



Molecular Epidemiology Questions Transmission Pathways Identified During the Year 2000 Outbreak of Classical Swine Fever in the UK

Rebecca Strong, Stephen McCleary, Sylvia Grierson, Bhudipa Choudhury, Falko Steinbach and Helen R. Crooke*

Virology Department, Animal and Plant Health Agency, APHA-Weybridge, Addlestone, United Kingdom

OPEN ACCESS

Edited by:

Matheus Nunes Weber,
Feevale University, Brazil

Reviewed by:

Erpeng Zhu,
Guizhou University, China
Fernando R. Spilki,
Feevale University, Brazil

*Correspondence:

Helen R. Crooke
Helen.Crooke@apha.gov.uk

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 31 March 2022

Accepted: 06 May 2022

Published: 28 June 2022

Citation:

Strong R, McCleary S, Grierson S,
Choudhury B, Steinbach F and
Crooke HR (2022) Molecular
Epidemiology Questions Transmission
Pathways Identified During the Year
2000 Outbreak of Classical Swine
Fever in the UK.
Front. Microbiol. 13:909396.
doi: 10.3389/fmicb.2022.909396

The last outbreak of classical swine fever (CSF) in the UK occurred in 2000. A total of 16 domestic pig holdings in the East Anglia region were confirmed as infected over a 3-month period. Obtaining viral genome sequences has since become easier and more cost-effective and has accordingly been applied to trace viral transmission events for a variety of viruses. The rate of genetic evolution varies for different viruses and is influenced by different transmission events, which will vary according to the epidemiology of an outbreak. To examine if genetic changes over the course of any future CSF outbreak would occur to supplement epidemiological investigations and help to track virus movements, the E2 gene and full genome of the virus present in archived tonsil samples from 14 of these infected premises were sequenced. Insufficient changes occurred in the full E2 gene to discriminate between the viruses from the different premises. In contrast, between 5 and 14 nucleotide changes were detected between the genome sequence of the virus from the presumed index case and the sequences from the other 13 infected premises. Phylogenetic analysis of these full CSFV genome sequences identified clusters of closely related viruses that allowed to corroborate some of the transmission pathways inferred by epidemiological investigations at the time. However, other sequences were more distinct and raised questions about the virus transmission routes previously implicated. We are thus confident that in future outbreaks, real-time monitoring of the outbreak *via* full genome sequencing will be beneficial.

Keywords: classical swine fever virus, pigs, full genome sequencing, molecular epidemiology, outbreak transmission pathways

INTRODUCTION

Classical swine fever (CSF) is a contagious febrile disease of pigs that remains a severe socio-economic threat to the global pig industry. The virus has been eradicated from many developed countries, but it continues to cause economic losses in endemic regions. The reintroduction of the virus into disease-free areas has a major impact, as highlighted by the re-emergence of the disease in Japan in 2018 after 26 years of freedom (Ganges et al., 2020; Shimizu et al., 2021). The virus is transmitted directly between pigs, either by vertical or horizontal transmission, including by contact with wild boar, or indirectly by contaminated food, objects, or people. Timely interventions

can rapidly identify infected premises, reduce the spread and length of an outbreak, provide a significant economic and welfare benefit, and facilitate the earlier restoration of trade in pig meat and products (Postel et al., 2018).

The last outbreak of CSF in the UK occurred in 2000 in the East Anglia region, a high pig dense area that was not known to have a feral pig population. A total of 264 premises were investigated for CSF suspicion although only 16 premises (designated IP01 to IP16) were confirmed as infected over a 3-month period. The outbreak was controlled by culling of infected and contact premises and vaccination was not applied. The first case (IP01) was confirmed on 8 August 2000. However, substantial illness, attributed to porcine circovirus (PCV-2)-associated disease, had been present since 11 July when weaner animals were introduced from an outdoor breeder unit (IP02) which was, at the time, accordingly thought to have been the origin of the outbreak (Figure 1). A routine veterinary visit of IP02 on 19th June did not find evidence of clinical disease. Signs noted were attributed to heat stress, and production records and mortality rates did not give cause for concern. However, it is thought that the disease was present on this premise since at least from early June (Mackinnon, 2001). Records indicated that the first signs of illness could have been found in a sow which was recorded as being unwell on 20 June and died on 25 July. Weaners from IP02 were sold to a large pig producing business that owned numerous farms, including IP01, IP03, IP04, and IP05. A batch of weaners was transferred from IP02 on 27 June to a rearing unit in Essex (IP03). IP04 and IP05, located in Suffolk and Norfolk, received animals from IP02 on 18 July and 4 July, respectively (Paton, 2002).

In early September, a cluster of further cases was detected in the south of Suffolk, close to IP01. IP07, an outdoor breeding herd with land adjacent to IP01, was confirmed as infected on 5 September. IP08, which had 5,000 growers and finishers, lies to the north of IP07 and had outdoor growers housed on land contiguous to IP07. IP09 also had outdoor sows, gilts, and boars located close to IP01. It was concluded that lateral local spread infected these premises, possibly due to the scavenging of dead pigs from IP01 by wild animals before CSF was confirmed (Mackinnon, 2001; Paton, 2002). A feature leading to delays in the detection of disease, especially in outdoor breeding units, was the insidious nature of clinical signs induced by the moderately virulent strain in adult pigs.

Two distantly located premises, IP10 and IP11, were confirmed on the 10th and 11th of September. Both these north Norfolk premises had received pigs from IP07; IP10 received weaner pigs on 8 August, whereas IP11 received breeding animals in both July and August. Pigs were sent from IP11 to IP15, which was later diagnosed on serological evidence. There was less information on potential transmission routes into the remaining infected premises; IP06, IP13, IP14, and IP16 were all located within 10 km of IP5, and IP12 is located close to IP04, suggesting that local spread may have occurred.

Genetic analysis of virus genomes can help us to understand disease transmission pathways to inform effective disease control policy decisions. However, for genetic data to be useful for molecular epidemiology, sufficient genetic diversity needs to

occur over a relevant epidemiological timescale. Traditionally, parts of the 5'NTR, E2, and sometimes the NS5B encoding regions of the CSFV genome, are sequenced to determine the CSFV genotype (Paton et al., 2000). These target regions are easily obtained and provide rapid information on the potential source of a newly introduced virus. However, the short sequence length does not allow discrimination between closely related viruses and is of limited use to trace virus movements within an outbreak. This led Postel and colleagues to propose an improved phylogenetic strategy, based on the analysis of the full-length E2 sequences (Postel et al., 2012). In the past decade, sequencing technologies have advanced such that obtaining full-length sequence of viral genomes is equally cost-efficient and rapid. This enhances the potential for full genome analysis to help inform about transmission events occurring during an outbreak.

To gain an understanding of the utility of generating full genome sequence data during a CSF outbreak, we obtained sequences of the E2 gene and viral genomes from archived tissue samples. Phylogenetic analysis of full genome data supported many, but not all, of the likely transmission pathways identified by epidemiological investigations at the time. The data also reveal additional details to the transmission pathways discovered through field epidemiology at the time.

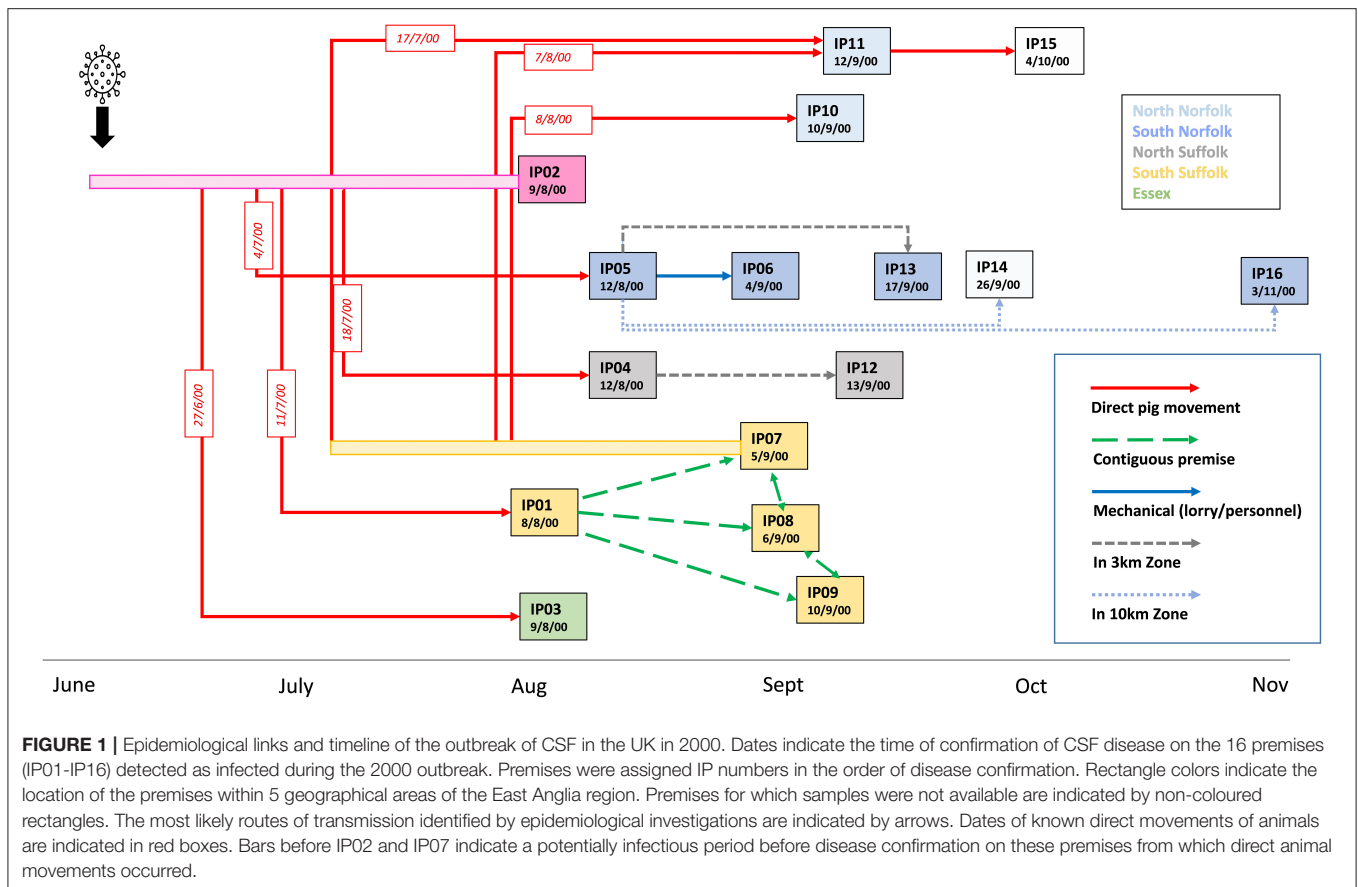
MATERIALS AND METHODS

RNA Extraction From Tonsil

Archived tonsil samples that had been stored at -80°C were available from individual animals from premises IP01-IP05 and IP16, whereas samples from premises IP06-IP13 were stored as pools of tissues from up to 4 pigs. A pooled sample (IP01-7.2) was also available from IP01. No samples were available from IP14 and IP15. To maximize the proportion of sequence data derived from the viral genome, as opposed to contaminating host nucleic acid, tonsil samples ($\sim 3\text{--}5\text{ mm}^3$) were added to 300 μl of lysis buffer (5 mM MgCl_2 , 50 mM Trizma, in PBS) containing 5 U/ μl of the endonuclease Omnicleave (Cambio). After incubation at room temperature for 30 min, the samples were homogenized in a Precellys 24 homogenizer in tubes with ceramic beads (CR28) with 3 cycles of 5,000 rpm for 45 s, with 30 s pauses between each cycle. Alternatively, the samples were homogenized using a GentleMax homogenizer and dissociated for 54 s using the manufacturer's program RNA_01.01. The homogenized samples were then clarified by centrifugation at $14,000 \times g$ for 2 min, the supernatant was mixed with 1 ml of the Trizol reagent (ThermoFisher), and RNA was extracted according to the manufacturer's instructions.

Sequencing of E2 and the Full Coding Region of the CSFV Genome

The full E2 coding region was amplified from one sample from each premise, except for IP06, with Phusion polymerase (ThermoFisher) with primers 2250F and 3710R (Postel et al., 2012) after preparation of cDNA with thermoscript III reverse transcriptase (Fisher scientific). The DNA sequence was obtained by Sanger sequencing using primers 3010F, 2919F, and 3000R as described (Postel et al., 2012). Almost full-length viral



genome sequences were obtained by preparation of cDNA with a Superscript IV first-strand synthesis system (ThermoFisher) followed by NEBNext® Ultra II non-directional RNA second-strand synthesis (NEB). To prepare a sequencing library, 1 ng of cDNA was added to the Nextera XT library kit and the samples were subsequently sequenced with an Illumina MiSeq. All paired-end sequence reads were trimmed using Trimmomatic (v0.39) (Bolger et al., 2014). Adapter sequences were removed and the headcrop option was used to remove 16 bases from the start of reads. Sliding window size and quality settings of 5:20 and minimum length trimming of 40 were applied. The CSFV genome sequence was obtained by *de novo* assembly (IVA v1.0.8 using default settings) (Hunt et al., 2015) of the sequence from IP02, the assumed index case, and checked by reference guided mapping against itself (BWA-MEM) v0.7.17) (Li and Durbin, 2009). The SF02 sequence was then used as a reference sequence for reference guided mapping of the remaining sequences. The sequences obtained from IP01–IP16 were designated SF01–SF16, respectively. The sequence from the pooled sample from IP01 (IP01-7.2) was designated SF01-7.2. Consensus sequences were called using iVar (Grubaugh et al., 2019a) with a minimum quality score of 20 and a depth of 1 and a frequency threshold of 0.6. A small region in the 3'UTR that was not obtained by Illumina sequencing for the sample from IP12 was amplified by

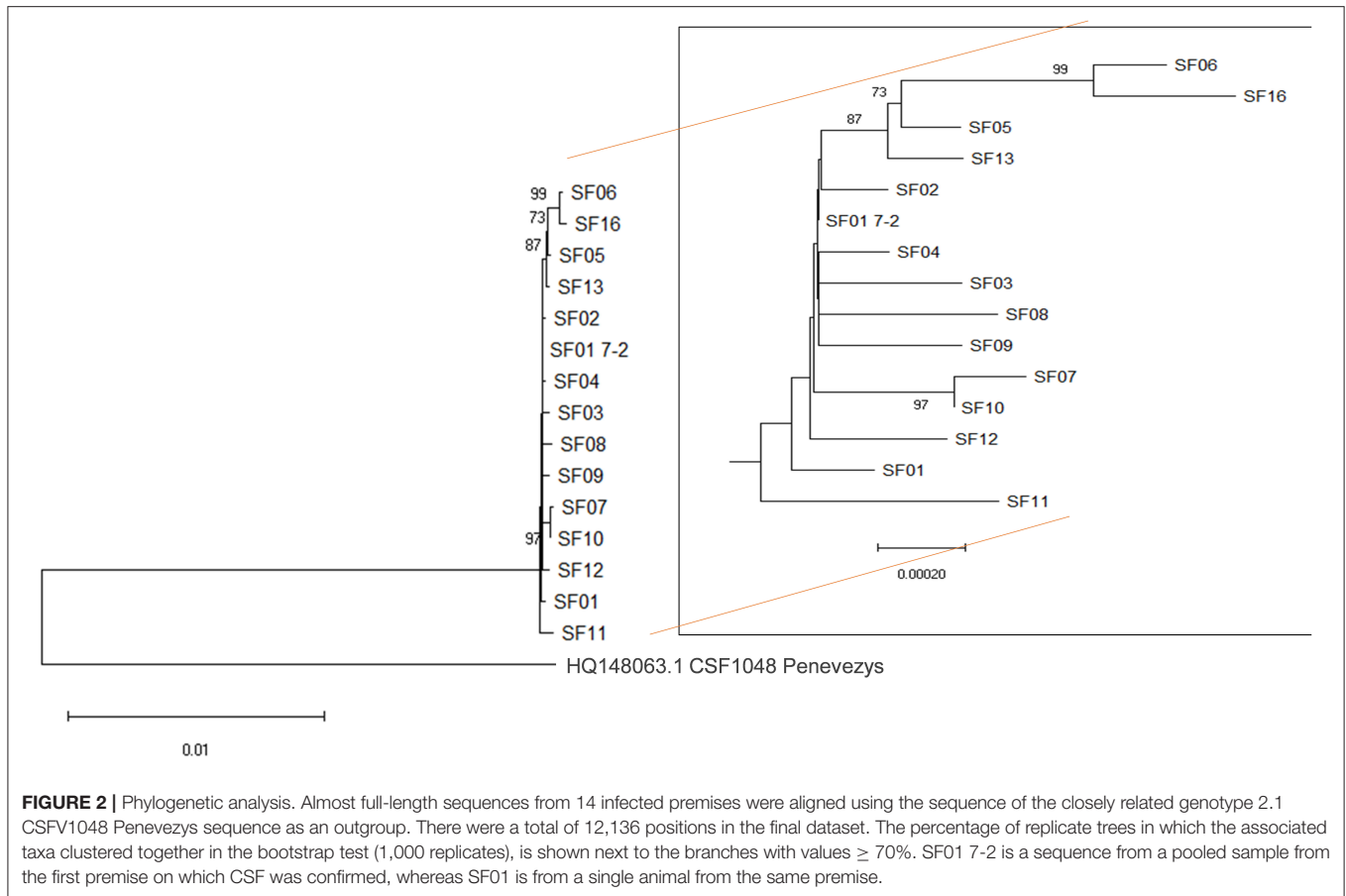
PCR using primers 5' GCAAGACGGGAAATAGGTAC 3' and 5' TTTTCCCTCCAGCTAAAGTG 3' and the resulting product was sequenced by Sanger sequencing. The resultant sequences were submitted to DDBJ.

Phylogenetic Analysis

Sequences were aligned using MegaX using the Clustal W algorithm. The evolutionary history was inferred using the Neighbor-Joining method. The bootstrap test was applied to infer confidence values; the percentage of trees in which the associated taxa clustered together in 1,000 replicates was expressed as a percentage, with values $\geq 70\%$ displayed at the tree nodes. The evolutionary distances were computed using the Maximum Composite Likelihood method. The sequence of the genotype 2.1 Penevezys strain CSF1048, isolated in Lithuania in 2009, was used as an outgroup (Genbank accession HQ148063).

RESULTS

The analysis of the complete E2 gene indicated that most of the samples contained viruses with identical E2 sequences which, in turn, were also identical to the previously published sequence of the virus isolate CSF0708 obtained from IP03 (JQ441582). However, the samples from IP08 and IP12, which were confirmed



positive in September, and IP16, which was the last premise detected 3 months after the first case, had viruses with one (IP12 and IP16) or two nucleotide changes (IP08) compared to the sequence from IP02. Accordingly, the use of the E2 gene would not suffice for in-depth molecular epidemiological approaches (data not shown).

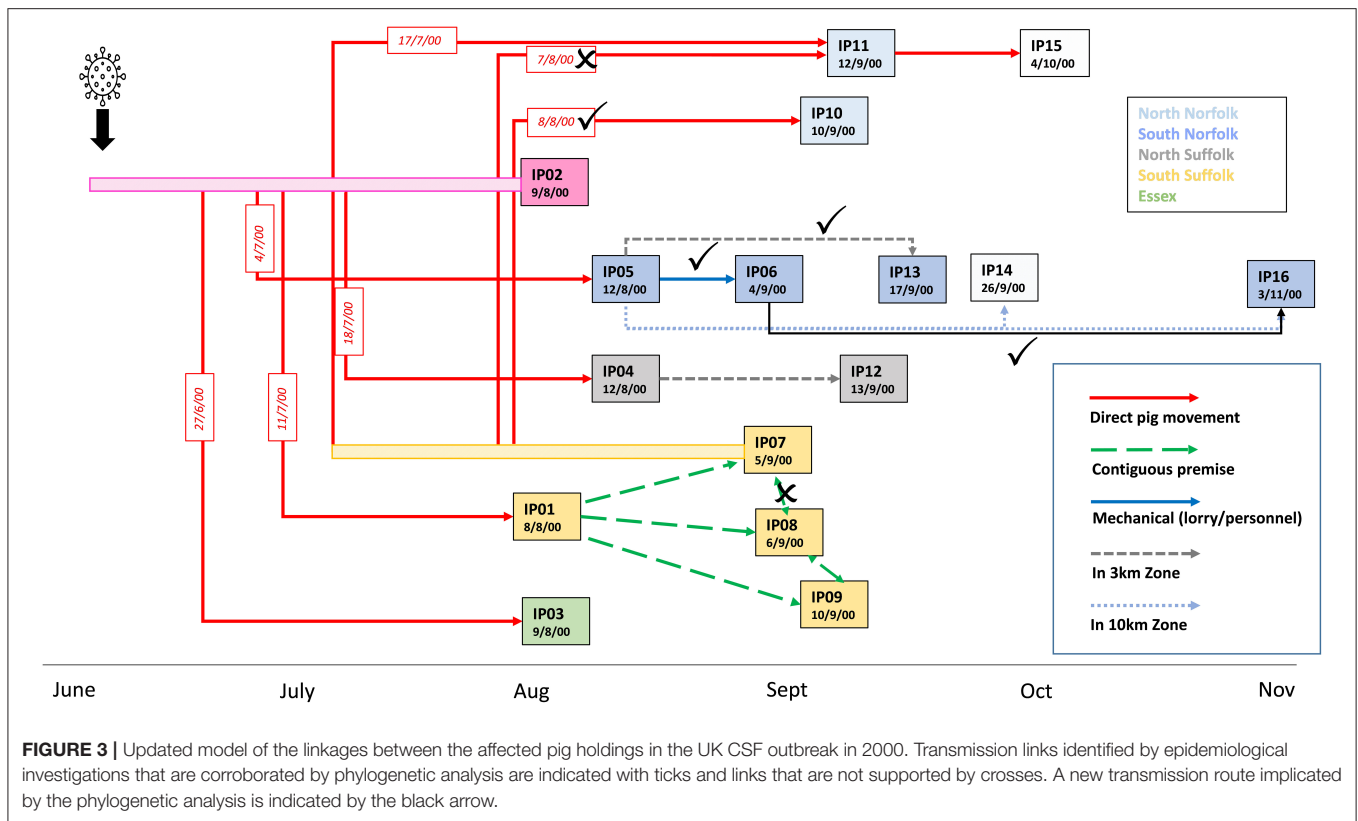
Near full-length sequences were obtained for all 15 samples: 12,136 nucleotides encoding the polyprotein region and 349/102 nucleotides of the 5'/3' UTRs, respectively (DDBJ accession numbers LC702327–LC702341). The changes in the E2 gene in SF12, SF16, and SF08 identified by sequencing of PCR product were present in the Illumina generated sequences, confirming that they were not the result of a Phusion polymerase error.

The alignment of the sequences encoding the almost full-length genome confirmed that the viruses in all 15 samples were highly similar ($>99.8\%$ identical), with only 5–14 nucleotide changes compared to the sequence from IP02, which was taken as the baseline as it was considered to be the first infected premise (**Supplementary Table 1**). Nucleotide changes were present in all regions of the genome, except for the regions encoding the core and NS4A proteins. The changes in the sequence SF01-7.2 from the pooled sample were degenerate, indicating that animals contributing to the pool may have been carrying slightly different viruses. Ten non-synonymous changes were identified but, as these were

not shared in sequences from different premises, there was no evidence that these amino acid changes imparted a competitive advantage.

Although virus genome sequences from all infected premises were highly similar, the phylogenetic analysis revealed succinct clusters between the viruses, with sequences from IP05, IP13 IP06, and IP16 forming a separate group, and IP06 and IP16 forming a subgroup therein (**Figure 2**). In addition, the sequences obtained from IP07 and IP10 were distinct from the main group but were closely related to each other. Furthermore, the sequence from IP11 was distinct from all others. These phylogenetic clusters, which are supported by high bootstrap values ($>80\%$), were not influenced by the nature of the samples (pooled vs. individual animals). In this context, it is interesting to note that there were 9 nucleotide differences between the sequences obtained from the sample from an individual animal from IP01, and the pooled sample (IP01-7.2). The SF01 sequence from the individual animal sits slightly outside the main sample group formed by sequences SF01-04, 08, and 09 but is not sufficient to form a bootstrap-supported group. It is, perhaps, not surprising that more changes are apparent in the sequence from an individual animal as changes in only one animal may be masked within the consensus sequence of a pooled sample.

Phylogenetic analysis of the sequences (SF01-SF16) obtained from the different infected premises highlighted clusters of



related sequences that were supportive of virus transmission routes suggested by epidemiological investigations (Figure 3). The sequences obtained from IP05, IP06, IP13, and IP16, all located in the south Norfolk area, support the hypothesis that virus spread by local transmission within the 10-km zone surrounding IP05. The sequence data suggest that IP06 was the most likely source of virus infecting IP16, which was the last premise detected, although IP14, for which samples were not available, may also have been involved in this transmission chain. The close relatedness of viruses found in IP07 and IP10 also strongly implicates the transfer of weaner pigs from IP07 to IP10 as being the source of virus introduction to IP10.

In contrast, although a direct transfer of animals between IP07 and IP11 occurred on 7 August, the sequence data indicate that the viruses on the two premises are diverse and an alternative route of spread to IP11 appears more likely. For example, the virus may have been transmitted earlier *via* the transport of breeding sows on 17 July from IP07, which would also explain the higher degree of virus evolution and outside clustering of sequence SF11, or transmission could have occurred *via* another unidentified contact with any of the earlier infected premises. Sequence information was less informative for clarifying virus transmission routes within the south Suffolk cluster of farms, IP01, IP07, IP08, and IP09, although the molecular analysis is not inconsistent with the theory of lateral spread by wildlife scavenging on carcasses left on IP01 prior to disease confirmation (Mackinnon, 2001). With SF07 being part of a genetic subgroup, the idea that there was a transmission from IP07 to the adjacent

IP08 is not corroborated by the molecular analysis. Similarly, the sequence data from IP04 and IP12 do not strongly support the theory that IP12 was infected directly by the virus from the closely located IP04. What also cannot be resolved further by this molecular approach are transmissions between the premises IP01-04, 08, 09, and 12 and how the infection spreads from there to the two premises forming cluster SF07/10, or in fact which premise was directly ancestral to SF11 or the spread to the farms with viruses in phylogenetic cluster SF05/13/06/16.

DISCUSSION

Molecular epidemiology has long been established as a useful addition to the interrogative epidemiology used to describe the movements of animals and people. There is a significant difference, however, between the global analysis of strain evolution and movements, such as has been done for influenza or SARS-CoV-2 in recent years (Petrova and Russell, 2018; COVID-19 Genomics UK (COG-UK), 2020) and the detailed analysis of a localized outbreak, such as the CSF outbreak in the UK in the year 2000. To achieve a high resolution with certainty for such localized outbreaks, a minimum number of sequence changes need to be present to obtain a phylogenetic resolution. In addition, the resources available for control of animal viruses are limited compared to human viruses, such as influenza and SARS-CoV-2, especially in developing regions, thus it is important to know if efforts to gain sequence data will provide useful information.

Analysis of the full-length E2 gene, which encodes the major immunogenic glycoprotein, has been applied to study the epidemiology of CSFV and provides useful discrimination between related strains circulating over a longer period within endemic regions (Jiang et al., 2013; Postel et al., 2013). Obtaining this 1,119 nucleotide sequence by PCR and Sanger sequencing is a straightforward and efficient method to study strains isolated over a prolonged period and/or between geographical regions. However, in this study, insufficient genetic diversity was found within the E2 coding region to assist in tracing virus movements over the relatively short time frame of this outbreak. Indeed, the almost full-length genome sequences contained a maximum of 14 nucleotide changes only, indicating that as much sequence information as possible will be beneficial. Different sequences from different animals on the same premise would be indicative of the time since the incursion to the farm, but multiple incursions (e.g., through multiple infected animals) also need to be considered. To delineate true differences over a slightly longer period a farm consensus sequence might be beneficial. The data analyzed in this study indicate that significant evolutionary changes that lead to the formation of new clusters can be detected through both approaches, but further studies will be required to resolve the mutation rates of CSFV per pig against time.

The relatively high level of differences found between the sequence from a single animal on IP01 and the sequence of pooled samples from the same farm likely reflect the substantial circulation of the virus within this farm, where a diagnosis of PCV-2-associated disease was initially made, and CSF was not suspected for weeks until about one-third of the 3,600 pigs were sick and 200 had died (Paton, 2002). Similarly, we can conclude a rather early transmission of the CSFV to farm IP11, even though the origin of transmission cannot be resolved further at this moment.

Sequencing of full virus genomes is becoming faster and increasingly cost-effective and its utility for large-scale molecular epidemiology has been demonstrated for numerous viruses (Grubaugh et al., 2019b). A particularly pertinent example of tracing a single outbreak was the molecular epidemiology of the foot and mouth disease virus (FMDV) outbreaks in 2001 and 2007 in the UK, which have been studied both retrospectively and in real time (Cottam et al., 2008a,b). The real-time analysis of FMDV genomes from the 2007 outbreak identified 9 nucleotide changes in the sequence between the viruses detected in early infected premises and a later cluster of infection that occurred 5 weeks after the initial premises were culled. This number of changes was higher than the predicted mean number of changes between premises and inferred the potential for the existence of intermediate, undetected, infected premises. Field investigations subsequently identified an infected premise that was the likely link between the two clusters.

However, interpreting the significance of small numbers of nucleotide changes is difficult and there are substantial differences in the rate of accumulation of genetic mutations between viruses (Campbell et al., 2018) and, for example, FMDV has a slightly higher mutation rate than CSFV (Freimanis et al., 2016; Rios et al., 2017). Epidemiological situations also vary and may impact virus evolution. For example, the presence of wildlife virus reservoirs, other susceptible species, length of

the outbreak, number of infected individuals or premises, and transmission routes will vary vastly and hence no two outbreaks, even of the same disease, will be the same. Knowledge on the extent and rate of virus evolution for different viruses under different circumstances is, therefore, important to assist in the interpretation of the significance of genetic changes.

The CSF outbreaks from various epidemiological situations have been analyzed by full genome sequencing. Leifer and colleagues sequenced isolates from wild boar in Germany in a region close to a previous vaccination zone (Leifer et al., 2010). Comparing the genome sequences of two viruses isolated in 2009 to the sequence of a putative ancestor strain, isolated 4 years earlier, inferred that, despite intense surveillance of this wildlife population, the virus had persisted undetected. Analysis of 11 genomes from viruses isolated over a 5-year period from wild boar in northern France indicated that the virus evolved relatively slowly, with an estimated substitution rate of 1.5×10^{-3} substitutions per site per year (Goller et al., 2016). This rate of evolution was sufficient to discriminate two spatial and temporal clusters that were separated by physical barriers and provided insight into the virus spread in wild boar over this extended time period.

More recently, full genome analysis has been applied to the CSF outbreak that started in 2018 in Japan (Sawai et al., 2021). A genotype 2.1 virus was initially detected in a farm in the Gifu prefecture and later found to be present in wild boar in the surrounding region. By 2020, the virus had spread to 65 farms and over 200 wild boar cases had been detected, with long distance transmissions occurring. Analysis of full-length genome sequences of 155 isolates, from farms and wild boar, indicated a similar virus evolution rate to earlier studies and molecular clock analysis allowed estimates of when the virus was introduced into the country, and when subsequent long distant transmission events may have occurred. Phylogeographic analysis indicated that the virus movement occurred mainly between neighboring regions, although some movements between non-adjacent areas were inferred, implying that both wild boar-mediated spread and transmission by anthropogenic factors occurred. A novel method of analysis of shared nucleotide variants allowed a greater degree of resolution between the highly similar Japanese isolates (Yamamoto et al., 2021). This method agreed well with standard phylogenetic analysis and helped define the borders between closely related groups of viruses thus enabling more detailed inferences about transmission pathways.

In summary, our study supports other recent studies that the full-length CSFV genome sequence can be useful to inform about disease control strategies, in conjunction with traditional epidemiological investigations, and that efforts to obtain and analyze such data in real time during an outbreak are warranted. Had this sequence information been available in the year 2000, it could have informed the field of epidemiological investigations substantially. However, interpretation of the significance of a small number of genetic changes remains challenging, especially if data are accumulating in real time (Villabona-Arenas et al., 2020). Greater knowledge on the rate of genetic changes occurring in the CSFV genome during different transmission events, particularly during genetic bottlenecks, such as during vertical or fomite transmission, will assist in interpreting the

significance of genetic data within the epidemiological context. The potential to enhance the resolution through the analysis of intra- and inter-host single nucleotide variants will be interesting to examine in further studies that could allow to better inform on CSFV transmission pathways in real time.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found online at: <https://www.ncbi.nlm.nih.gov/>, accession numbers LC702327 to LC702341.

AUTHOR CONTRIBUTIONS

FS and HC conceived the project and acquired funding. RS, SM, and SG contributed to sample processing, generation, and analysis of sequence data. BC performed phylogenetic analysis. HC prepared the manuscript. All authors critically reviewed the manuscript.

REFERENCES

- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Campbell, F., Strang, C., Ferguson, N., Cori, A., and Jombart, T. (2018). When are pathogen genome sequences informative of transmission events? *PLoS Pathog.* 14, e1006885. doi: 10.1371/journal.ppat.1006885
- Cottam, E. M., Thebaud, G., Wadsworth, J., Gloster, J., Mansley, L., Paton, D. J., et al. (2008b). Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc Biol Sci.* 275, 887–895. doi: 10.1098/rspb.2007.1442
- Cottam, E. M., Wadsworth, J., Shaw, A. E., Rowlands, R. J., Goatley, L., Maan, S., et al. (2008a). Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS Pathog.* 4, e1000050. doi: 10.1371/journal.ppat.1000050
- COVID-19 Genomics UK (COG-UK). (2020). An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe.* 1, e99–e100. doi: 10.1016/S2666-5247(20)30054-9
- Freimanis, G. L., Di Nardo, A., Bankowska, K., King, D. J., Wadsworth, J., Knowles, N. J., et al. (2016). Genomics and outbreaks: foot and mouth disease. *Rev. Sci. Tech.* 35, 175–189. doi: 10.20506/rst.35.1.2426
- Ganges, L., Crooke, H. R., Bohorquez, J. A., Postel, A., Sakoda, Y., Becher, P., et al. (2020). Classical swine fever virus: the past, present and future. *Virus Res.* 289, 198151. doi: 10.1016/j.virusres.2020.198151
- Goller, K. V., Gabriel, C., Dimna, M. L., Potier, M. L., Rossi, S., Staubach, C., et al. (2016). Evolution and molecular epidemiology of classical swine fever virus during a multi-annual outbreak amongst European wild boar. *J. Gen. Virol.* 97, 639–645. doi: 10.1099/jgv.0.000376
- Grubaugh, N. D., Gangavarapu, K., Quick, J., Matteson, N. L., De Jesus, J. G., Main, B. J., et al. (2019a). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* 20, 8. doi: 10.1186/s13059-018-1618-7
- Grubaugh, N. D., Ladner, J. T., Lemey, P., Pybus, O. G., Rambaut, A., Holmes, E. C., et al. (2019b). Tracking virus outbreaks in the twenty-first century. *Nat. Microbiol.* 4, 10–19. doi: 10.1038/s41564-018-0296-2
- Hunt, M., Gall, A., Ong, S. H., Brener, J., Ferns, B., Goulder, P., et al. (2015). IVA: accurate *de novo* assembly of RNA virus genomes. *Bioinformatics* 31:2374–2376. doi: 10.1093/bioinformatics/btv120
- Jiang, D. L., Gong, W. J., Li, R. C., Liu, G. H., Hu, Y. F., Ge, M., et al. (2013). Phylogenetic analysis using E2 gene of classical swine fever virus

FUNDING

The authors and these studies were financially supported by the UK Department for Environment, Food and Rural Affairs (Defra), the Scottish Government, and the Welsh Government projects SE2210 and SE2215.

ACKNOWLEDGMENTS

We thank M. Khatri for technical assistance and K. Lloyd-Jones for optimization of the omnicleave extraction method.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.909396/full#supplementary-material>

- reveals a new subgenotype in China. *Infect. Genet. Evol.* 17, 231–238. doi: 10.1016/j.meegid.2013.04.004
- Leifer, I., Hoffmann, B., Hoper, D., Bruun Rasmussen, T., Blome, S., Strebelow, G., et al. (2010). Molecular epidemiology of current classical swine fever virus isolates of wild boar in Germany. *J. Gen. Virol.* 91, 2687–2697. doi: 10.1099/vir.0.023200-0
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Mackinnon, J. D. (2001). Some clinical and epidemiological aspects of the outbreak of classical swine fever in East Anglia in 2000. *The State Vet J.* 11, 2–7.
- Paton, D. (2002). The reappearance of classical swine fever in England in 2000. *Trends Emerging Viral Infect. Swine.* Chapter 5.3, 153–158. doi: 10.1002/9780470376812.ch5c
- Paton, D. J., McGoldrick, A., Greiser-Wilke, I., Parchariyanon, S., Song, J. Y., Liou, P. P., et al. (2000). Genetic typing of classical swine fever virus. *Vet. Microbiol.* 73, 137–157. doi: 10.1016/S0378-1135(00)00141-3
- Petrova, V. N., and Russell, C. A. (2018). The evolution of seasonal influenza viruses. *Nat. Rev. Microbiol.* 16, 47–60. doi: 10.1038/nrmicro.2017.118
- Postel, A., Austermann-Busch, S., Petrov, A., Moennig, V., and Becher, P. (2018). Epidemiology, diagnosis and control of classical swine fever: recent developments and future challenges. *Transbound Emerg. Dis.* 65(Suppl 1), 248–261. doi: 10.1111/tbed.12676
- Postel, A., Schmeiser, S., Bernau, J., Meindl-Boehmer, A., Pridotkas, G., Dirbakova, Z., et al. (2012). Improved strategy for phylogenetic analysis of classical swine fever virus based on full-length E2 encoding sequences. *Vet. Res.* 43, 50. doi: 10.1186/1297-9716-43-50
- Postel, A., Schmeiser, S., Perera, C. L., Rodriguez, L. J., Frias-Lepoureau, M. T., and Becher, P. (2013). Classical swine fever virus isolates from Cuba form a new subgenotype 1.4. *Vet. Microbiol.* 161, 334–338. doi: 10.1016/j.vetmic.2012.07.045
- Rios, L., Coronado, L., Naranjo-Feliciano, D., Martinez-Perez, O., Perera, C. L., Hernandez-Alvarez, L., et al. (2017). Deciphering the emergence, genetic diversity and evolution of classical swine fever virus. *Sci Rep.* 7:17887. doi: 10.1038/s41598-017-18196-y
- Sawai, K., Nishi, T., Fukai, K., Kato, T., Hayama, Y., and Yamamoto, T. (2021). Phylogenetic and phylodynamic analysis of a classical swine fever virus outbreak in Japan (2018–2020). *Transbound Emerg. Dis.* 69, 1529–1538. doi: 10.1111/tbed.14117
- Shimizu, Y., Hayama, Y., Murato, Y., Sawai, K., Yamaguchi, E., and Yamamoto, T. (2021). Epidemiological analysis of classical swine fever in wild boars in Japan. *BMC Vet. Res.* 17, 188. doi: 10.1186/s12917-021-02891-0

- Villabona-Arenas, C. J., Hanage, W. P., and Tully, D. C. (2020). Phylogenetic interpretation during outbreaks requires caution. *Nat. Microbiol.* 5, 876–877. doi: 10.1038/s41564-020-0738-5
- Yamamoto, T., Sawai, K., Nishi, T., Fukai, K., Kato, T., Hayama, Y., et al. (2021). Subgrouping and analysis of relationships between classical swine fever virus identified during the 2018-2020 epidemic in Japan by a novel approach using shared genomic variants. *Transbound Emerg. Dis.* 69, 1166–1177. doi: 10.1111/tbed.14076

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Crown Copyright © 2022 Strong, McCleary, Grierson, Choudhury, Steinbach and Crooke. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.