Infection, Genetics and Evolution 32 (2015) 440–448

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

The impact of within-herd genetic variation upon inferred transmission trees for foot-and-mouth disease virus

Begoña Valdazo-González [a], Jan T. Kim [a], Samuel Soubeyrand [b], Jemma Wadsworth [a], Nick J. Knowles [a], Daniel T. Haydon [c], Donald P. King [a,*]

[a] The Pirbright Institute, Ash Road, Pirbright, Surrey GU24 0NF, United Kingdom
[b] INRA, UR546 Biostatistics and Spatial Processes, F-84914 Avignon, France
[c] Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form 5 March 2015
Accepted 26 March 2015
Available online 8 April 2015

Keywords:
Foot-and-mouth disease virus
Full genome
Within-herd genetic variation
Transmission trees

ABSTRACT

Full-genome sequences have been used to monitor the fine-scale dynamics of epidemics caused by RNA viruses. However, the ability of this approach to confidently reconstruct transmission trees is limited by the knowledge of the genetic diversity of viruses that exist within different epidemiological units. In order to address this question, this study investigated the variability of 45 foot-and-mouth disease virus (FMDV) genome sequences (from 33 animals) that were collected during 2007 from eight premises (10 different herds) in the United Kingdom. Bayesian and statistical parsimony analysis demonstrated that these sequences exhibited clustering which was consistent with a transmission scenario describing herd-to-herd spread of the virus. As an alternative to analysing all of the available samples in future epidemics, the impact of randomly selecting one sequence from each of these herds was used to assess cost-effective methods that might be used to infer transmission trees during FMD outbreaks. Using these approaches, 85% and 91% of the resulting topologies were either identical or differed by only one edge from a reference tree comprising all of the sequences generated within the outbreak. The sequence distances that accrued during sequential transmission events between epidemiological units was estimated to be 4.6 nucleotides, although the genetic variability between viruses recovered from chronic carrier animals was higher than between viruses from animals with acute-stage infection: an observation which poses challenges for the use of simple approaches to infer transmission trees. This study helps to develop strategies for sampling during FMD outbreaks, and provides data that will guide the development of further models to support control policies in the event of virus incursions into FMD free countries.

1. Introduction

The poor fidelity and lack of proofreading activity of the polymerases of RNA viruses cause high rates of spontaneous mutation during virus replication. These rates are estimated to range from 10^{-5} to 2×10^{-3} mutations per nucleotide per replication event (Thebaud et al., 2010). As a consequence, these viruses evolve rapidly and have high degrees of genome variability, which is a constant challenge for molecular diagnosic tests, as well as for prophylaxis and control methods such as vaccines and antivirals. Nevertheless, these high evolution rates have been exploited to understand the transmission of human and animal RNA virus infections across fine spatial and temporal scales (Cottam et al., 2006; Baillie et al., 2011; Bataille et al., 2011; Cotten et al., 2013; Gray et al., 2011; Hughes et al., 2012; Li et al., 2010; Orton et al., 2013). These studies help to increase the knowledge on virus evolution and to identify and analyse the potential origins, patterns of transmission and spread and risks of virus infections to be ready for the prediction, early detection and/or control of the disease.

Foot-and-mouth disease virus (FMDV) is a non-enveloped, single-stranded positive-sense RNA virus from the genus Aphthovirus within the family Picornaviridae which rapidly spreads among cloven-hoofed animals. Full genome sequences of FMDV can be generated and analysed in real-time to discern the origin of outbreaks, the transmission links between infected premises, and to predict undisclosed infection to support control and eradication policies in free-without-vaccination countries (Cottam et al., 2008b; Valdazo-González et al., 2012). Furthermore, these approaches have also been used to monitor the genetic evolution of FMD viruses at the finest scales: such as within an individual animal (Wright et al., 2011) and during animal-to-animal

* Corresponding author at: The Pirbright Institute, Ash Road, Pirbright, Woking, Surrey GU24 0NF, United Kingdom. Tel.: +44 (0)1483 231129; fax: +44 (0)1483 231142.
E-mail address: donald.king@pirbright.ac.uk (D.P. King).

http://dx.doi.org/10.1016/j.meegid.2015.03.032
1567-1348/

transmission in experimental studies (Juleff et al., 2013). The interpretation of these data can be enhanced by using a range of models that have been recently developed that integrate sequence data with epidemiological information (Cottam et al., 2008a; Morelli et al., 2012). However, the practical use of these tools to confidently reconstruct transmission trees during FMD outbreaks is limited by our understanding of the genetic diversity of viruses that exist within different epidemiological units under field conditions (within-herd diversity) (Orton et al., 2013).

This study has investigated the genetic variability of viruses from field samples collected from the FMDV outbreaks that occurred in the Southeast of the United Kingdom (UK) between the 3rd of August and the 30th of September 2007 (Cottam et al., 2008b; Ryan et al., 2008). FMDV sequences from the O/EURO-SA topotype were generated and analysed from samples within each of the eight infected premises (IPs) from 10 separate locations (with individual herds/flocks of animals grazing together) confirmed in a series of FMD outbreaks that occurred in two phases (that were geographically 17 km and temporally 34 days apart). Information regarding these herds and the clinical and laboratory investigations of these outbreaks has been described previously (Cottam et al., 2008b; Reid et al., 2009; Ryan et al., 2008). A particular focus of this work has been to consider the impact of sequencing only a single sample from each epidemiological unit upon the inferred transmission trees in order to help to design rapid and cost effective sequencing approaches that can be used in the event of FMD outbreaks, when sequencing all of the infected animals within the outbreak might not be possible.

2. Material and methods

2.1. Selection of samples

In total, 34 FMD virus-positive clinical samples from 26 animals infected during the 2007 outbreak in UK (Table 1) were processed in this study, and were jointly analysed with a further 11 previously published full-genome sequences from these outbreaks (Cottam et al., 2008b). These samples had been selected from the samples sent to the UK National Reference Laboratory for FMD (The Pirbright Institute, United Kingdom) during the 2007 outbreak in UK on basis of the cycle threshold (CT) values ($\leqslant$27) generated by a real-time RT-PCR which targets the region encoding the FMDV non-structural protein 3D (Reid et al., 2009). These samples included vesicular epithelium (10% suspension, prepared as described (Cottam et al., 2008b), whole blood (collected in EDTA tubes), sera and oesophageal/pharyngeal scrapings (probangs).

2.2. Full genome (FG) amplification and sequencing strategies

Samples were processed individually on separate days to prevent potential cross-contamination. Viral RNA was extracted using either the RNeasy Mini Kit (Qiagen, Crawley, West Sussex, UK), or TRIzol Reagent (Invitrogen, Paisley, UK) for those samples such as oesophageal-pharyngeal scrapings with high CT values. Reverse transcription (RT) and complete FMDV genome amplification [except for the poly(C) region] were performed with one oligo-dT reverse RT primer and 23 tagged PCR primers pairs as previously described (Cottam et al., 2008b), but using a cDNA purification step (Illustra GFX™ PCR DNA and Gel Band Purification Kit, GE Healthcare UK Limited, Buckinghamshire, UK) prior to PCR amplification. Additional PCR reactions were carried out using oligo dT reverse primers to amplify the 3′ terminus of the virus, as described (Valdazo-González et al., 2012). Negative control reactions were performed in parallel and were included in all steps and for each of the amplification reactions.

Amplified PCR products were separated by gel electrophoresis (1.8% agarose gels), stained with ethidium bromide (0.2–0.5 μg/mL) and visualized under ultraviolet light. After purification (Illustra GFX™ PCR DNA and Gel Band Purification Kit, GE Healthcare UK Limited, Buckinghamshire, UK), cycle sequencing was carried out using M13 universal forward and reverse primers (Cottam et al., 2008b) or the corresponding specific forward and reverse primers for each PCR product. One of the two following Sanger sequencing reagents and sequencers were used: the Beckman DTCS Kit (Beckman Coulter, USA) on a Beckman Coulter CEQ 8000 sequencer and the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) on an ABI PRISM®-3730 analyzer. Sequences were assembled, proof-read and edited using Lasergene® v11.0 package (DNASTAR Inc., Madison, WI). These sequences have been submitted to GenBank and have been assigned the following accession numbers: KJ560276–KJ560309.

2.3. Complete genome sequences of foot-and-mouth disease virus

The sequences generated in this study were aligned (BioEdit, Version 7.0.5.3 (Hall, 1999)) together with 11 previously published sequences from this outbreak (Cottam et al., 2008b). In total, 45 complete genome sequences from samples from 33 animals (28 cattle and 5 sheep) from eight infected premises (IPs) (ten separate herds) were analysed. This analysis comprised one to five animals and up to eight sequences per herd. Eleven out of these 33 animals were represented by two or three sequences obtained from different clinical samples within the animal (see Table 1).

2.4. Positive selection and recombination analysis

Detection of potential selection pressures at amino acid sites was calculated using the CODEML programme in the PAML 4.1 software package (Yang, 2007). Briefly, the dN/dS ratio (ω value) was obtained using eight different models (M0 to M8). Comparison of likelihood values for nested models by likelihood ratio tests (LRTs) determined if models of positive selection (M2a, M3 and M8) were significantly more likely than models of neutral evolution (M1a and M7). Bayesian methods were used to locate specific sites that have $\omega > 1$ with high posterior probabilities. Preliminary data for the analysis (transition/transversion ratio and phylogenetic relationship between taxa) were carried out using TREE-PUZZLE version 5.2. (Schmidt et al., 2002). Detection of potential recombination between sequences was carried out using low linkage disequilibrium (observed data versus corresponding null distributions from 500 simulated datasets) as implemented in a test statistic, as described (Haydon et al., 2004).

2.5. Bayesian Markov chain Monte Carlo (MCMC) analysis (BEAST) analysis

Bayesian evolutionary analysis using Markov chain Monte Carlo (MCMC) sampling (30,000 trees from 30 million generations), as implemented using BEAST software, Version 1.6.1 (Drummond and Rambaut, 2007), was carried out to infer the phylogenetic relationships between the 45 complete sequences, to estimate the age of their most recent common ancestor (MRCA) and their rate of molecular evolution. Sampling collection dates were used to calibrate the molecular clock. The HKY model of base substitution with the gamma model of site heterogeneity was selected as described (Orton et al., 2013) and used with different combinations of molecular clocks, demographic and phylogeographic diffusion models to check the robustness of the parameters. The resulting output was checked in Tracer, Version 1.5 and visualized with FigTree (Rambaut, 2010), Version 1.3.1 (Lemey et al., 2010).