Virus Research 178 (2013) 392–397

Contents lists available at ScienceDirect

Virus Research

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

Phylogeny and evolution of porcine parvovirus

Xiaofeng Ren[a,*], Ye Tao[a], Jin Cui[a], Siqingaowa Suo[a], Yingying Cong[a], Peter Tijssen[b]

[a] Department of Preventive Medicine, College of Veterinary Medicine, Northeast Agricultural University, 59 Mucai Street, Xiangfang District, Harbin 150030, China
[b] INRS-Institut Armand-Frappier, Université du Québec, 531 boulevard des Prairies, Laval, QC H7V 1B7, Canada

ARTICLE INFO

Article history:
Received 23 July 2013
Received in revised form 4 September 2013
Accepted 5 September 2013
Available online 17 September 2013

Keywords:
Porcine parvovirus
Nucleotide substitution rate
Recombination
Selection pressure

ABSTRACT

Porcine parvovirus (PPV), a member of the genus *Parvovirus*, family *Parvoviridae*, is a significant causative agent in porcine reproductive failure, causing serious economic losses in the swine industry. Previous phylogenetic studies based on the NS1 or VP2 genes indicated that current PPV strains diverged 30 years ago and that VP2 was under neutral or positive selection. Our analysis of NS1, VP2 and complete ORFs indicated that the most recent common ancestor of PPV strains existed about 250 years ago and that the 127-nt repeat in the 3′NTR was present in viruses of some subclades that evolved about 80 years ago. Nucleotide substitution rates of NS1 and VP2 genes were 3.03×10^{-5} and 1.07×10^{-4}, respectively. Both the NS1 and VP2 proteins were under purifying selection and recombination did not contribute to the genetic diversity of PPV. As expected, surface amino acids are hydrophilic and make up the majority of mutations in the VP2 protein; residues in VP2 interfaces were substituted gradually, often in conjunction with complementary substitutions in the neighboring VP2.

1. Introduction

Porcine parvovirus (PPV) is one of the major etiological agents responsible for porcine reproductive failure. This SMEDI syndrome is characterized by infertility, early embryonic and fetal death, mummified fetuses, stillbirth, and delayed return to oestrus (Szelei et al., 2006; Truyen and Streck, 2012). PPV is an autonomous parvovirus belonging to the genus *Parvovirus*, a group of viruses that also infect rodents and carnivores, and belongs to the *Parvovirinae* subfamily within the *Parvoviridae* family (Tijssen et al., 2011).

PPV has a negative, single-stranded DNA of about 5 kb (Molitor et al., 1983) with at least two open reading frames (ORF) (Bergeron et al., 1993, 1996). The major structural protein, VP2 is the main target for neutralizing antibodies to PPV (Kamstrup et al., 1998; Martinez et al., 1992). The near-atomic structure of the viral capsid has been resolved using X-ray crystallography, and was found to be similar to that of minute virus of mice, feline parvovirus, and canine parvovirus (Simpson et al., 2002).

Porcine parvovirus (PPV), first isolated from sows in Germany (Mayr et al., 1968), has been found to occur worldwide (Cui et al., 2012; Kim and Chae, 2004; Mengeling et al., 1991). In recent years, new strains, such as ZJ and BQ, have been isolated from pigs in China (Cui et al., 2009). Although inactivated and attenuated vaccines are used widely in the swine industry, PPV outbreaks occurred in many regions, causing serious economic losses in the industry.

In this study, we investigated the PPV ORFs to understand the most recent common ancestor of PPV, the nucleotide substitution rates in the different ORFs, the existence of recombination, and the type of selection pressure.

2. Materials and methods

2.1. Sequence data

Nucleotide sequences of 46 PPV genomes were obtained from GenBank (as of November, 2011). These sequences contained 31 (nearly) complete genomes and a Kilham Rat Virus complete genome, also from the *Parvovirus* genus and used as the outgroup, as well as sequences of 39 VP2 and 38 NS1 genes. Eleven genomes had complete or partial repeats downstream of the VP gene, whereas 22 lacked these repeats. The PPV2010 strain was isolated by our lab and has been submitted to GenBank (accession number JN872448). Strain names and isolation locations are summarized in Supplementary Table 1. For each data set, sequences were aligned using MUSCLE (Edgar, 2004) and then manually adjusted using Se–Al (available at http://tree.bio.ed.ac.uk/software/seal/) to preserve reading frame integrity.

2.2. Estimation of substitution rates and TMRCA

Rates of nucleotide substitutions per site per year and the time to most recent common ancestor (TMRCA) were estimated using the Bayesian Markov Chain Monte Carlo (MCMC) method available in the BEAST package v1.5.3 (Drummond and Rambaut, 2007).

* Corresponding author. Tel.: +86 451 55191974.
E-mail addresses: renxf@neau.edu.cn, rxfemail@gmail.com (X. Ren).

0168-1702/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2013.09.014

The best-fit model of nucleotide substitution for each data set was determined using Model Test 3.7 (Posada and Crandall, 1998). The relaxed uncorrelated lognormal molecular clock and the Bayesian skyline plot (BSP) models were chosen as coalescent priors. In each case, chain lengths were run for a sufficient time to ensure effective sample size (ESS) above 200, with the first 10% of MCMC chains removed as burn-in. Uncertainty in parameter estimates is reflected as the 95% highest probability density (HPD) values. Maximum clade credibility trees were summarized with TreeAnnotator, and were depicted using FigTree (Drummond and Rambaut, 2007).

2.3. Recombination analysis

For detection of potential recombination events, we aligned the complete ORFs of 31 PPV strains. Identification of potential recombinant breakpoints was performed using the GARD (http://www.datamonkey.org/GARD/) and Recombination Detection Program (RDP) 4.13 (Martin et al., 2010) including RDP, GENECONV, BOOTSCAN, MaxChi, CHIMAERA, SISCAN, PhylPro, LARD, and 3seq algorithms (Cui et al., 2011; Martin et al., 2010). Phylogenetic incongruence was used as the gold-standard bioinformatics method for the identification of recombinant sequences (Boni et al., 2010).

2.4. Selection pressures

To investigate selection pressures acting on the NS1 and VP2 genes of PPV, we estimated average numbers of nonsynonymous substitutions (d*N*) and synonymous substitutions (d*S*) per site (d*N*/d*S* ratio) using the single likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), and random-effects likelihood (REL) methods implemented in the HYPHY platform (Kosakovsky Pond and Frost, 2005), accessed through the DataMonkey web-server (http://www.datamonkey.org). In each method, the confidence level was set to a Bayes factor of 50 and a *P* value of 0.05.

2.5. Amino acid mutation and hydrophilic forecast of PPV VP2

The VP2 amino acid sequence of PPV strain PPV2010 (Cui et al., 2012) was compared to 38 PPV isolates using ClustalW. Amino acid identity and the hydropathic nature of its residues was assessed with DNAStar and online software. In the Hphob./Kyte&Doolittle analysis (Kyte and Doolittle, 1982) (http://web.expasy.org/protscale/) a window of nine residues was chosen and a weight variation model of 100% (linear). The location of mutations in the PPV structure was analyzed by Chimera 1.6.1 software and online in the http://viperdb.scripps.edu database.

3. Results

3.1. Divergence of PPV

A maximum clade credibility (MCC) tree of the complete ORFs of PPV was constructed (Fig. 1). According to our results, the PPV strains had a common ancestor 252 years ago, with 95% highest posterior density (HPD) values ranging 91–482 years ago. Three major distinct lineages (termed Group I, Group II and Group III, respectively) were identified from the MCC tree: Group I originated 153 years ago (95% HPD, 57 to 293 years ago), and contained 3 European strains, 4 American strains, 1 European strain, 9 Asian strains; Group II originated 157 years ago (95% HPD, 56–291 years ago), and contained 1 Chinese strain, and 13 European strains. Group III originated 114 years ago (95%HPD, 37– 214 years ago). Interestingly, no significant geographic divergence was associated with the clades.

The tandem repeats downstream from the VP gene in some PPV strains were not included in our phylogeny analysis. Nevertheless, this repeat has significant impacts on the replication efficiency of PPVs (Fernandes et al., 2011) and it would be of biological interest to determine whether PPVs with repeats are present in both Group I; Group II and Group III. PPV repeats are about twice the size of those observed in MVM, CPV, and H1 (Fernandes et al., 2011) with one repeat unit at 127 nts and a tandem repeat at 254 nts (1–127, 1–127). Interestingly, PPVs with sequences of one full repeat unit and at least a large repeat clustered all in 2 subclades of Group II. The SR-1, N, China and NADL2 strains with complete repeats (1–127, 1–127) were in one subclade and PPV2010 (1–127, 1–127) and BQ (1–66, 3–127) were in the different subclade from the same clade, the ZJ (1–127, 13–94), Nanjing200801 (1–47, 1–127) and LZ (1–84, 63–127) strains with large, but incomplete repeats in another subclade. In group III, in the same subclade, are two PPVs, 21620005_1 h with a very small repeat (1–127, 105–127) and IDT, with two incomplete repeats (1–83, 1–95).

Table 1
Rates of nucleotide substitutions of PPV.

Dataset	Mean rate	95% HPD lower	95% HPD upper
NS1	3.03E−5	4.61E−8	7.12E−5
VP2	1.07E−4	4.95E−5	1.71E−4
Complete CDS	4.19E−5	1.03E−7	8.05E−5

3.2. Evolutionary rates

Based on the MCMC method, the mean nucleotide substitution rates for the NS1, VP2, and complete coding sequences were determined (Table 1). For the NS1, the estimated rate was 3.03×10^{-5} substitutions/site/year (95% HPD, 4.61×10^{-8} to 7.12×10^{-5}), for the VP2, the estimated rate was 1.07×10^{-4} substitutions/site/year (95% HPD, 4.95×10^{-5} to 1.71×10^{-4}) and for the complete coding sequences, the estimated rate was 4.19×10^{-5} substitutions/site/year (95% HPD, 1.03×10^{-7} to 8.05×10^{-5}).

3.3. Recombination analysis

To assess the role of homologous recombination in evolution of PPVs, we gathered 31 (nearly) full-length genome sequences of PPV and performed recombination analyses using the GARD and RDP programs. The two programs detected some possible recombinants (data not shown). However, when phylogenetic incongruence was tested to confirm whether those results were true or false positive, none of the potential recombinations was supported. No significant recombination events were detected in the complete PPV ORFs.

3.4. Selection pressures

To better understand viral evolutionary dynamics, we assessed selection pressures on NS1 and VP2 (Table 2). The mean ratio of nonsynonymous substitutions (d*N*) and synonymous substitutions (d*S*) rates (d*N*/d*S*) was used to identify signatures of negative (d*N*/d*S* < 1), neutral (d*N*/d*S* = 1) or positive (d*N*/d*S* > 1) selection. Rates for NS1 and VP2 were 0.19 and 0.29, respectively. By REL analysis, 10 and 15 positively selected sites were detected in NS1 and VP2 genes, respectively, but those sites were not supported

Table 2
Selection pressures of PPV.

Dataset	Mean rate of d*N*/d*S*	Negative selected sites
NS1	0.19	260, 263, 598
VP2	0.29	58, 118, 119, 149, 467