Infection, Genetics and Evolution 16 (2013) 369–376

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Molecular evolutionary genetic analysis of emerging parvoviruses identified in pigs

Chao-Ting Xiao, Patrick G. Halbur, Tanja Opriessnig *

Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA

ARTICLE INFO

Article history:
Received 1 February 2013
Received in revised form 7 March 2013
Accepted 12 March 2013
Available online 22 March 2013

Keywords:
Parvovirus
Pigs
Phylogenetic analysis
Porcine bocavirus
Nomenclature

ABSTRACT

Parvoviruses infect a wide variety of vertebrates and arthropods and are associated with various clinical manifestations. Due to the advent of new sequence-independent PCR methods and high-throughput sequencing, several novel members of parvoviruses within the subfamily *Parvovirinae* were recently described. Several of these viruses do not fit in the current classification and others now have confusing or contradictory nomenclature because two or more names were used for similar or identical groups of parvoviruses or identical names were used for distinct virus groups. In this study, recently described vertebrate parvoviruses with emphasis on those identified in pigs were classified through phylogenetic analyses based on the sequences of their complete or near complete genomes, open reading frame (ORF) 1 (non-structural protein, NS1), ORF2 (capsid protein, VP1), and ORF3 (nuclear phosphoprotein, NP1) genes by using Bayesian Markov chain Monte Carlo (MCMC), Maximum Likelihood (ML) and Neighbor-Joining (NJ) methods. Among all available vertebrate *parvovirus* sequences, eight distinct clades were identified, corresponding to the five well established genera *Parvovirus*, *Erythrovirus*, *Denpendovirus*, *Amdovirus* and *Bocavirus*. Moreover, three novel clades were identified and tentatively designated as *PARV4-like virus*, *novel clade 1* and *novel clade 2*. Parvoviruses in pigs were found to be distributed across four different clades including *Parvovirus*, *Bocavirus*, *PARV4-like virus* and the *novel clade 2.* All pig parvoviruses identified to date were organized based on the current analysis. The present analysis will assist to clarify the nomenclature of parvoviruses in pigs and facilitate future uniform assignment of names for new parvoviruses within the subfamily *Parvovirinae.*

1. Introduction

Parvoviruses infect a wide range of vertebrates and invertebrates and are associated with many disease manifestations including reproductive failure, enteritis, panleukopenia, hepatitis, erythrocyte aplasia, immune complex-mediated vasculitis, and cerebellar ataxia (Anderson, 1990; Goddard and Leisewitz, 2010; Manteufel and Truyen, 2008; Mengeling et al., 2000; Tijssen et al., 2011). Parvoviruses are small non-enveloped isometric viruses, with a linear, non-segmented, small, single-stranded DNA genome of about 4–6.3 kb in size (Tijssen et al., 2011). The family *Parvoviridae* is classified into two subfamilies: *Parvovirinae* and *Densovirinae*. Viruses infecting vertebrates and vertebrate cell cultures are assigned to the subfamily *Parvovirinae* and viruses infecting arthropods are assigned to the subfamily *Densovirinae*. The *Parvovirinae* subfamily is currently further divided into five genera: *Dependovirus* (from the Latin word *dependo* which means "to hang down"), *Bocavirus* (from bovine and canine), *Erythrovirus* (from the Greek word *erythros* which means "red"), *Parvovirus* (from the Latin word *parvus* which means "small") and *Amdovirus* (from Aleutian mink disease) (Tijssen et al., 2011).

Recently, due to wide usage of sequence-independent PCR methods and high-throughput sequencing, many new members of the subfamily *Parvovirinae* were discovered both in humans and animals. Since not all of these newly identified viruses fit in the current classification scheme, they have remained unclassified (Allander et al., 2001; Blomström et al., 2009; Canuti et al., 2011; Cheng et al., 2010; Hijikata et al., 2001; Jones et al., 2005; Lau et al., 2008; McKillen et al., 2011; Shan et al., 2011a,b). In order to simplify recognition, novel virus names have been introduced to describe emerging lineages of viruses. Unfortunately, independent research groups have not consistently used the same names to classify similar if not identical groups of parvoviruses. To further complicate this, the same names have apparently been used for different groups of viruses within the *Parvovirinae* subfamily.

To add clarity and to obtain an overview of the phylogenetic relationship of parvoviruses identified in pigs, in this study we performed phylogenetic analyses using GenBank available complete or nearly complete genome sequences of parvoviruses detected in pigs and other vertebrates. Then further phylogenetic analysis was performed with complete nucleotide and amino acid

* Corresponding author. Address: 2654 Vet Med Building, College of Veterinary Medicine, Iowa State University, 1600 S. 16th Street, Ames, IA 50011, USA. Tel.: +1 515 294 1137; fax: +1 515 294 3564.
E-mail address: tanjaopr@iastate.edu (T. Opriessnig).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.03.017

sequences of open reading frame (ORF) 1 (non-structural protein 1, NS1), ORF2 (capsid protein, VP1), and ORF3 (nuclear phosphoprotein, NP1). The results of these phylogenetic analyses aided in classification of presently identified parvoviruses in pigs and will better inform changes to the nomenclature going forward.

2. Materials and methods

2.1. Sequence data

Four-hundred-nineteen complete or nearly complete genome sequences of parvoviruses detected in pigs and other vertebrates in the *Parvovirinae* subfamily available in the GenBank were analyzed. Moreover, to include individual parvoviruses recently identified in pigs with only the NS1 or the VP1 sequence available, and to further define the relationship of pig parvoviruses with each other and to members from other species, additional phylogenetic analysis was performed using 118 complete nucleotide and amino acid sequences of NS1, including 88 sequences from pigs and 30 sequences from other vertebrates, and 148 complete nucleotide and amino acid sequences of VP1 including 119 sequences from pigs and 29 sequences from other vertebrates. Furthermore, since three ORFs are present in the genomes of member of the genus *Bocavirus* and porcine *parvovirus* 4 (PPV4), phylogenetic analyses using the complete nucleotide and amino acid sequences of the ORF3 (nuclear phosphoprotein, NP1) were performed to determine whether the topologies of the trees constructed using the ORF3 are similar to those using other ORFs and genomes. This analysis was done using 15 porcine bocavirus (PBoV) sequences, 8 bocavirus sequences from other vertebrates, and 12 PPV4 sequences.

2.2. Phylogenetic analysis

The sequences were aligned with ClustalW by using the software BioEdit (Hall, 1999). For comparison, phylogenetic analyses were carried out by BEAST 1.7 using the method of Bayesian Markov chain Monte Carlo (MCMC) (Drummond et al., 2012) and by MEGA5 using the Neighbor-joining (NJ) and Maximum-Likelihood (ML) methods (Tamura et al., 2011). Specifically, for the analysis of nucleotide sequences, the General Time Reversible substitution model (GTR) with a proportion of invariant sites and gamma distributed rate heterogeneity (GTR+I+Γ) was used by the methods of ML and Bayesian inference (Drummond et al., 2012; Tamura et al., 2011). Furthermore, the Maximum Composite Likelihood (MCL) substitution model with gamma distribution for rate variation among sites was used by the NJ method. For the amino acid sequence analysis, the Poisson correction model was used with the ML method and the Blosum62 substitution model was used with the Bayesian method with a proportion of invariant sites and gamma distributed rate heterogeneity (Drummond et al., 2012; Tamura et al., 2011). During the analysis of the Bayesian inference, a chain length of 1×10^7 generations with sampling every 1000 generation was performed to estimate the posterior probability, using an extended Bayesian skyline plot as a tree prior, and the burn-in was set at 10% of the sampled states. The trees were assessed by the program Tracer (http://beast.bio.ed.ac.uk/Tracer) and were viewed by FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

2.3. Determination of nucleotide distances between and within clades to indicate genus demarcation

To quantify the nucleotide distances between and within the clades or genera identified in the phylogenetic analysis, the average pairwise distances (between and within the clades) based on the genome sequences were calculated by MEGA5 using the Kimura 2-parameter model (Tamura et al., 2011). The following criteria were used for delimiting clades corresponding to different genera: (I) Maintaining the previous demarcation of the five genera (Tijssen et al., 2011). (II) A bootstrap value (based on 1000 bootstrap replicates for the NJ method and 500 bootstrap replicates for the ML method) >60% and Bayesian posterior probabilities >60% as clade-defining mode. (III) For each delimited genus, the average distances between the genera must be >twice the average distance within the genus. (IV) Simplification of the classification by designating less genera and including more taxa.

3. Results and discussion

3.1. Phylogenetic analysis based on the genomes of vertebrate parvovirus

The phylogenetic analysis demonstrated that the topology of each clade was almost identical and the major clades were consistently identified in the evolutionary trees constructed by the Bayesian inference, ML and NJ methods. All parvoviruses analyzed in this study were effectively grouped in eight clades (Table 1), supported by high bootstrap or posterior probability (>70%), with only two exceptions: Bovine *parvovirus* (BPV) type3 (BPV3) (AF406967) and *Artibeus jamaicensis parvovirus* 1 (JQ037754). These two viruses were not found to belong to any of the identified clades. The evolutionary trees constructed by Bayesian inference and the ML method are shown in Fig. 1 and Fig. S1, respectively, while the tree constructed by NJ is not shown to avoid redundancy.

Based on the topology of the evolutionary trees and the distances between and within the eight identified clades, they were then assigned to eight corresponding genera including the five known genera, *Parvovirus*, *Erythrovirus*, *Dependovirus*, *Amdovirus*,

Table 1
Average evolutionary divergence over sequence pairs within genera and between genera. The number of base substitutions per site from averaging over all sequence pairs within each genus and between genera is shown. Analyses were conducted using the Kimura 2-parameter model. The analysis involved 417 complete and near complete genome nucleotide sequences of viruses within the subfamily *Parvovirinae*. All positions containing gaps and missing data were eliminated. There were a total of 1652 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.

Genus	Average pairwise distances within genera	Average pairwise distances between genera						
		Parvovirus	*Erythrovirus*	*Dependovirus*	*Amdovirus*	*Bocavirus*	*PARV4-like*	*Novel clade 2*
Parvovirus	0.298							
Erythrovirus	0.257	1.474						
Dependovirus	0.391	1.57	1.345					
Amdovirus	0.102	0.947	1.989	2.144				
Bocavirus	0.269	1.339	1.053	1.162	1.758			
PARV4-like	0.425	1.645	1.25	1.204	2.109	1.227		
Novel clade 2	0.286	1.574	1.186	1.225	1.926	1.218	1.036	
Novel clade 1	0.130	1.587	1.237	1.194	1.957	1.01	1.391	1.326