Emerging novel porcine parvoviruses in Europe: origin, evolution, phylodynamics and phylogeography

Dániel Cadar,¹,² Márta Lőrincz,² Timea Kiss,¹ Dinko Novosei,³ Katarzyna Podgorska,⁴ Zsolt Becskei,⁵ Tamás Tuboly² and Attila Cságola²

¹Department of Infectious Diseases, Faculty of Veterinary Medicine, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania
²Department of Microbiology and Infectious Diseases, Faculty of Veterinary Science, Szent István University, Budapest, Hungary
³Department of Pathology, Croatian Veterinary Institute, Zagreb, Croatia
⁴National Veterinary Research Institute, Department of Swine Diseases, Pulawy, Poland
⁵Department of Animal Breeding and Genetics, Faculty of Veterinary Medicine, University of Belgrade, Belgrade, Serbia

To elucidate the spatiotemporal phylodynamics, dispersion and evolutionary processes underlying the emergence of novel porcine parvovirus 2 (PPV2), PPV3 and PPV4 species, we analysed all available complete capsid genes, together with ours, obtained in Europe. Bayesian phylogeography indicates that Romania (PPV2 and PPV4) and Croatia (PPV3) are the most likely ancestral areas from which PPVs have subsequently spread to other European countries and regions. The timescale of our reconstruction supported a relatively recent history of the currently circulating novel PPV species (1920s to 1980s) in the domestic or sylvatic host. While PPV2 strains exhibited a large genetic exchange characterized by significant recombination and gene flow between distinct regions and hosts, PPV3 and PPV4 showed a diversification reflected by the accumulation of geographically structured polymorphisms. The RNA-like evolutionary rates detected inter- and intrahost recombination and the positive selection sites provided evidence that the PPV2–4 capsid gene plays a prominent role in host adaptation.

Recently, several new members of the subfamily Parvovirinae have been discovered in animals, particularly in domestic pigs. The first such novel parovirus, provisionally designated porcine parovirus 2 (PPV2), has been described worldwide (Hijikata et al., 2001; Wang et al., 2010; Xiao et al., 2013a). PPV3 (Cheung et al., 2010), initially designated porcine hokovirus (Lau et al., 2008) and also known as porcine partetravirus (Tse et al., 2011), porcine PARV4-like virus (Szelei et al., 2010) and porcine PARV4 (Xiao et al., 2012), shows a global presence in domestic pigs. Among newly described paroviruses of swine, PPV4, a novel genetically divergent member of the subfamily Parvovirinae was discovered initially in the USA in porcine circovirus-associated disease (PCVAD)-affected pigs (Cheung et al., 2010) and subsequently reported in Asia, Europe and Africa (Huang et al., 2010; Zhang et al., 2011; Cságola et al., 2012; Cadar et al., 2013; Ndze et al., 2013).

With little knowledge on their pathogenicity, these emerging paroviruses are currently receiving more attention due to the tentative association with diseases like PCVAD (Xiao et al., 2012; Li et al., 2013; Opriessnig et al., 2013) or ‘high fever disease’ (Wang et al., 2010), which highlights the need to investigate the molecular evolution, epidemiology and genetic diversity of these novel PPVs in Suidae hosts. The aim of this study was to reconstruct the phylogeny and evolution of PPV2–4 on a spatiotemporal scale in order to detect the driving forces shaping their evolution, and estimate the time of origin and patterns of geographical dispersal of the different strains in general and in Europe in particular. These data will help to fill several gaps in the understanding of dispersion, evolution and phylodynamics of these emerging novel porcine paroviruses.

To detect the presence of PPV2–4 in different regions of Europe, tissues and serum samples collected (2006–2011)
from domestic pigs from Croatia (n=89), Poland (n=185) and Serbia (n=78) were used in this study. In addition, Hungarian and Romanian samples from our previous studies (Csághola et al., 2012; Cadar et al., 2013) were also used. The detection and amplification of the complete PPV2–4 VP gene sequences were performed using previously described specific PCR protocols (Cadar et al., 2011, 2013; Csághola et al., 2012). We retrieved from GenBank all available PPV2–4 complete VP gene sequences. Their accession numbers and other additional information including PPV2–4 sequences of this study are listed in Table S1 (available in JGV Online).

To reconstruct the evolutionary history of each dataset (global and European), maximum likelihoods (MLs) using Treefinder (Jobb et al., 2004) and the Bayesian Markov chain Monte Carlo (MCMC) method implemented in the BEAST v1.6.2 package (Drummond & Rambaut, 2007) were applied. Moreover, we estimated the rate of evolutionary change (substitutions per site per year), the time to most recent common ancestor (tMRCA) and $Ne.g$ (i.e. the effective number of transmission events, $Ne$, times the generation time, $g$, of the pathogen) under a Bayesian skyline coalescent tree prior (Drummond et al., 2005). The extent and pattern of geographical structure in PPV2–4 VP gene phylogenies were assessed using the program BaTS (Parker et al., 2008).

We used a timescaled phylogeny to reconstruct the PPV2–4 spread patterns using a standard continuous-time Markov chain process over discrete sampling locations with the Bayesian stochastic search variable selection (BSSVS) model, which allows the diffusion rates to be zero with some positive prior probability (Lemey et al., 2009). The putative spatiotemporal pattern of PPV2–4 spread in

**Fig. 1.** Bayesian MCC trees that summarized global PPV2 (a), PPV3 (b) and PPV4 (c) complete VP gene datasets were generated using geospatial Bayesian analysis. We coloured branches according to the most probable location state of their descendent nodes. The location s.p.p. distributions of ancestral nodes corresponding to PPV2–4 are on the left. ML bootstrap scores (>70 %) and BPP (>90 %) are shown above the branches. Time is reported in the axis below the tree. The BSPs including ten coalescent interval groups obtained by analysing the PPV2–4 sequences sampled at different times are shown in the background. The thick dashed line indicates the median estimates, and the light blue area shows the 95 % highest posterior density (HPD). The x-axis is the timescale in years, and the y-axis is a logarithmic scale of $Ne.g$ (where $Ne$ is the effective population size and $g$ is the generation time).
Europe was analysed using the spread (Bielejec et al., 2011) program.

The selection pressures analyses were performed using the slac, fel, rel (Kosakovsky Pond & Frost, 2005) and fubar (Murrell et al., 2013) methods of the HyPhy package. In order to detect both episodic and pervasive positive selection at the level of individual sites, the mixed effects model of evolution (Murrell et al., 2012) method was implemented. We used the automated algorithms contained within the rdp3 program (Martin et al., 2010) to screen for putative recombination events. The mosaic structures of any recombinants were also inferred by means of bootscanning using gard (Kosakovsky Pond et al., 2006). The SplitsTree program v4.12.3 (Huson & Bryant, 2006) was also employed to confirm the phylogenetic relationship of the recombiant samples of each dataset.

Fig. 1 shows the Bayesian MCMC trees, the estimation of viral population dynamics and the tMRCA of global PPV2–4 datasets. The root of the PPV2 tree had a tMRCA of 86 years before present (y.b.p.). From this ancestor, three main clades (A–C) were identified, having tMRCA of 63, 61 and 65 y.b.p., respectively. The phylogeographic reconstruction clearly showed that all clades and subclades had a MRCA location in Romania [state posterior probability (s.p.p.)=0.24] (Fig. 1a). While the root of the PPV3 tree had a tMRCA of 81 y.b.p., the seven major clades identified had a mean ranging between 23 and 56 y.b.p. The phylogeographic reconstruction was unable to identify a single location for the root of the tree of the PPV3 global dataset, due to almost identical s.p.p. for two localities (China and UK) (Fig. 1b). The PPV4 phylogeny showed a more recent tMRCA (30 y.b.p.) in comparison with the other two PPV species studied (Fig. 1c and Table 1). Five distinct clades were identified, which diverged at about 8–21 y.b.p. The AI and PS statistics using the program BaTS indicated that there was very strong geographical clustering of PPV3 (except in Germany and Poland) and PPV4 (except in Croatia) strains by country of origin (P<0.05), reflecting a significant population subdivision, except for the PPV2 dataset, for which there was evidence of significant gene flow between distinct regions (P>0.05) (Table S3).

Fig. 2 illustrates the Bayesian MCC tree for PPV2–4 in Europe, and the colour of each lineage and internal node represents its most probable geographical locality. The spatial reconstruction suggested that PPV2 and PPV4 in Europe most likely originated in Romania (highest s.p.p. values), but we could not accurately pinpoint the geographical origin of PPV3 due to almost identical s.p.p. for two localities (Romania and Croatia) (Fig. 2b). Thus, investigation of the diffusion patterns showed a well-supported connection [Bayes factor (BF) >3] for Croatia using the BF test under BSSVS analysis (Fig. S3b). Our analysis recognized six, seven and four main clades corresponding to PPV2, PPV3 and PPV4 European datasets having the root of the tree tMRCA of 71, 45 and 32 y.b.p. (Fig. 2). Bayesian skyline plot (BSP) analysis of the PPV2 European dataset showed similar Ne.g values and dynamics as observed for the global dataset (Figs 1a and 2a). PPV3 dynamics showed an increased growth phase that started around the mid-1960s, with a plateau from the 1970s to 1990s associated with a decreased phase, until the mid-2000s followed by a phase when Ne.g values remained constant and still persist (Fig. 2b). PPV4 population dynamics are characterized by an initial constant phase from the 1980s to 1990s, followed by a decreasing level of Ne.g values towards the 2000s, increased with a stabilized value persisting until today (Fig. 2c). The histories of dispersal of PPV2–4 among European countries reconstructed using the equal rates model
are shown in Fig. S2. Fig. S3 shows only those linkages that are statistically well supported (i.e. non-zero rates supported by a BF >3). Our reconstruction of discrete ancestral states (Fig. 2) had the higher BPP values for Romania (PPV2, PPV4) and Croatia (PPV3).

Table S2 shows the prevalence of PPV2–4 infections detected in this study and previously reported worldwide. Multiple nucleotide alignments of PPV3 datasets revealed one codon deletion (position 714) in two strains from Serbia (13-RS, 18-RS). The overall $d_S/d_K$ ratios in VPs were 0.134 (PPV2), 0.176 (PPV3) and 0.234 (PPV4), indicating that most sites are subject to strong purifying selection. However, positive selection sites were detected in each parvovirus species under the five implemented algorithms (Table S5). Evolutionary fingerprints of PPV2–4 VP gene alignments clearly reflected and supported the presence of positively selected individual sites and purifying selection (Fig. S4). The paraplyetic Pro$_{99}$Ala and Ala$_{100}$Val/His mutations located in the C terminus of the PPV2 VP gene appeared only in European domestic and sylvatic strains, while the paraplyetic Ser$_{101}$Thr mutation was detected worldwide. The geographical structure of the Pro$_{233}$Ser mutation indicated the presence of this amino acid replacement only in wild boar origin PPV3 VP gene sequences, while the PPV4 Gln$_{722}$Pro mutation appeared only in domestic and sylvatic strains from Romania and domestic strains from China. Strong recombination signals were detected only in the PPV2 dataset, and confirmed by GARD and SplitTree network analysis (Table S4, Fig. S1). Almost all PPV2 strains from the European dataset exhibited potential recombination events. These events were detected both within and between clades and within and between countries, and also between domestic and sylvatic (wild boars) strains (Table S4).

In this study, we sought to elucidate the possible origin, spatiotemporal dynamics and factors that shape the evolution of novel PPV2–4 species in general, and in Europe especially. Although several distinct clades in the phylogenetic trees of PPV2–4 were detected, there is so far

Fig. 2. Bayesian MCC trees summarized for European PPV2 (a), PPV3 (b) and PPV4 (c) complete VP gene datasets were generated using geospatial Bayesian analysis. Branches are coloured according to the most probable location state of their descendent nodes. The location s.p.p. distributions of ancestral nodes corresponding to PPV2–4 are on the left. ML bootstrap scores (>70%) and BPP (>90%) are shown above the branches. Time is reported in the axis below the tree. The BSPs, including ten coalescent interval groups obtained by analysing the PPV2–4 sequences sampled at different times, are shown in the background. The thick dashed line indicates the median estimates, and the light blue area shows the 95 % HPD. The x-axis is the timescale in years, and the y-axis is a logarithmic scale of $Ne.g$ (where $Ne$ is the effective population size and $g$ is the generation time).
no available classification methodology for PPV2–4. Very recently, Xiao et al. (2013b) proposed three new tentatively designated genera within the subfamily Parovirinae, from which the tentative genus PARV4-like included PPV2 and PPV3, and a currently unnamed novel clade 2 for PPV4.

Our study is the first attempt to reconstruct the phylodynamics and phylogeography history of PPV2–4 species. Although PPV2–4 have received more attention only recently, our estimates suggest that they must have been around at least since the 1920s (PPV2), 1930s (PPV3) and 1980s (PPV4), respectively, in the domestic or sylvatic host. Recently, it has been shown that PPV1 originated approximately 120 years ago, with the main divergence occurring in the past 20–60 years (Cadar et al., 2012). This suggests that the novel PPV2–4 and PPV1 have a relatively recent and similar evolutionary history.

These results suggest the occurrence of interspecies transmissions of PPV2–4 between domestic and sylvatic hosts and are supported by our previous study (Cadar et al., 2013), but they do not allow us to pinpoint the host of origin for each PPV species, and by which the viruses were introduced into specific geographical regions. An extended surveillance, especially in wild boars among these regions, will be necessary to answer this question. The spread of PPV2 and PPV4 in Europe indicates Romania as the initial source of their diffusion. For PPV3, we were unable to detect the most probable ancestral location due to the almost identical state probability for two basal localities, but based on BSSVS analysis, the diffusion patterns showed well-supported connections (BF >3) for Croatia as the likely source of PPV3 diffusion in Europe. Although there is clear evidence of gene flow of PPV2 among countries within Europe, and worldwide as well, BaTS analysis also detected strong structuring of the PPV3 and PPV4 phylogeny by country, suggesting that in situ evolution also plays an important role in the maintenance of these viruses worldwide. Indeed, few genetic clusters included viruses isolated from single geographical areas, suggesting the existence of epidemiological and commercial connections among different countries.

Evolutionary rates have been calculated for the VP gene of several ssDNA viruses and range from more than $1.2 \times 10^{-3}$ to as low as $9.4 \times 10^{-5}$ substitutions per site per year (Shackelton et al., 2005; Duffy & Holmes, 2008; Hoelzer et al., 2008; Firth et al., 2009; Streck et al., 2011; Cadar et al., 2012, 2013). The evolutionary rates of the VP genes of PPV2–4 ($2.57\text{--}3.38 \times 10^{-4}$ substitutions per site per year) fall in the middle of this spectrum and could maintain evolutionary dynamics of these viruses closer to those of ssRNA viruses than to those of dsDNA viruses. A larger sample of viruses from both hosts is needed to confirm whether the immune escape and tropism shifts are responsible for the high substitution rates. The overall low $d_S/d_\text{N}$ ratio in these novel PPVs indicates that most amino acid residues are subject to purifying selection with adaptive evolution restricted to specific residues within VP genes. It is unknown if these mutations are a consequence of the mechanisms of antigenic escape or further adaptation to the host, but they may affect the antigenic profile and confer an evolutionary advantage to the viruses.

It has been shown that parvoviruses are able to emerge in new hosts (Parrish & Kawaoka, 2005; Shackelton et al., 2005). Recombination events of PPV2 VP genes, including the likely intra- and interspecies and within and between country strain recombinations, are reported in this study. All lineages included closely related PPV2 variants from distant locations and different host species together with the radiation pattern observed for diversification of PPV2 strains, suggesting that genetic flow has occurred. Our analysis provides novel data about the spatiotemporal phylodynamics, dispersion and evolutionary scenario shaping these emerging PPVs in general and in Europe in particular, and supports the notion that Romania (PPV2 and PPV4) and Croatia (PPV3) are possible sources of PPVs that have subsequently spread to other countries of the continent. While circulating PPV2 strains exhibited a large genetic exchange characterized by significant gene flow between distinct regions, PPV3 and PPV4 showed a diversification of viral lineages reflected by the accumulation of geographically structured polymorphisms. The present study also provides data on the evolutionary dynamics of PPVs, serving as a useful substrate for further studies regarding the effect of specific mutations, geographical genetic diversity and the possible implication or association as a co-factor in the development of swine diseases such as PCVAD.

Acknowledgements

Research was funded by OTKA PD/100405, a KTIA-AIK grant of the Hungarian Government and a János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Special thanks go to the Balassi Institute-Hungarian Scholarship Board Office for financial support. We thank Herbik Józsefné for excellent technical assistance.

References


