

# A Shift in *Porcine Circovirus 3* (PCV-3) History Paradigm: Phylodynamic Analyses Reveal an Ancient Origin and Prolonged Undetected Circulation in the Worldwide Swine Population

Giovanni Franzo,\* Wanting He, Florencia Correa-Fiz, Gairu Li, Matteo Legnardi, Shuo Su, and Joaquim Segalés

The identification of a new circovirus (*Porcine circovirus 3*, PCV-3) has raised a remarkable concern because of some analogies with *Porcine circovirus 2* (PCV-2). Preliminary results suggest an extremely recent PCV-3 emergence and high mutation rate. Retrospective studies prove its circulation at least since the early 1990s, revealing that PCV-3 could have been infecting pigs for an even longer period. Therefore, a new evaluation, based on an updated collection of PCV-3 sequences spanning more than 20 years, is performed using a phylodynamic approach. The obtained results overrule the previous PCV-3 history concept, indicating an ancient origin. These evidences are associated with an evolutionary rate far lower ( $10^{-5}$ – $10^{-6}$  substitution/site/year) than the PCV-2 one. Accordingly, the action of selective pressures on PCV-3 open reading frames (ORFs) seems to be remarkably lower compared to those acting on PCV-2, suggesting either a reduced PCV-3 plasticity or a less efficient host-induced natural selection. A complex and not-directional viral flow network is evidenced through phylogeographic analysis, indicating a long lasting circulation rather than a recent emergence followed by spreading. Being recent emergence has been ruled out, efforts should be devoted to understand whether its recent discovery is simply due to improved detection capabilities or to the breaking of a previous equilibrium.


## 1. Introduction

The *Circovirus* genus claimed veterinarian attention shortly after its discovery, since its members were considered responsible of relevant diseases in birds.<sup>[1–3]</sup> However, it was only in the middle 1990s, when *Porcine circovirus 2* (PCV-2) and its related clinically and economically relevant syndromes (thereafter named Porcine circovirus diseases; PCVD) were recognized, that this genus became the focus of an intensive research activity.<sup>[4,5]</sup>

Since then, a remarkable collection of data and knowledge has been gathered on PCV-2 biology, pathogenesis, epidemiology, control, and evolution.<sup>[6,7]</sup> One of the most astonishing findings was the evidence that, in spite of the sudden PCVD emergence, PCV-2 has shared a long path with domestic and wild swine populations.<sup>[8,9]</sup> Not only this but other swine infection scenarios have pointed

Dr. G. Franzo, M. Legnardi  
Department of Animal Medicine  
Production and Health (MAPS)  
University of Padua  
Viale, dell'Università 16, 35020 Legnaro (PD), Italy  
E-mail: giovanni.franzo@unipd.it

Dr. W. He, Dr. G. Li, Prof. S. Su  
MOE International Joint Collaborative Research Laboratory  
for Animal Health & Food Safety  
Jiangsu Engineering Laboratory of Animal Immunology  
Institute of Immunology  
College of Veterinary Medicine  
Nanjing Agricultural University  
Nanjing 210000, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.201901004>.

© 2019 The Authors. Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.201901004

Dr. F. Correa-Fiz  
IRTA  
Centre de Recerca en Sanitat Animal (CRESA, IRTA-UAB)  
Campus de la Universitat Autònoma de Barcelona  
Bellaterra 08913, Spain  
Prof. J. Segalés  
UAB  
Centre de Recerca en Sanitat Animal (CRESA, IRTA-UAB)  
Campus de la Universitat Autònoma de Barcelona  
Bellaterra 08913, Spain  
Prof. J. Segalés  
Departament de Sanitat i Anatomia Animals  
Facultat de Veterinària  
Universitat Autònoma de Barcelona  
Bellaterra 08913, Spain

out the role of the modern farming system in the rising of new multifactorial diseases.<sup>[4,10]</sup>

The history seems to repeat itself for a new, recently discovered porcine circovirus: *Porcine circovirus 3* (PCV-3).<sup>[11]</sup> This virus is featured by a circular ssDNA genome of  $\approx 2000$  bases containing three open reading frames (ORFs) identified so far,<sup>[11]</sup> although only ORF1 and ORF2 have been characterized. ORF1, located on the positive strand, apparently codes for a single replicase protein of 296–297 aa.<sup>[11,12]</sup> ORF2 is located on the negative DNA viral strand and encodes the capsid protein.<sup>[12]</sup> Despite the common genomic organization, PCV-3 is distantly related to other known circoviruses, although a certain relation with bat and avian circoviruses has been suggested based on phylogenetics, codon bias and genome composition analysis.<sup>[13,14]</sup>

PCV-3 was first identified in the USA in 2015 using a metagenomics approach in tissues from animals displaying porcine dermatitis and nephropathy syndrome (PDNS) and reproductive disorders.<sup>[11]</sup> Thereafter, it has been identified all over the world, including Asia,<sup>[15–18]</sup> Europe<sup>[19–22]</sup> and South America,<sup>[11,23]</sup> in presence of several clinical syndromes like PDNS,<sup>[11]</sup> reproductive disorders,<sup>[24,25]</sup> respiratory disease,<sup>[26,27]</sup> and myocarditis.<sup>[28]</sup> Moreover, PCV-3 genome has been detected through *in situ hybridization* and immunohistochemistry in different tissue lesions,<sup>[11,28,29]</sup> supporting the potential etiological role of PCV-3. However, its identification with comparable prevalence both in healthy pigs<sup>[30,31]</sup> and wild boar<sup>[32–34]</sup> questions the pathogenic role of this virus or at least suggests the need of other concomitant factors to trigger overt disease.

Nevertheless, based on the “PCV-2 experience,” a pressing concern has been directed toward the study of the evolution and origin of PCV-3: is PCV-3 a newly originated viral species or is it an ancient one that has been circulating for a long time, only recently emerging as a potential threat for swine industry?

The first attempts to answer this question, based on a limited number of samples collected over a short time period, pointed out a recent PCV-3 origin (approximately in the new millennium) coupled with a noteworthy evolutionary rate.<sup>[35]</sup> However, retrospective studies performed in Sweden<sup>[20]</sup> in 1993 and Spain<sup>[36]</sup> and China<sup>[15]</sup> in 1996 demonstrated that PCV-3 has been circulating during several decades in domestic pigs. Remarkably, PCV-3 has been detected in the oldest samples so far tested in these studies, showing a marked limit in our knowledge due to scarce data availability and suggesting that this virus could have infected pigs for even a longer period.<sup>[12]</sup>

Because of the relevance in terms of PCV-3 epidemiology understanding and control strategies development, the present study attempts to re-evaluate PCV-3 history, population dynamics and spreading patterns based on a wider sequence dataset, spanning the broader collection time window currently available. The herein reported results throw a new light on PCV-3 history and its evolutionary pathways.

## 2. Results

### 2.1. Dataset

A total of 187 complete genome sequences were included in the *dataset1*, 208 in the *dataset2* and 421 (427 nucleotide

long, spanning the region between nucleotides 1347 and 1803 based on the reference genome KT869077.1) in the *dataset3*. All databases comprised sequences originating from 14 countries over a time period between 1996 and 2018, although with different number per country (Datasets S1–S3, Supporting Information).

### 2.2. PCV-3 Origin, Evolutionary Rate, and Population Dynamics

The analysis performed on the *dataset1* provided a time to most recent common ancestor (tMRCA) of 1372.23 years before present (ybp), although with a high uncertainty [95 high posterior density [95HPD]: 666.30–3780.91] (Figure 1a and Figure S1, Supporting Information). The lower tMRCA estimate was provided by *dataset3* (156.9 ybp [95HPD: 75.72–318.28]) (Figure 1b and Figure S2, Supporting Information). *Dataset2* provided an intermediate tMRCA estimation ( $\approx 300$  ybp) when the results of the ten independent runs were averaged. However, a substantial overlap in the confidence intervals compared to *dataset1* was observed [95HPD: 138.79–1333.4]. The different runs, performed on 10 randomly generated sequence datasets provided consistent results (Figure 1c and Figure S3, Supporting Information).

Molecular clock estimates for the complete genome dataset reported  $2.35 \times 10^{-5}$  [95HPD:  $8.79 \times 10^{-6}$ – $4.71 \times 10^{-5}$ ],  $4.71 \times 10^{-5}$  [95HPD:  $1.75 \times 10^{-5}$ – $9.38 \times 10^{-5}$ ], and  $5.13 \times 10^{-5}$  [95HPD:  $1.32 \times 10^{-5}$ – $1.92 \times 10^{-5}$ ] substitution/site/year for the ORF1, ORF2, and intergenic regions, respectively (Figure 1a).

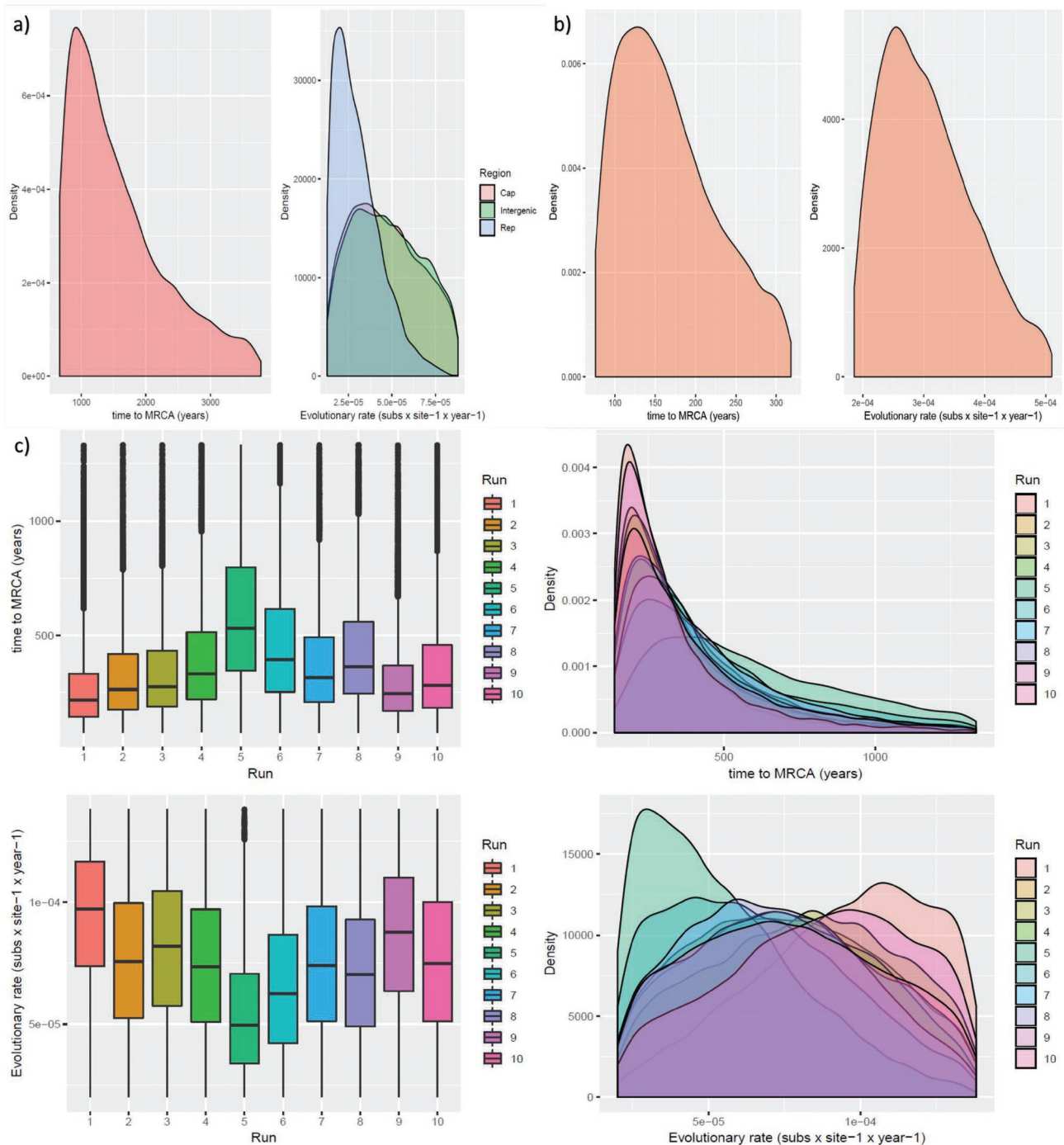
Accordingly, the *dataset2* estimates showed an evolutionary rate in the range  $10^{-4}$ – $10^{-5}$  substitution/site/year, consistent among different runs (Figure 1c). Finally, the evolutionary rate estimated using the *dataset3* was  $2.88 \times 10^{-4}$  [95HPD:  $1.85 \times 10^{-4}$ – $4.41 \times 10^{-4}$ ] (Figure 1b).

The relative genetic diversity ( $N_e \times t$ ) was featured by a broad 95HPD, independently from the considered dataset or run. Nevertheless, a trend toward a rise in the viral population size was observed, peaking approximately in the 1980s (Figure 2).

### 2.3. Phylogeographic Analysis

Reconstruction of the viral spread over time demonstrated a relevant uncertainty (i.e., posterior probability lower than 0.9) in the ancestral country estimation (Figure 3). Therefore, accounting for this uncertainty, relationship among strains were estimated in terms of well-supported (i.e., Bayesian Factor (BF) > 10) contacts among countries.

Except for a certain interdataset variability, the overall picture supports the presence of three interconnected nuclei of viral spread, corresponding to Asia, Europe, and America, with Asia acting as an intermediary between Europe and America (Figure 4 and Figures S4 and S5, Supporting Information). Within those main areas, several local migration routes were proven significant, being China the most likely responsible for viral spread to other Asian countries (e.g., South Korea, Thailand, Japan, and Russia). A more complex and less directional network could be inferred for European countries. Viral spread appears also to occur between USA

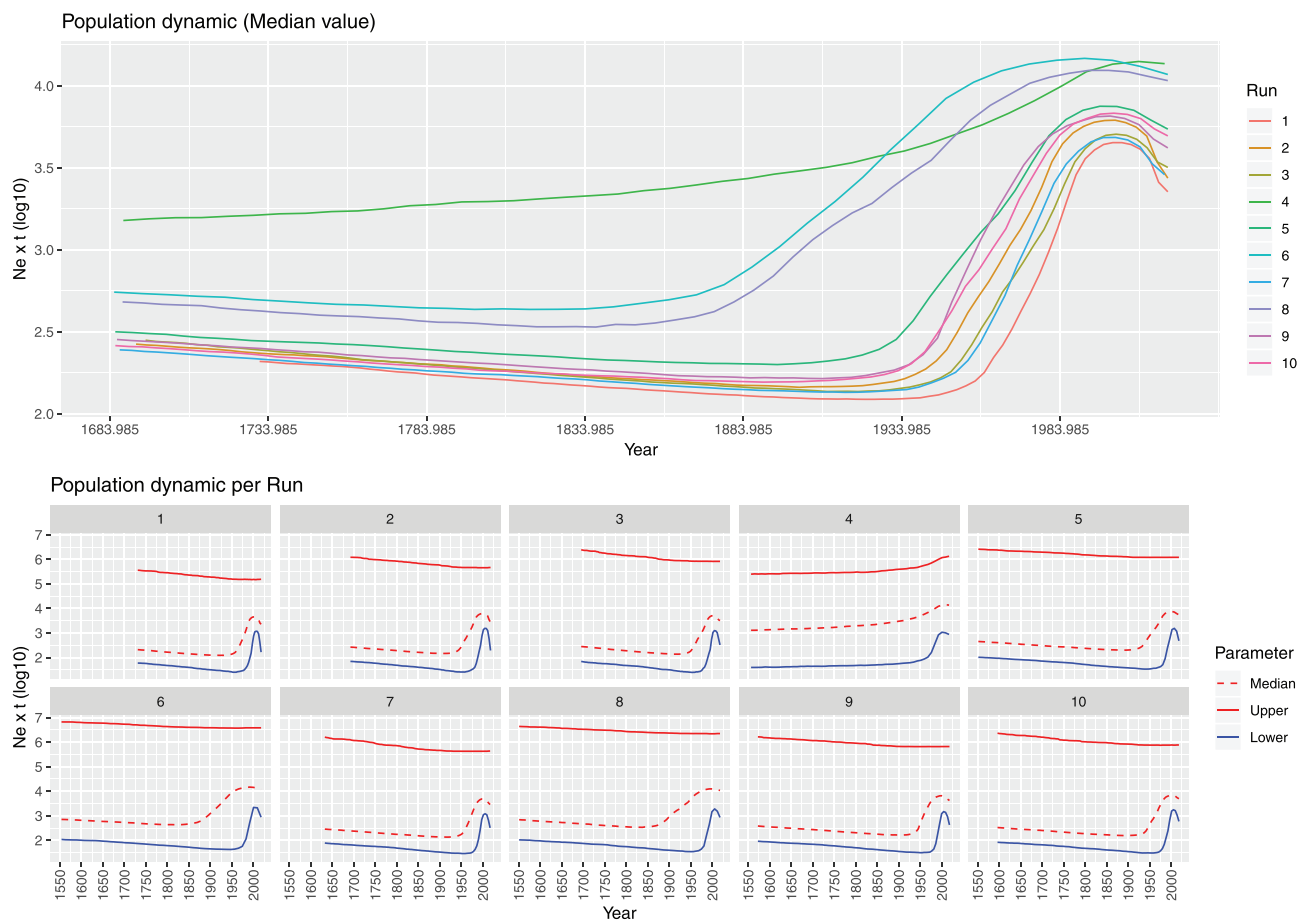


**Figure 1.** a) PCV-3 tMRCA and evolutionary rate based on *dataset1*. Left figure: density plot of the MRCA posterior probability. Right figure: density plot of the mean evolutionary rate posterior probability. Evolutionary rates of different genomic regions have been color coded. The 95HPD intervals are reported for both figures. b) PCV-3 tMRCA and evolutionary rate based on *dataset3*. Left figure: density plot of the MRCA posterior probability. Right figure: density plot of the mean evolutionary rate posterior probability. The 95HPD intervals are reported for both figures. c) PCV-3 tMRCA and evolutionary rate based on *dataset2*. Upper figure: box plot (left) and density plot (right) of the MRCA posterior probability. Lower figure: box plot (left) and density plot (right) of the mean evolutionary rate posterior probability. Results have been estimated performing ten independent runs based on randomly sampled sequences. The 95HPD intervals are reported for both figures.

and Mexico. However, the contact with the South American country included in the study (i.e., Brazil) was more likely mediated by viral strain exchange with China rather than other American countries.

#### 2.4. Selective Pressure Analysis

Selective pressure analysis demonstrated a clear dominance of sites under negative pressure both in the Rep and Cap proteins



**Figure 2.** PCV-3 genotype population dynamics reconstructed based on *dataset2*. Upper figure: mean relative genetic diversity ( $N_e \times t$ ) of the worldwide PCV-3 population overtime. The results of the ten independent runs have been color coded. Lower figure: median and upper and lower 95HPD values are reported for each run.

(Figure 5). Limited evidences were detected of pervasive diversifying selection in both proteins: positions 19, 44, 45, and 122 of the Rep protein were proven statistically significant with the Fast Unconstrained Bayesian AppRoximation (FUBAR) method only, while position 5 of the Cap was detected by fixed effects likelihood (FEL) and FUBAR, and sites 56 and 137 by the latter method only.

On the other hand, episodic diversifying selection was detected in positions 5, 44, and 97 in the Rep, and in positions 5, 56, and 102 in the Cap proteins.

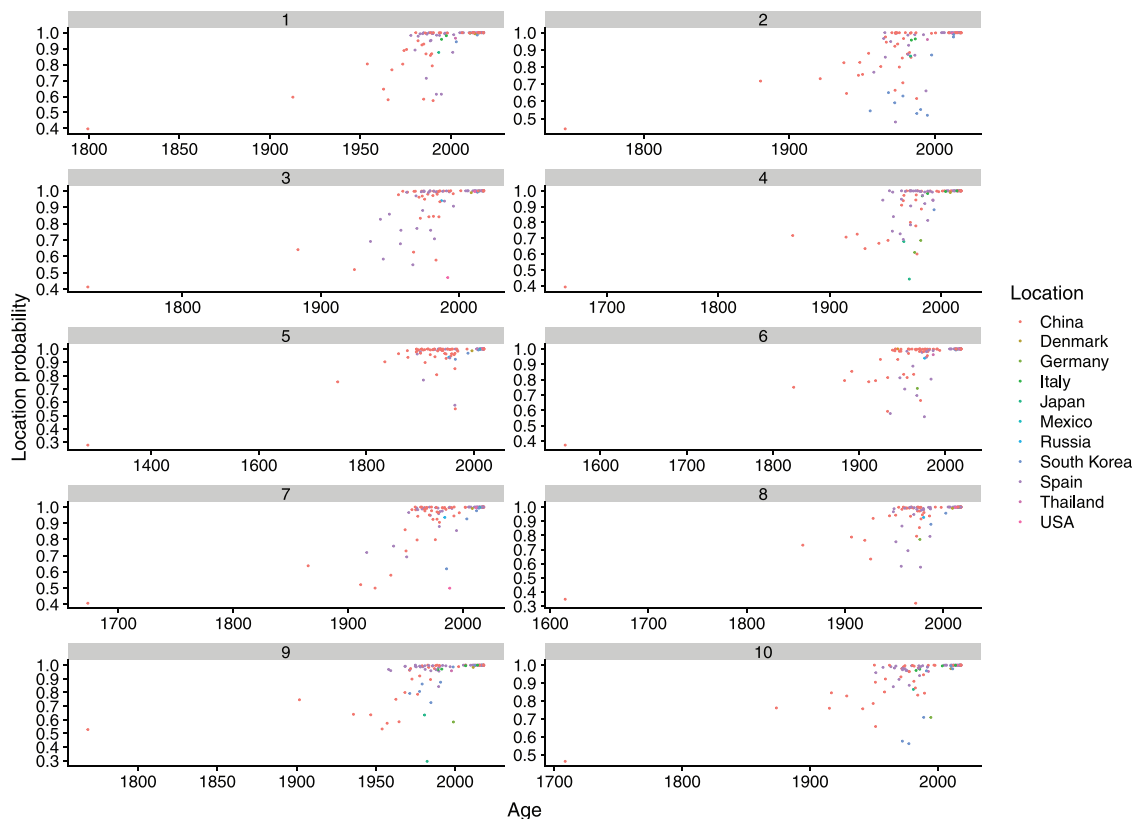
A comparison of the selective pressures acting on the two genes showed a statistically significant different selective regimen (i.e., different selection strength and proportion of sites under selection) acting on both genes ( $p = 0.003$ ), being the overall dN/dS and proportion of sites higher in the ORF2 alignment.

Homology modeling allowed an approximate reconstruction of the protein tertiary structure. Two sites (i.e., 5 and 44) under episodic diversifying selection were located on the Rep protein surface, as well as positions 19 and 45 (detected by FUBAR) (Figure 6a). The remaining sites under diversifying selection were buried in the protein predicted structure. In the Cap protein, all the amino-acids under episodic diversifying selection were located on the capsid surface (Figure 6b), while sites detected by FUBAR were located inside the protein structure.

### 3. Discussion

The identification of PCV-3, a new porcine circovirus resembling the significant pig pathogen PCV-2 from several perspectives, has raised a great interest toward the epidemiology and biology of this virus. One of the most relevant questions to solve is about its origin and evolution. The number of recognized swine viruses has remarkably increased in the last years<sup>[10,12]</sup> raising a dichotomous question to this phenomenon: have these viruses been recently introduced in a new host species?, or have they circulated for a long period, undetected in the domestic pig population and emerged only recently as a major threat because of concomitant factors or improved detection and research technologies? While the second explanation appears more likely for several pathogens, including PCV-2,<sup>[4,9]</sup> their natural prior-to-emergence history remains largely unknown.

Predictably, similar questions arose on the PCV-3 origin and some studies have meritoriously attempted to investigate this issue.<sup>[35,37]</sup> Preliminary results suggested a recent PCV-3 emergence, which was located approximately at the beginning of the new millennium. However, all the mentioned studies included only sequences of PCV-3 strains collected over a short time frame; i.e., after 2015. Therefore, a poor precision on tMRCA



**Figure 3.** Ancestral location scatter plot. Scatter plot representing the posterior probability of each ancestral location (color coded) prediction over time. The results of ten independent BEAST run are reported.

estimation could be forecasted, especially considering the low resolution in the collection date annotation (i.e., collection year resolution). In fact, tMRCA underestimation can also severely bias the substitution rate estimation. A recent tMRCA implies the current genetic heterogeneity originated over a limited time

period, imposing a relatively fast evolutionary rate. Accordingly, PCV-3 was reported to be the fastest evolving circovirus,<sup>[35]</sup> which appears suspicious considering the limited genetic variability reported so far. The detection of PCV-3 in retrospective samples collected during the 1990s in several countries from Europe and Asia confirmed its underestimated origin and claimed further analysis based on an updated genetic information availability.<sup>[15,20,36]</sup>

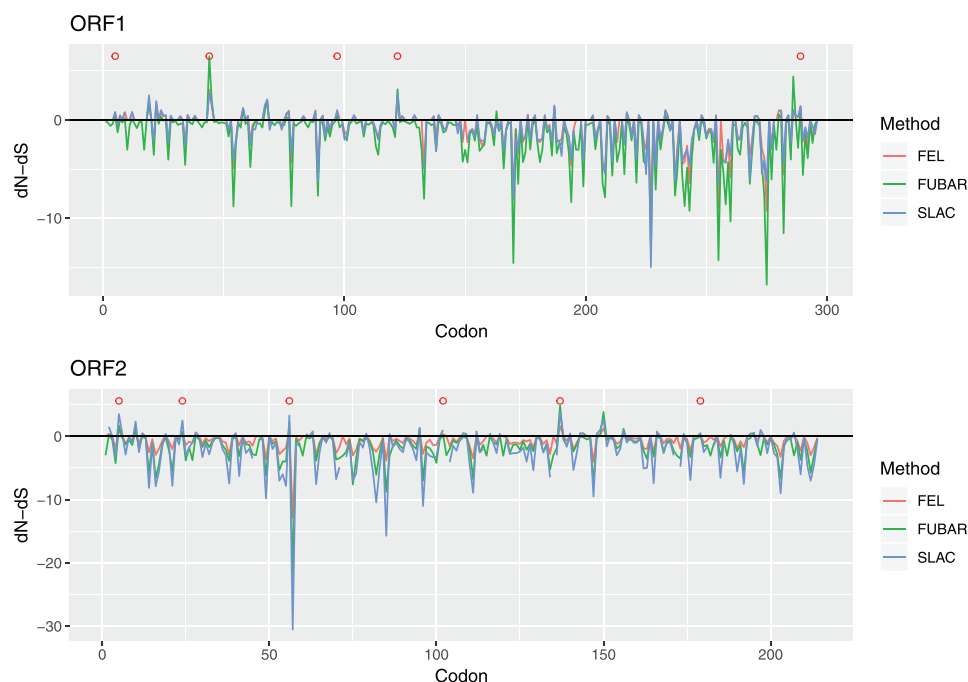
Remarkably, although in Fu et al., estimation of the “PCV-3 only clade” still showed a recent origin, they were able to anticipate PCV-3 emergence in the middle of the previous century by including the PorkNW2/USA/2009 strain in the coalescent analysis.<sup>[37]</sup> This strain is genetically similar to PCV-3 but shows a lower genome size and, although it could be reasonably classified as a defective PCV-3 or a replicative intermediate,<sup>[38]</sup> its classification remains controversial, mining the reliability of the results. Substantially comparable results were obtained by Saraiva et al.,<sup>[38]</sup> however, the PCV-3 strains included in this report were collected after 2015 and only Asia and America were represented.

Based on these premises, the present study aimed to reconstruct the evolution history and population dynamics of PCV-3 based on a high quality and updated sequence dataset spanning a sampling time longer than 20 years (1996–2018). Despite our efforts, the currently available data are still limited and potentially biased by the different diagnostic and sequencing activities performed in different countries. Therefore, particular care was dedicated to evaluate result consistency by analyzing



**Figure 4.** Network reporting the well-supported migration routes (BF > 10). PCV-3 spreading path among different countries estimated using ten independent BEAST runs (color coded) based on *dataset2*. The arrows' size is proportional to the BF value.

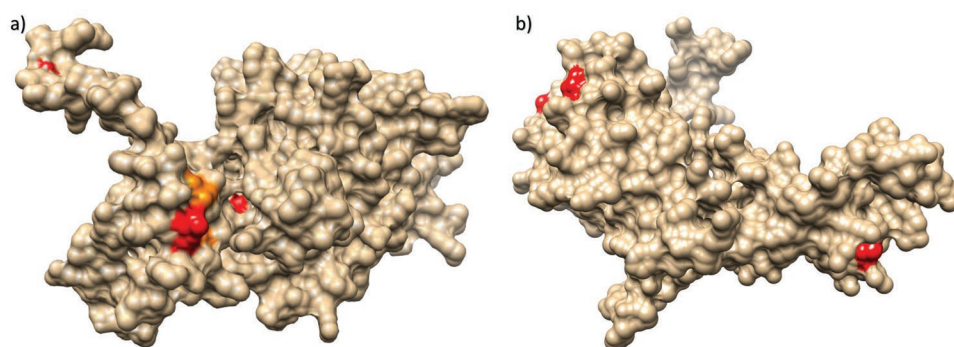




**Figure 5.** Line graph reporting dN–dS values estimated for each codon position of ORF1 (upper figure) and ORF2 gene (lower figure) using different methods. Sites detected to be under statistically significant positive selection by MEME are reported as red circles.

different genome regions and randomly generated, down-sampled, datasets with different features (i.e., alignment length, number of sequences, presence of complete coding regions, etc.).<sup>[39]</sup> Independently from the considered dataset, PCV-3 origin was always backdated before 1900 at least. However, with the *dataset3* being the only exception, a far more ancient origin was supported, in the order of several centuries (*dataset2*) or millennia (*dataset1*). The large overlapping in 95HPD between the two datasets and analysis runs, further support the consistency of the results. The trend toward an increase in tMRCA with increasing sequence length supports the usefulness of adding informative sites to improve parameter estimation accuracy.<sup>[40]</sup> Particularly, the full genome dataset displays some features that could have additionally contributed to tMRCA back-estimate. The inclusion and independent modeling of more conserved regions, like the ORF1, could have allowed

the reconstruction of more ancient events. Moreover, the use of the protein coding region allowed to implement a model that, although it cannot be considered an actual codon model (accounting for differential synonymous and nonsynonyms substitution rates), depicted the heterogeneous substitution rates among different alignment regions and codon positions in a more effective way. Several studies have pointed out the underestimation of the origin of rapidly evolving viruses and the occurrence of a “time-dependent rate phenomenon,” where viral evolutionary rates appear to vary over time, continuously decreasing along with the timescale of rate measurement.<sup>[41–43]</sup> Among the possible causes of these phenomena, substitution saturation, poor modeling of natural selection and inability to deal with the vast majority of substitutions occurring multiple times at a limited subset of sites (i.e., high rate heterogeneity among sites), have been advocated.<sup>[41–44]</sup> Therefore, because



**Figure 6.** a) Tertiary structure of the Rep protein estimated through homology model. When located on the protein surfaces, sites under pervasive and episodic diversifying selection have been highlighted in orange and red, respectively. b) Tertiary structure of the Cap protein estimated through homology model. When located on the protein surfaces, sites under pervasive and episodic diversifying selection have been highlighted in red.

of the higher number of informative sites and more realistic model, the complete genome-based estimations can likely be considered a more reliable tMRCA approximation. However, a potential underestimation of PCV-3 tMRCA can still not be excluded based on its high diversity compared to other known circoviruses, stressing the need for further improvements in our mathematical modeling capabilities.

Apart from these considerations, which are far beyond the scope of the study, the achieved results consistently demonstrated that PCV-3 origin should have occurred centuries ago. This scenario is further supported by the worldwide distribution pattern of the virus, featured by strain collected in different countries widely interspersed in the phylogenetic tree (Figures S1–S3, Supporting Information), similarly to what has previously been reported by other authors.<sup>[12,38,45,46]</sup>

The root and ancestral node location posterior probability was often low (as indicated in Figure 3), revealing a largely expected uncertainty considering the large time frame between the estimated tMRCA and the oldest available sequences. Long branches and the lack of historical data hinder the inference of the spatial history of older viral lineages with confidence. Additionally, the long branch length is likely to conceal additional spatial movements between multiple locations.<sup>[47]</sup> Consequently, migration patterns were evaluated in terms of ‘contact’ among countries, minimizing the risk of over-interpretation of their timing and directionality. Altogether, three major nuclei of local transmission (i.e., North America, Europe and Asia), connected by long distance transmission events were consistently identified. Such uncontrolled viral circulation can be easily explained by the recent PCV-3 identification and by its frequent detection in healthy animals. The strain distribution along the tree and the best fitting of a symmetric migration model over the asymmetric one, poses in favor of a long lasting viral circulation rather than a recent emergence followed by progressive introduction in different countries. Nevertheless, based on the network structure and following a parsimony criterion, China seems to have played a pivotal role in the spreading of PCV-3 both within and between continents, which could seem surprising being China a minor exporter of live swine. Despite our attempt to limit the effect of uneven sequence availability from different countries by down-sampling the original dataset and creating randomly generated ones, a certain bias in spreading pattern inference due to the more intense sequencing activity in China cannot be totally excluded. However, fully comparable long and short range spreading pattern has already been described for other livestock pathogens, like PCV-2 and *Infectious bronchitis virus* (IBV),<sup>[9,48,49]</sup> supporting the plausibility of a comparable scenario for PCV-3. Therefore, the actual presence of some preferential livestock-virus “highways” can be suggested and further efforts should be dedicated to highlight the underlying causes, investigating the factors affecting viral dispersal and introducing effective control strategies, if necessary. While our hypothesis followed a parsimony criterion, other patterns could represent a more accurate depiction of PCV-3 spreading. However, the current lack of accurate data in pig flow from most of the countries considered in the present study (especially in the time period when initial PCV-3 dispersal likely occurred) prevents additional investigations of an association between swine or swine products trades and viral spread.

Despite the long lasting PCV-3 circulation in the swine population, an increase in the viral relative genetic diversity ( $N_e \times t$ ; i.e., a proxy of population size dynamics) was observed in the last decades using a skyline plot, similarly to what previously described for PCV-2.<sup>[9]</sup> Common epidemiological causes could thus be hypothesized, like the alteration in the long lasting equilibrium due to pig raising conditions in the context of expanding intensive farming.<sup>[4]</sup> The estimated increase in viral population size, largely anticipate the identification of PCV-3 in animals showing clinical signs. Therefore, the answer to whether the detected rise mirrors an increased pathogenic role of PCV-3 or is simply due to a wider viral circulation (of no clinical significance) in a bigger and more connected animal population world, remains elusive. Although some evidences appear to support a certain association between PCV-3 and clinical disease, contradictory reports have been published up to date and more extensive studies should be performed.

The ancient origin of PCV-3 implies also a lower evolutionary rate compared to PCV-2 and previous PCV-3 estimation, both reporting a substitution rate in the order of magnitude of  $10^{-3}$  substitution/site/year.<sup>[9,35,50]</sup> On the contrary, the present study results showed a far lower rate (i.e.,  $\approx 10^{-5}$  substitution/site/year), which is consistent with the limited genetic variability so far observed and with the high similarity between recent sequences and those obtained from early-mid 1990s.<sup>[12]</sup> This evidence could suggest a lower intensity of diversifying selective pressures shaping the evolution of this virus.

Accordingly, compared with previous studies evaluating the forces acting on PCV-2, the number of sites under positive selection in the Cap was remarkably lower in PCV-3.<sup>[9,51]</sup> Although further confirmation will be needed, this scenario could be due to a lower plasticity of PCV-3 or to a less intense host-induced natural selection, which could be tentatively considered as an evidence of a lower PCV-3 virulence and/or prolonged virus-host co-evolution, leading to a decreased immune response stimulation. Unfortunately, no reliable experimental in vitro and/or in vivo model is currently available to investigate the PCV-3 immunopathogenesis and interaction within the host, which are likely the most relevant determinants of its evolution. Overall, the field of PCV-3 immunology is still in its infancy and further confirmation must be provided to shed light on this hypothesis plausibility. Nevertheless, the phylodynamic approach implemented in the present study, based on a global viral sampling spanning more than 20 years, can provide a useful and consistent depiction of the overall patterns and determinants of PCV-3 evolution, avoiding assumptions and constraints of viral biology induced by experimental conditions.

In fact, the ORF1 gene demonstrated a tendency toward a lower substitution rate compared to the ORF2, which was reflected by a statistically significant difference in diversifying selection acting on the two coding regions. These results, coupled with the location on the protein surface of sites under diversifying selection, suggest at least a limited action of the host immune response in shaping PCV-3 evolution. However, it must be stressed that the homology based estimation of Rep and Cap protein conformation should be evaluated with caution because of the absence of closely related experimentally derived tertiary structures.

Overall, the present study provides an updated representation of PCV-3 origin, population dynamics and evolution, pointing out a quite ancient viral origin and a low evolutionary rate compared to other circoviruses of clinical relevance. These results could contribute to the evaluation of the actual PCV-3 relevance for swine industry and, possibly, to the planning of effective control strategies. It must be emphasized that the present work just scratched the surface of PCV-3 history and biology and future and constant re-evaluation of the present results will be mandatory to update and improve the knowledge of this emergent virus behavior.

## 4. Experimental Section

**Dataset Preparation:** All currently available PCV-3 sequences were downloaded from Genbank (accessed on 2018 November 29) and annotated with collection year and country when available. Sequences lacking of these data were removed from the dataset. Despite the fact that the obtained sequence collection was the broader currently available, the molecular epidemiology information was still limited and sparse.

To deal with unbalanced sequence availability, which could bias the results, different sequence datasets were prepared to alternatively benefit from the higher sequence length or number (i.e., representativeness): *dataset1* included all available complete genome; *dataset2* comprised long PCV-3 sequences (1000bp), including the Spanish ones obtained during a retrospective study conducted from the mid-1990s onward;<sup>[36]</sup> *dataset3* was based on a region where the higher number of PCV-3 sequence had a full coverage.

Each dataset was designed using the following approach:

**Dataset1:** Complete genome sequences were divided in ORF1, ORF2, and intergenic regions (both intergenic regions were merged in a single partition). Coding regions were aligned at amino-acid level and then back translated to nucleotide sequence using the MAFFT algorithm<sup>[52]</sup> implemented in TranslatorX.<sup>[53]</sup> Poorly aligned sequences as well as those showing premature stop codons or frame-shift mutation were excluded from the analysis. Recombination analysis was performed using RDP4<sup>[54]</sup> on complete genome alignment and each ORF independently.

**Dataset2:** This dataset was primarily designed to benefit from all the 1990s samples, including the partial ones describe by Klaumann et al.<sup>[36]</sup> Additionally, *dataset2* was used to evaluate the impact of sampling bias in analysis results. To this purpose, ten independent datasets were generated by randomly sampling a maximum of ten sequences for each country-year pair, as described by Franzo et al.<sup>[48,49]</sup> The obtained sequences were aligned using MAFFT version 7.271<sup>[52]</sup> and scanned for recombination events using RDP4.<sup>[54]</sup>

**Dataset3:** All the available, partial and complete, PCV-3 sequences were aligned using MAFFT<sup>[52]</sup> and the alignment region with the highest sequence coverage was selected for further analysis. All sequences spanning the selected region were extracted, realigned, and scanned for recombination events using RDP4.<sup>[54]</sup>

**Population Parameters Estimation/Phylogenetic Analysis:** On each of the above-mentioned datasets, a tip dated serial coalescent analysis was performed using the Bayesian approach implemented in BEAST 1.8.2<sup>[55]</sup> to primarily estimate tMRCAs, evolutionary rate, and population dynamics over time. Additionally, a discrete state phylogeography was performed as described by Lemey et al.<sup>[56]</sup> Additionally, the implementation of the Bayesian stochastic search variable selection (BSSVS) allowed a BF test that identified the most parsimonious description of the spreading process. A BF > 10 was considered as suggestive of a significant migration pattern between country pairs.

For the *dataset1*, the three partitions (ORF1, ORF2, and intergenic regions) were allowed independent substitution and clock models, while a single tree topology for the three regions was constrained. Additionally, ORF1 and ORF2 regions were further partitioned allowing independent evolution models for each codon position. On the other hand, a single

partition was used for *dataset2* and *dataset3*. The substitution model was selected based on the BIC scores calculated using Jmodeltest version 2.1.7. Molecular clock, population dynamics model, and discrete trait substitution model (i.e., symmetric vs asymmetric migration rate) were selected by evaluation of the BF (i.e., the ratio of the compared model marginal likelihoods, estimated using a Path sampling and Stepping stone approach) as suggested by Baele et al.<sup>[57]</sup> Relaxed lognormal molecular clock,<sup>[58]</sup> skyline population model,<sup>[59]</sup> and symmetric migration rate were selected.<sup>[56]</sup>

The tree and the model parameters were estimated over a 100 million generation Markov Chain Monte Carlo (MCMC) chain, sampling them every 10 000 generations. Run results were accepted only if the mixing and convergence, visually inspected using Tracer,<sup>[60]</sup> were adequate and the estimated sample size (ESS) was higher than 200, after discharging the first 20% generations as burn-in. Parameter estimation were summarized as median and 95HPD. The maximum clade credibility tree was estimated using the treeannotator tool of the BEAST 1.8.2<sup>[55]</sup> package. BF of well-supported migration rates was calculated using Spread3.<sup>[61]</sup>

**Selective Pressure Analysis:** The presence of protein sites under diversifying selection was evaluated using several dN–dS based methods. To this purpose, all complete ORF1 and ORF2 sequences were downloaded from Genbank and aligned at codon level using TranslatorX.<sup>[53]</sup> The achieved alignments were tested for pervasive diversifying selection using single-likelihood ancestor counting (SLAC), FEL, and FUBAR<sup>[62,63]</sup> methods implemented in HyPhy.<sup>[64]</sup> Significance level was set to  $p < 0.05$  for SLAC and FEL and to a posterior probability higher than 0.9 for FUBAR. Sites were considered under pervasive diversifying selection when detected by at least two methods.

Since adaptive evolution often occurs in episodic bursts,<sup>[65]</sup> i.e., affecting a small subset of branches at individual sites, episodic diversifying selection was also investigated using MEME.<sup>[66]</sup> The MEME significance level was set  $p < 0.05$ .

The tertiary structure of PCV-3 Rep and Cap proteins were estimated by homology modeling using Phyre 2.0<sup>[67]</sup> to predict sites location under diversifying selection.

The action of selective pressures was also compared among different genes (ORF1 and ORF2) using the dNdSDistributionComparison.bf implemented in HyPhy.<sup>[64]</sup> Particularly, the presence of a difference between the two ORFs in selective pressure strength, proportion of sites under selective pressure and selective regime (i.e. a combination of the two factors) was evaluated.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

G.F., W.H., S.S., and J.S. contributed equally to this work. W.T.H., G.R.L., and S.S. were financially supported by the National Key Research and Development Program of China (2017YFD0500101), the Natural Science Foundation of Jiangsu Province (BK20170721), and the China Association for Science and Technology Youth Talent Lift Project (2017–2019). J.S. and F.C.-F. thank the financial support of the E-RTA2017-00007-00-00 project from the Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (Spanish Government). G.F. thanks the financial support of the Department of Animal Medicine, Production and Health, University of Padua, (grant BIRD187958/18).

## Conflict of Interest

The authors declare no conflict of interest.



## Keywords

epidemiology, evolution, porcine circovirus 3 (PCV-3), phylodynamic, phylogeography

Received: April 29, 2019

Revised: July 12, 2019

Published online: September 30, 2019

- [1] I. Tischer, R. Rasch, G. Tochtermann, *Zentralbl. Bakteriell., Parasitenkd., Infektionskrankh. Hyg., Abt. 1: Orig., Reihe A* **1974**, 226, 153.
- [2] I. Tischer, H. Gelderblom, W. Vettermann, M. A. Koch, *Nature* **1982**, 295, 64.
- [3] D. Todd, *Vet. Microbiol.* **2004**, 98, 169.
- [4] J. Segalés, T. Kekkarainen, M. Cortey, *Vet. Microbiol.* **2013**, 165, 13.
- [5] X. J. Meng, *Virus Res.* **2012**, 164, 1.
- [6] F. Madec, N. Rose, B. Grasland, R. Cariolet, A. Jestin, *Transboundary Emerging Dis.* **2008**, 55, 273.
- [7] M. A. Ssemadaali, M. Ilha, S. Ramamoorthy, *Res. Vet. Sci.* **2015**, 103, 179.
- [8] G. Franzo, M. Cortey, A. M. M. G. de Castro, U. Piovezan, M. P. J. Szabo, M. Drigo, J. Segalés, L. J. Richtzenhain, *Vet. Microbiol.* **2015**, 178, 158.
- [9] G. Franzo, M. Cortey, J. Segalés, J. Hughes, M. Drigo, *Mol. Phylogenet. Evol.* **2016**, 100, 269.
- [10] G. Fournié, L. Kearsley-Fleet, J. Otte, D. U. Pfeiffer, *Vet. Res.* **2015**, 46, 114.
- [11] R. Palinski, P. Piñeyro, P. Shang, F. Yuan, R. Guo, Y. Fang, E. Byers, B. M. Hause, *J. Virol.* **2017**, 91, e01879.
- [12] F. Klaumann, F. Correa-Fiz, G. Franzo, M. Sibila, J. I. Núñez, J. Segalés, *Front. Vet. Sci.* **2018**, 5, 315.
- [13] G. Franzo, J. Segales, C. M. Tucciarone, M. Cecchinato, M. Drigo, *PLoS One* **2018**, 13, e0199950.
- [14] G. Li, H. Wang, S. Wang, G. Xing, C. Zhang, W. Zhang, J. Liu, J. Zhang, S. Su, J. Zhou, *Virulence* **2018**, 9, 1301.
- [15] J. Sun, L. Wei, Z. Lu, S. Mi, F. Bao, H. Guo, C. Tu, Y. Zhu, W. Gong, *Transboundary Emerging Dis.* **2018**, 65, 607.
- [16] X. Ku, F. Chen, P. Li, Y. Wang, X. Yu, S. Fan, P. Qian, M. Wu, Q. He, *Transboundary Emerging Dis.* **2017**, 64, 703.
- [17] S. Qi, M. Su, D. Guo, C. Li, S. Wei, L. Feng, D. Sun, *Transboundary Emerging Dis.* **2019**, 66, 1004.
- [18] W. Wang, W. Sun, L. Cao, M. Zheng, Y. Zhu, W. Li, C. Liu, X. Zhuang, J. Xing, H. Lu, T. Luo, N. Jin, *BMC Vet. Res.* **2019**, 15, 60.
- [19] G. Franzo, M. Legnardi, C. K. Hjulsager, F. Klaumann, L. E. Larsen, J. Segales, M. Drigo, *Transboundary Emerging Dis.* **2018**, 65, 602.
- [20] X. Ye, M. Berg, C. Fossum, P. Wallgren, A. L. Blomström, *Virus Genes* **2018**, 54, 466.
- [21] T. Stadejek, A. Woźniak, D. Miłek, K. Biernacka, *Transboundary Emerging Dis.* **2017**, 64, 1350.
- [22] P. J. Collins, J. McKillen, G. Allan, *Vet. Rec.* **2017**, 181, 599.
- [23] C. Tochetto, D. A. Lima, A. P. M. Varela, M. R. Loiko, W. P. Paim, C. M. Scheffer, J. I. Herpich, C. Cerva, C. Schmitt, S. P. Cibulski, A. C. Santos, F. Q. Mayer, P. M. Roehe, *Transboundary Emerging Dis.* **2018**, 65, 5.
- [24] Y. Zou, N. Zhang, J. Zhang, S. Zhang, Y. Jiang, D. Wang, Q. Tan, Y. Yang, N. Wang, *Arch. Virol.* **2018**, 163, 2841.
- [25] S. Faccini, I. Barbieri, A. Gilioli, G. Sala, L. R. Gibelli, A. Moreno, C. Sacchi, C. Rosignoli, G. Franzini, A. Nigrelli, *Transboundary Emerging Dis.* **2017**, 64, 1661.
- [26] S. L. Zhai, X. Zhou, H. Zhang, B. M. Hause, T. Lin, R. Liu, Q.-L. Chen, W.-K. Wei, D. H. Lv, X. H. Wen, F. Li, D. Wang, *Virol. J.* **2017**, 14, 222.
- [27] H. Shen, X. Liu, P. Zhang, L. Wang, Y. Liu, L. Zhang, P. Liang, C. Song, *Transboundary Emerging Dis.* **2018**, 65, 264.
- [28] T. G. Phan, F. Giannitti, S. Rossow, D. Marthaler, T. Knutson, L. Li, X. Deng, T. Resende, F. Vannucci, E. Delwart, *Virol. J.* **2016**, 13, 184.
- [29] B. Arruda, P. Piñeyro, R. Derscheid, B. Hause, E. Byers, K. Dion, D. Long, C. Sievers, J. Tangen, T. Williams, K. Schwartz, *Emerging Microbes Infect.* **2019**, 8, 684.
- [30] S. Zheng, X. Wu, L. Zhang, C. Xin, Y. Liu, J. Shi, Z. Peng, S. Xu, F. Fu, J. Yu, W. Sun, S. Xu, J. Li, J. Wang, *Transboundary Emerging Dis.* **2017**, 64, 1337.
- [31] G. Franzo, M. Legnardi, C. M. Tucciarone, M. Drigo, F. Klaumann, M. Sohrmann, J. Segalés, *Vet. Rec.* **2018**, 182, 83.
- [32] G. Franzo, C. M. Tucciarone, M. Drigo, M. Cecchinato, M. Martini, A. Mondin, M. L. Menandro, *Transboundary Emerging Dis.* **2018**, 65, 957.
- [33] F. Klaumann, A. Dias-Alves, O. Cabezon, G. Mentaberre, R. Castillo-Contreras, M. López-Béjar, E. Casas-Díaz, M. Sibila, F. Correa-Fiz, J. Segalés, *Transboundary Emerging Dis.* **2019**, 66, 91.
- [34] G. Franzo, L. Grassi, C. M. Tucciarone, M. Drigo, M. Martini, D. Pasotto, A. Mondin, M. L. Menandro, *Transboundary Emerging Dis.* **2019**, 66, 1548.
- [35] G. Li, W. He, H. Zhu, Y. Bi, R. Wang, G. Xing, C. Zhang, J. Zhou, K.-Y. Yuen, G. F. Gao, S. Su, *Adv. Sci.* **2018**, 5, 1800275.
- [36] F. Klaumann, G. Franzo, M. Sohrmann, F. Correa-Fiz, M. Drigo, J. I. Núñez, M. Sibila, J. Segalés, *Transboundary Emerging Dis.* **2018**, 65, 1290.
- [37] X. Fu, B. Fang, J. Ma, Y. Liu, D. Bu, P. Zhou, H. Wang, K. Jia, G. Zhang, *Transboundary Emerging Dis.* **2018**, 65, e296.
- [38] G. L. Saraiva, P. M. P. Vidigal, J. L. R. Fietto, G. C. Bressan, A. Silva Júnior, M. R. de Almeida, *Virus Genes* **2018**, 54, 376.
- [39] M. D. Hall, M. E. J. Woolhouse, A. Rambaut, *Virus Evol.* **2016**, 2, vew003.
- [40] J. Felsenstein, *Mol. Biol. Evol.* **2006**, 23, 691.
- [41] J. O. Wertheim, D. K. W. Chu, J. S. M. Peiris, S. L. Kosakovsky Pond, L. L. M. Poon, *J. Virol.* **2013**, 87, 7039.
- [42] E. C. Holmes, *J. Virol.* **2003**, 77, 3893.
- [43] P. Aiwasakun, A. Katzourakis, *J. Virol.* **2016**, 90, 7184.
- [44] S. Y. W. Ho, R. Lanfear, L. Bromham, M. J. Phillips, J. Soubrier, A. G. Rodrigo, A. Cooper, *Mol. Ecol.* **2011**, 20, 3087.
- [45] G. Franzo, M. Legnardi, C. K. Hjulsager, F. Klaumann, L. E. Larsen, J. Segales, M. Drigo, *Transboundary Emerging Dis.* **2018**, 65, 602.
- [46] R. Fux, C. Söckler, E. K. Link, C. Renken, R. Krejci, G. Sutter, M. Ritzmann, M. Eddicks, *Virol. J.* **2018**, 15, 25.
- [47] M. I. Nelson, C. Viboud, A. L. Vincent, M. R. Culhane, S. E. Detmer, D. E. Wentworth, A. Rambaut, M. A. Suchard, E. C. Holmes, P. Lemey, *Nat. Commun.* **2015**, 6, 6696.
- [48] G. Franzo, P. Massi, T. M. Tucciarone, I. Barbieri, G. Tosi, L. Fiorentini, M. Ciccozzi, A. Lavazza, M. Cecchinato, A. Moreno, *PLoS One* **2017**, 12, e0184401.
- [49] G. Franzo, M. Cecchinato, G. Tosi, L. Fiorentini, F. Faccin, C. M. Tucciarone, T. Trogu, I. Barbieri, P. Massi, A. Moreno, *PLoS One* **2018**, 13, e0203513.
- [50] C. Firth, M. A. Charleston, S. Duffy, B. Shapiro, E. C. Holmes, *J. Virol.* **2009**, 83, 12813.
- [51] G. Franzo, C. M. Tucciarone, M. Cecchinato, M. Drigo, *Sci. Rep.* **2016**, 6, 39458.
- [52] K. Katoh, D. Standley, *Mol. Biol. Evol.* **2013**, 30, 772.
- [53] F. Abascal, R. Zardoya, M. J. Telford, *Nucleic Acids Res.* **2010**, 38, W7.
- [54] D. P. Martin, B. Murrell, M. Golden, A. Khoosal, B. Muhire, *Virus Evol.* **2015**, 1, vev003.
- [55] A. J. Drummond, M. A. Suchard, D. Xie, A. Rambaut, *Mol. Biol. Evol.* **2012**, 29, 1969.
- [56] P. Lemey, A. Rambaut, A. J. Drummond, M. A. Suchard, *PLoS Comput. Biol.* **2009**, 5, e1000520.

- [57] G. Baele, P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard, A. V. Alekseyenko, *Mol. Biol. Evol.* **2012**, 29, 2157.
- [58] A. J. Drummond, S. Y. W. W. Ho, M. J. Phillips, A. Rambaut, *PLoS Biol.* **2006**, 4, e88.
- [59] A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, *Mol. Biol. Evol.* **2005**, 22, 1185.
- [60] A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, *Syst. Biol.* **2018**, 67, 901.
- [61] F. Bielejec, G. Baele, B. Vrancken, M. A. Suchard, A. Rambaut, P. Lemey, *Mol. Biol. Evol.* **2016**, 33, 2167.
- [62] S. L. K. Pond, S. D. W. Frost, *Mol. Biol. Evol.* **2005**, 22, 1208.
- [63] B. Murrell, S. Moola, A. Mabona, T. Weighill, D. Sheward, S. L. Kosakovsky Pond, K. Scheffler, *Mol. Biol. Evol.* **2013**, 30, 1196.
- [64] S. L. K. Pond, S. D. W. Frost, S. V. Muse, *Bioinformatics* **2005**, 21, 676.
- [65] S. L. K. Pond, B. Murrell, M. Fourment, S. D. W. Frost, W. Delpont, K. Scheffler, *Mol. Biol. Evol.* **2011**, 28, 3033.
- [66] B. Murrell, J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, S. L. K. Pond, *PLoS Genet.* **2012**, 8, e1002764.
- [67] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**, 10, 845.