Molecular Phylogenetics and Evolution 100 (2016) 269–280

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Phylodynamic analysis of porcine circovirus type 2 reveals global waves of emerging genotypes and the circulation of recombinant forms

Giovanni Franzo [a,*], Marti Cortey [b,*], Joaquim Segalés [b], Joseph Hughes [c], Michele Drigo [a]

[a] Department of Animal Medicine, Production and Health (MAPS), University of Padua, Viale dell'Università 16, 35020 Legnaro (PD), Italy
[b] Centre de Recerca en Sanitat Animal (CReSA), UAB-IRTA, Barcelona, Spain
[c] MRC-University of Glasgow Centre for Virus Research, Glasgow, Scotland, United Kingdom

ARTICLE INFO

Article history:
Received 9 January 2016
Revised 20 April 2016
Accepted 21 April 2016
Available online 23 April 2016

Keywords:
PCV2
Phylodynamics
Evolution
Population dynamics
Phylogeography
Recombination

ABSTRACT

Since the first description of *Porcine circovirus type 2* (PCV2), four genotypes (PCV2a, PCV2b, PCV2c and PCV2d) have been recognized and three of them have been shown to exhibit worldwide distribution. Here, the population dynamics of PCV2 has been reconstructed over time and the factors that have shaped its evolution determined. The results obtained confirm that PCV2 originated approximately at the beginning of the 20th century. The most recent common ancestor of genotypes PCV2a, PCV2b, PCV2c and PCV2d circulated in the 1950s, 1980s, 1960s and 1950s, respectively, and the population sizes of the individual genotypes remained low until the mid 90s, coinciding with the identification of PCV2 as a major pathogen of the pig industry.

The population dynamics of PCV2 have been characterized by the appearance of periodic waves of distinct genotypes that, after an initial rise, spread following major swine commercial routes and were then superseded by subsequent emerging genotypes. Various recombinant forms displayed comparable population dynamics and spreading routes to those of major genotypes, suggesting that recombinant strains are able to compete with parental ones. The capsid gene is subjected to immune selection and evasion of the host immune response seems to be a major force for the emergence and spread of new genotypes. In contrast, the evolution of other genes appears to be constrained by the particular genomic organization of PCV2. In summary, obtained results suggest that changes in farming strategies, international trade, host population immunity, recombination and the constraints imposed by genome organization have all played a major role in the evolutionary dynamics of PCV2.

1. Introduction

Porcine Circovirus type 2 has emerged as one of the most devastating viral diseases of pigs, causing severe economic losses due to clinical and subclinical syndromes and costs associated to infection control (Alarcon et al., 2013). PCV2 is a small non-enveloped circular single-stranded DNA virus (genome size about 1.7 kb) that belongs to the family *Circoviridae*, genus *Circovirus*. This genus encompasses a group of other animal viruses such as *goose circovirus, canary circovirus, psittacine beak and feather disease virus, chicken anemia virus, and pigeon or columbid circovirus*. Although eleven PCV2-specific RNA transcripts have been predicted, only four of them (*ORF1, ORF2, ORF3* and *ORF4*) have been shown to be translated into proteins with relevant biological roles (Lv et al., 2014). *ORF1* encodes – through alternative splicing – two proteins (i.e. *Rep and Rep′*), which are required for genome replication. *ORF2* encodes the *Cap* protein, the only component of the viral capsid and the major target of the host immune response (Khayat et al., 2011; Nawagitgul et al., 2000). *ORF3* and *ORF4* are located in the same region as *ORF1* but in the antisense strand, and encode for two non-structural proteins involved in the regulation of viral replication and apoptosis (Lv et al., 2014; Gao et al., 2014; Karuppannan and Kwang, 2011; Mankertz et al., 2004). Two intergenic regions (IR) oriented in opposite directions are located between *ORF1* and *ORF2*: the shorter is located between the 3′ end of *ORF1* and *ORF2*, while the longer is located between their 5′ ends and contains the origin of genome replication (Lv et al., 2014).

PCV2 has been traditionally divided into four major genotypes: PCV2a, PCV2b, PCV2c and PCV2d. All but PCV2c have been reported worldwide (Xiao et al., 2015; Franzo et al., 2015c), albeit with different prevalences. It is believed to be transmitted mainly through

* Corresponding authors.
E-mail addresses: giovanni.franzo@unipd.it (G. Franzo), marti.cortey@irta.cat (M. Cortey).

http://dx.doi.org/10.1016/j.ympev.2016.04.028
1055-7903/

the oronasal route, even though any secretions and excretions can be involved in the viral transmission (Grau-Roma et al., 2011). Although PCV2 has been reported to infect some wild population (i.e. wild boars, feral pigs and Peccaries) (Franzo et al., 2015a), it affects mainly domestic pigs and has been a major challenge for the swine industry in the last 20 years. PCV2 causes a group of diseases collectively named porcine circovirus associated diseases (PCVD). One of them, PCV2-systemic disease (PCV2-SD) showed epidemic proportions in Europe and South East Asia by the late 1990s and in North-America by 2004–05 (Segales et al., 2013). The increase in the frequency of outbreaks substantially mirrored an increase in PCV2b prevalence, which replaced PCV2a as the most widespread genotype (Dupont et al., 2008; Carman et al., 2008; Cheung et al., 2007; Cortey et al., 2011; Grau-Roma et al., 2008; Wang et al., 2009). More recently, the PCV2d genotype was identified in China and its wide circulation in that country is currently recognized (Ge et al., 2012). Nevertheless, other studies have confirmed its presence, before its first recognition as an independent genotype, in Switzerland (1998) and its circulation in several countries has been reported (Xiao et al., 2015; Franzo et al., 2015c). Some reports suggest that PCV2d may evade vaccine-induced immunity (Seo et al., 2014; Xiao et al., 2012), which had been traditionally considered fully protective against previously circulating genotypes (Beach and Meng, 2012; Chae, 2012; Fort et al., 2009). However, a recent report indicated that PCV2a based vaccines are also able to protect against PCV2d infection (Opriessnig et al., 2014).

In the past years, other studies have provided remarkable knowledges of PCV2 epidemiology and evolution, investigating its origin, dynamics and routes of spread (Xiao et al., 2015; Firth et al., 2009; Vidigal et al., 2012). Nevertheless, a comprehensive study on the epidemiology and evolutionary dynamics of all known PCV2 genotypes is still lacking. Here, we aimed to elucidate the molecular epidemiology of PCV2 at the global scale and determine the factors that have shaped its evolution at the genomic level.

2. Material and methods

2.1. Dataset

A total of 925 PCV2 complete genome sequences with known collection dates and country of origin were downloaded from GenBank (accessed 06/10/2014) and aligned using the MAFFT method (Katoh and Standley, 2013). All poorly aligned sequences and those displaying degenerate nucleotides or indels which caused reading frame alterations, suggesting sequencing errors, were removed from the dataset.

2.2. Recombination analysis

The whole dataset was tested for recombination using two programs based on different approaches: RDP4 (Martin et al., 2010) and GARD (Kosakovsky Pond et al., 2006). To decrease the computational burden, the dataset was reduced using CD-HIT (Li and Godzik, 2006) to cluster together sequences with a nucleotide identity threshold of 99% and a single sequence representative of each cluster was selected. GARD analysis was performed using the program implemented in datamonkey (http://www.datamonkey.org/) and, to cope with the circular structure of the PCV2 genome, the complete genome alignment was concatenated twice. The RDP4 settings for each method were adjusted to account for the dataset features according to the RDP manual recommendations. In particular RDP, GENECONV, Chimaera and 3Seq were used in a primary scan while the full set of available methods was used for the analysis refinement. Only recombination events detected by more than 2 methods with a significance value lower than 10^{-5} (p-value < 10^{-5}) and Bonferroni correction were accepted. The non-recombinant sequences as well as those sharing recombination events were split into separate datasets and expanded to their original size.

2.3. Genotyping and database preparation

The non-recombinant sequences were classified into genotypes PCV2a, PCV2b, PCV2c or PCV2d according to Franzo et al. (2015b).

The most appropriate nucleotide substitution model was selected according to the results of the Akaike information criterion (AIC) score calculated using JModel Test 2.1.2 (Darriba et al., 2012). A phylogenetic tree was reconstructed using the Maximum likelihood (ML) approach implemented in PhyML (Guindon et al., 2010). The best tree search method included the combination of two branch swapping algorithms: nearest neighbor interchange (NNI) and subtree pruning and regrafting (SPR). The robustness of the monophyly of the taxa subsets was estimated with the fast non-parametric version of the aLRT (Shimodaira–Hasegawa [SH]-aLRT), developed and implemented in PhyML 3.0 (Anisimova et al., 2011). On the basis of the recombination and phylogenetic analyses, sequences were divided into independent datasets, corresponding to different genotypes and CRFs (i.e. those including more than 30 sequences collected in two or more countries). Every dataset was further divided in three regions, namely *ORF1*, *ORF2* and intergenic region (obtained merging together the major and the minor intergenic regions) and a new alignment was generated on each dataset. The coding regions were aligned at the amino acid level and then the nucleotide sequences were back-translated using the MAFFT algorithm implemented in TranslatorX (Abascal et al., 2010).

2.4. BEAST analysis

The time to most recent common ancestor (tMRCA), substitution rates and population dynamics were jointly estimated using a Bayesian serial coalescent approach implemented in BEAST 1.8.1 (Drummond and Rambaut, 2007). For this purpose, datasets for each genotype and each CRF were analyzed independently. Each region (i.e. *ORF1*, *ORF2* and intergenic regions) were allowed a different substitution model and molecular clock. Substitution and clock models were respectively selected according to the results of the Bayesian Information Criterion (BIC) score calculated using JModel Test 2.1.2 and to the Bayes Factor (BF) value, calculated through estimation of marginal likelihood of the different models using the path sampling (PS) and stepping stones (SS) methods (Baele et al., 2012). To reconstruct population dynamics over time, the non-parametric skyline model was selected (Drummond et al., 2005). The timing of viral dispersal patterns among countries were estimated with the same program using the discrete state phylogeographic approach described by Lemey et al. (2009a). This model allows character mapping in natural time scale, under a molecular clock assumption and accounting for population size changes while integrating them over phylogenetic uncertainty. 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. For this purpose, to avoid over-parametrization of the model and achieve as much information as possible two models were used. The first one, based on the complete datasets implemented a symmetric substitution model with BSSVS considering each country as a discrete state. For the second analysis, countries were grouped in macro-areas (i.e. Asia_other than China, China, Europe, Oceania, North_America, South_Africa and South_America) and the