knowles, N.J., Samuel, A.R., 2003. Molecular epidemiology of foot-and-mouth disease virus. Virus Res. 91, 65–80.

Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., Vandamme, A.M., 2003. Tracing the origin and history of the HIV-2 epidemic. Proc. Natl. Acad. Sci. U. S. A. 100, 6588–6592.

Li, Y., Carroll, D.S., Gardner, S.N., Walsh, M.C., Vitalis, E.A., Damon, I.K., 2007. On the origin of smallpox: correlating variola phylogenics with historical smallpox records. Proc. Natl. Acad. Sci. U. S. A. 104, 15787–15792.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005a. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retroviruses 21, 98–102.

Martin, D.P., Williamson, C., Posada, D., 2005b. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260–262.

Martinez, M.A., Dopazo, J., Hernandez, J., Mateu, M.G., Sobrino, F., Domingo, E., Knowles, N.J., 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.

Nakano, T., Lu, L., He, Y., Fu, Y., Robertson, B.H., Pybus, O.G., 2006. Population genetic history of hepatitis C virus 1b infection in China. J. Gen. Virol. 87, 73–82.

Njouom, R., Nerrienet, E., Dubois, M., Lachenal, G., Rousset, D., Vessiere, A., Ayouba, A., Pasquier, C., Pouillot, R., 2006. The hepatitis C virus epidemic in Cameroon: genetic evidence for rapid transmission between 1920 and 1960. Infect. Genet. Evol 104 (47), 18566–18570.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Pardi, F., Goldman, N., 2005. Species choice for comparative genomics: being greedy works. PLoS Genet. 1, e71.

Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U. S. A. 98, 13757–13762.

Pybus, O.G., Drummond, A.J., Nakano, T., Robertson, B.H., Rambaut, A., 2003. The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol. Biol. Evol. 20, 381–387.

Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature 453, 615–619.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. 18, 1001–1013.

Tully, D.C., Fares, M.A., 2006. Unravelling selection shifts among Foot-and-Mouth Disease Virus serotypes. Evolutionary Bioinformatics 2, 237–251.

Vallée, H., Carré, H., 1922. Sur la pluralité des virus aphteux. CR Acad. Sci. Paris 174, 1498–1500.

van Rensburg, H., Haydon, D., Joubert, F., Bastos, A., Heath, L., Nel, L., 2002. Genetic heterogeneity in the foot-and-mouth disease virus Leader and 3C proteinases. Gene 289, 19–29.

Verbeeck, J., Maes, P., Lemey, P., et al., 2006. Investigating the origin and spread of hepatitis C virus genotype 5a. J. Virol. 80, 4220–4226.

Waldmann, O., Trautwein, K., 1926. Experimentelle Untersuchungen über die Pluralität des Maul-und Klauenseuchevirus. Berl. Tierärztl. Wschr. 42, 569–571.