Molecular Phylogenetics and Evolution 66 (2013) 243–253

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Capsid protein evolution and comparative phylogeny of novel porcine parvoviruses

Dániel Cadar [a,*], Attila Cságola [b], Timea Kiss [a], Tamás Tuboly [b]

[a] Department of Infectious Diseases, Faculty of Veterinary Medicine, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania
[b] Department of Microbiology and Infectious Diseases, Faculty of Veterinary Science, Szent István University, Budapest, Hungary

ARTICLE INFO

Article history:
Received 2 July 2012
Revised 24 September 2012
Accepted 25 September 2012
Available online 5 October 2012

Keywords:
Porcine parvoviruses
Phylogeny
Evolution
Recombination
Wild boar
Domestic pig

ABSTRACT

In addition to the well known "classical" porcine parvovirus (PPV1; responsible for reproductive failure of susceptible sows) several new porcine parvoviruses have been recognized (PPV2, PPV3 and PPV4) in recent years. The genetic variation, characteristics and evolutionary factors shaping these novel PPVs were studied by comparing the complete capsid (*cap*) genes of PPVs from domestic pigs and wild boars. Using Bayesian coalescent methods we estimated the rate of nucleotide substitution for PPV2, PPV3 and PPV4 to be of the order of 3.86×10^{-4}–8.23×10^{-4} subs site^{-1} year^{-1}, similar to those commonly measured for RNA viruses, although this rate in case of PPV2 is probably influenced by frequent recombination events. Given such rapid evolutionary dynamics, it is likely that novel PPVs will continue to improve their capacity to spread among *Suidae* hosts worldwide. The mean time to the most recent common ancestor for the sampled genetic diversity of the newly discovered porcine parvoviruses was estimated. The results indicated that novel PPVs originated within approximately the last 70 years. Incongruent phylogenetic relationships of several strains suggested recombination events supported by several recombination-detecting methods and by split-decomposition phylogenetic networks. Analyses of the selective constraints acting on each codon suggest that some regions of PPV *cap* genes were under positive selection. This study showed that inter- and intraspecies recombination and diversifying selection pressures are prevalent across the *cap* genes of novel PPVs, and beside host switching and gene flow are important driving forces of their evolution and may be significant factors in the emergence of new viral variants.

1. Introduction

Parvoviridae is a large family of linear non-enveloped DNA viruses of approximately 4–6 kilobases (kb) that infect a wide range of vertebrates and intervertebrates. The clinical significance of parvoviruses is well known, some of them being responsible for severe clinical signs (porcine parvovirus, human parvovirus B19, human bocavirus, canine parvovirus-2) or only mild or subclinical infections (Shade et al., 1986; Allander et al., 2005; Sloots et al., 2006; Decaro and Buonavoglia, 2012). According to the latest taxonomic report of the International Committee on Taxonomy of Viruses (King et al., 2011), the *Parvoviridae* family is divided into two subfamilies: the subfamily *Parvovirinae*, whose members infect vertebrates, and includes five genera, namely *Parvovirus*, *Erythrovirus*, *Dependovirus*, *Amdovirus* and *Bocavirus*, whereas members of the subfamily *Densovirinae* are present in insects and other arthropods. Recently, due to sequence-independent PCR methods and high-throughput sequencing, several new porcine parvoviruses have been detected. The first such novel parvovirus, provisionally designated as porcine parvovirus 2 (PPV2), was discovered in swine sera during a hepatitis E survey in Myanmar (Hijikata et al., 2001). The 5-kb genome shared similarity with Muscovy duck parvovirus and bovine parvovirus 3 but was phylogenetically distinct and not closely related to any other known members of *Parvovirinae*. Several years later in China, genetically highly similar parvoviruses were discovered in serum samples of pigs affected with "high fever" disease (PRRS associated disease) and PMWS (Wang et al., 2010). They formed a distinct sublineage (proposed as *Cnvirus*, Wang et al., 2010) within the subfamily *Parvovirinae*. Studies show that the prevalence of PPV2 infection in pigs is relatively low, between 6.4% and 9.66% (Wang et al., 2010; Cságola et al., 2012). In 2008, a new porcine parvovirus, porcine hokovirus (PHoV) belonging to the newly proposed genus *Hokovirus* was identified in Hong Kong (Lau et al., 2008). The analysis of genome sequences showed that PHoV strains were closely related to human parvovirus 4 (PARV4) and bovine hokovirus (BHoV), and formed a distinct cluster within the family *Parvoviridae* (Lau et al., 2008). A very recent classification of porcine hokovirus as porcine parvovirus 3 (PPV3) was proposed by Cheung et al. (2010). Recent studies performed on PPV3 infection in domestic pigs in Great Britain (Szelei et al., 2010), Hungary (Cságola et al., 2012) and in wild boars in Germany (Adlhoch et al., 2010) and Romania (Cadar et al., 2011b) suggest a

* Corresponding author. Address: Department of Infectious Diseases, Faculty of Veterinary Medicine, University of Agricultural Sciences and Veterinary Medicine, Manastur 3-5, Box 400372, Cluj-Napoca, Romania. Fax: +40 264 593 792.
E-mail address: dimmuvet@yahoo.com (D. Cadar).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.09.030

widespread distribution of this virus with a very high prevalence mainly in wild boars (50.54%), and the overall presence of PPV3 in adults indicated a persistent infection. Another novel porcine parvovirus, PPV4 was detected in 2005 from PCVAD (porcine circovirus associated disease) affected swine herds in North Carolina, USA (Cheung et al., 2010) and subsequently reported in China (Huang et al., 2010; Zhang et al., 2011). The positivity rates of PPV4 infection among swine herds remains low worldwide, between 2.09% and 6.6% (Cheung et al., 2010; Zhang et al., 2011; Cságola et al., 2012). Each of these novel porcine parvoviruses shows a genome organization characteristic to the members of the subfamily *Parvovirinae*, encoding two major open reading frames (ORFs), the non-structural protein (NS) and the capsid proteins (VP1 and VP2), except PPV4 which has an additional ORF3, located between ORF1 and ORF2, characteristic to members of the *Bocavirus* genus. The nucleoprotein encoding ORF3 of PPV4 is, however, a quite distinct form and shares limited amino acid identity with the ORF3 encoded proteins of the members of the *Bocavirus* genus (Cheung et al., 2010; Huang et al., 2010). In a recent study, Zhang et al. (2011) proposed a new designation of PPV4, namely porcine bocavirus 2 (PoBoV2). This study showed high phylogenetic distances between PPV4 and known porcine bocaviruses based on the deduced amino acid sequence of ORF2. Until now, there have not been enough data or consensus on definition or classification of this virus, therefore, we consider it appropriate to keep the original name, proposed by Cheung et al. (2010), namely porcine parvovirus 4 (PPV4). There is increasing evidence suggesting that various evolutionary mechanisms such as virus-host co-evolution, inter- or intraspecies transmission, recombination and positive selection play important roles in the evolution of parvoviruses (Lukashov and Goudsmit, 2001; Shackelton et al., 2005, 2007; Streck et al., 2011; Cadar et al., 2012). However, the molecular epidemiology, genetic characteristics, evolution or even the presence of the newly identified PPVs in domestic pigs or wild boars are still unknown. In this study we report for the first time the presence of PPV2 and PPV4 in wild boars and the complete *cap* gene sequences of the novel PPVs in Europe. This is the first in-depth molecular characterization and analysis of complete *cap* gene sequences of novel PPVs. The availability of these genes allowed us to conduct a comprehensive phylogenetic and comparative analysis of the genetic characteristics of the *cap* gene and the evolutionary processes that have influenced the diversification of these novel PPVs in both domestic and wild hosts.

2. Materials and methods

2.1. Sample collection and viral DNA extraction

Tissue samples of shot free-living wild boars originated from the Western region of Romania (Transylvania) collected during 2006–2011 (Cadar et al., 2010, 2011a, 2011b) were included in this work. All 842 wild boar samples were grouped according to their geographic origin, hunting grounds and the date of collection. For comparative genetic characterization and phylogenetic analysis we used 120 domestic pig tissue samples of 10 different farms of the same sampling regions as for wild boars. Total viral DNA was extracted from pooled tissue samples separately for each animal using Viral Gene-spin™ Viral DNA/RNA Extraction Kit (Intron Biotechnology Inc., Korea) according to the manufacturer's instructions.

2.2. Detection and amplification of PPV2, PPV3 and PPV4 cap gene

The initial detection of PPV2, PPV3 and PPV4 from extracted DNA samples of domestic pig and wild boar tissues was performed using previously described specific PCR methods (Cadar et al., 2011b; Cságola et al., 2012). Based on geographic origin and the time of collection, we selected 11 PPV2 (6 wild boar and 5 domestic pig), 9 PPV3 (8 + 1) and 10 PPV4 (6 + 4) positive samples and the complete *cap* genes of these were amplified using the primers listed in Supplementary Table S1. In case of wild boar PPV3, we used already amplified and published sequences from our previous study (Cadar et al., 2011b). The 50 μL reaction mixture for the PCR amplification consisted of 10 μL 10× My*Taq* Red Buffer (Bioline GmbH, Germany), 0.2 μL of My*Taq* DNA Polymerase (Bioline GmbH, Germany), 0.5 μL of 25 μM of forward and reverse primers, 1 μL of DNA template and ddH_2O up to 50 μL. The PCR conditions used for each novel PPV included an initial denaturation at 94 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 60 °C (PPV2), 58 °C (PPV3), 57 °C (PPV4) for 45 s and extension at 72 °C for 3 min, and a final elongation step at 72 °C for 7 min. PCR products were visualized using 1.5% agarose gels stained with GR Safe Nucleic Acid Stain (Excellgen Inc., USA) and Dark Reader blue light transilluminator (Chemical Research, Inc., USA). The amplified ORF2 fragments were sequenced (Macrogen Europe Inc., Netherlands) at least twice in each direction using the PCR primers shown in Supplementary Table S1.

2.3. Nucleotide and protein sequence analysis

The obtained sequences were evaluated and assembled with Geneious v4.8.5 (Biomatters, Ltd.). Four PPV2, 29 PPV3 and 10 PPV4 that represent all of the available complete novel PPV capsid gene sequences were retrieved from NCBI (http://www.ncbi.nlm.nih.gov) and subjected to phylogenetic analysis. Their GenBank accession numbers and other additional information including those of the PPV2, PPV3 and PPV4 sequences of this study are listed in Supplementary Table S2. The complete *cap* gene nucleotide sequences obtained were compared to sequences available in the GenBank and alignment of the sequences was carried out using the ClustalW included in BioEdit v7.0.9. (Hall, 1999). Translation of the nucleotide sequences into amino acid sequences and the degree of similarity among the sequences at both nucleotide and amino acid levels were determined by using BioEdit. Searches for DNA, and protein motifs present in full-length *cap* genes of PPVs were performed with Geneious v4.8.5.

2.4. Phylogenetic analysis

The evolutionary relationships of novel porcine parvoviruses (PPV2, PPV3, PPV4) with the "classical" PPV1, between and within domestic pig and wild boar origin were determined using three different approaches, a Bayesian phylogenetic, maximum likelihood and a split network analysis. A multiple sequence alignment of novel PPVs, including PPV1 sequences, was performed for complete *cap* gene data sets. For each phylogenetic analysis we used jModeltest (Guindon and Gascuel, 2003; Posada, 2008) to determine the model of nucleotide substitution that best fits the data. The Akaike's Information Criterion was chosen as the model selection framework and Hasegawa–Kishino–Yano model of sequence evolution with a proportion of invariant sites and gamma distributed rate heterogeneity (HKY + I + Γ) as the best model. Bayesian inference analyses were performed using the Markov chain Monte Carlo (MCMC) coalescent approach implemented in BEAST v1.6.2 package (Drummond and Rambaut, 2007) based on the selected nucleotide substitution model obtained. Three independent runs of 5×10^6 generations each were performed to estimate the posterior probability distribution. Topologies were sampled every 1000 generations to ensure the independence of successive trees. The first 10% of each run was discarded as burn-in and to ensure that convergence of the Markov chain was assessed, we used the