INTRODUCTION

Recombination and point mutations are evolutionary mechanisms that are crucial for the generation of genomic diversity and adaptation. Nevertheless, the relative contribution of recombination to the evolution of different organisms has been explored in detail only recently (Awadalla, 2003; Hughes et al., 2007; Norberg et al., 2007; Papke et al., 2007; Worobey & Holmes, 2001). It is known that human immunodeficiency virus type 1 (HIV-1), hepatitis C virus (HCV) and hepatitis B virus have high mutation rates (Crotty et al., 2000; Leitner & Albert, 1999; Margeridon-Thermet & Shafer, 2010).

Recent studies have demonstrated that parvoviruses also have high substitution rates, although the contribution of these rates to parvovirus fitness remains to be clarified (López-Bueno et al., 2006; Norja et al., 2008; Pereira et al., 2007; Shackelton & Holmes, 2006; Streck et al., 2011). In general, point mutations are a well-established evolutionary force that shapes viral diversity and adaptation (Grigoras et al., 2010; Jayaraman et al., 2011; Lauring & Andino, 2010; Sanjuán et al., 2010). Recombination, on the other hand, has recently been suggested as a potential factor that may contribute to host adaptation and acquisition of drug resistance (Kellam & Larder, 1995; Mild et al., 2007; Moutouh et al., 1996; Nora et al., 2007; Rambaut et al., 2004; Suryavanshi & Dixit, 2007). For example, in geographical regions where HIV-1 is highly prevalent, inter-subtype recombination can be pervasive (Leal & Villanova, 2010; Martins et al., 2008; Rousseau et al., 2007). Recombination is also frequently reported in other RNA viruses, such as HCV (Noppornpanth et al., 2006), in DNA viruses (Kaneko et al., 2011; Schierup et al., 2005; Stainton et al., 2012; Zell et al., 2012) and even endogenous retroviruses (Bartosch et al., 2004).

Although parvoviruses have recently been associated with many diseases in animals (López-Bueno et al., 2006; Mengeling, 2006), the overall contribution of recombination to parvovirus diversity and disease onset needs to be better studied (Pereira et al., 2007; Shackelton et al., 2007; Shangjin et al., 2009; Streck et al., 2011). Recombination likely increases the genetic diversity of parvovirus populations, although evidence of in vivo recombination is quite rare and difficult to obtain. For example, in a recent study, phylogenetic incongruence between gene regions was used to show recombination in rodent parvovirus (Lukashov & Goudsmit, 2001). In addition, evidence of natural recombination has been observed in canine parvoviruses (Mochizuki et al., 2008). Another study applied unique methods to detect recombination in animal paroviruses, although
details about the rates and the pattern of mosaic genomes were not presented (Rambaut et al., 2004; Shackelton et al., 2007). Therefore, it is crucial to evaluate the effect of recombination on the evolution of pathogens because immune evasion and population level diversity may increase due to genomic interchange among organisms such as viruses (Rambaut et al., 2004). For these reasons, we performed a detailed study to expand the current knowledge on recombination in the Ns1 and Vp1Vp2 genes of porcine parvovirus (PPV).

RESULTS

Genetic characteristics of the PPV isolates

Initially, the full-length genomes of some of the PPV reference strains (i.e. Kresse_U44978, BQ_EU790641, ZJ_EU790642, NC_0017118, POVG_D00623, SR-1_DQ675456, China_AY583318 and Nanjing200801_FJ822038) were analysed to determine the nucleotide diversity among them, and the mean genetic distance was 0.007 ± 0.001. Next, based on the clustering pattern observed by phylogenetic analysis (see below), we estimated the genetic diversity of the POVNADL2 and Tornau/IDT groups. The mean genetic distances of the Ns1 gene were 0.003 ± 0.001 and 0.007 ± 0.001, respectively, for the POVNADL2 group and Tornau/IDT group. The mean genetic distances of the Vp1/Vp2 gene were 0.004 ± 0.001 and 0.007 ± 0.001 for the POVNADL2 group and Tornau/IDT group, respectively. Interestingly, high pairwise diversity was detected in both the Ns1 and Vp1/Vp2 genes of the Tornau/IDT group, which included PPV isolated in this study. To better characterize the PPV sequences, we analysed amino acid substitutions on a site-by-site basis. The amino acid composition of our isolates presented a low degree of variation. Specifically, we found 47 variable sites among the 453 aa of the Ns1 gene, there were 38 variable sites among the 612 residues of the Vp2 gene. In the Vp1 gene, there were 48 variable sites out of 664, and in the Vp2 gene, there were 38 variable sites among the 453 aa of the coding region. Although amino acid polymorphisms were observed throughout the sequences, insertions and deletions were rare. We found one stop codon at site 110 of the Vp2 protein in isolate 901-3. Perhaps the most important feature of the alignments was an extensive deletion identified between aa 29 and 109 in the Vp2 gene of the isolate 371-3. Sequences generated in this study were deposited in GenBank with the following accession numbers: JX568153–JX568158. Complete alignments of the Ns1 and Vp1/Vp2 genes from the PPV isolates in this study are available at http://www.biotorrents.net/

Phylogenetic analysis

Our first approach was to construct a tree using the entire 3847 bp genomic fragment (Ns1/Vp1Vp2) amplified by PCR. Fig. 1 shows a maximum-likelihood (ML) tree constructed using our sequences (filled dots) and the following PPV reference sequences: POVG_D00623, China_AY583318, POVNADL2_M38367, SR-1_DQ675456, ChinaNanning_HM989009, Kresse_U44978, Challenge_AY684866, China_10_JN872448, BQ_EU790641, VR1-1_AY390557, ZJ_EU790642, Nanjing200801_FJ822038, IDT_AY684872 and Tornau_AY684869. The clustering pattern of the PPV isolates indicated that there are two main clades (groups): one formed by the reference sequences IDT (AY684872) and Tornau (AY684869) and another formed by POVNADL2 (M38367). Sequences that we generated (i.e. Den_11, 626-1, 2733-5, 371-2, 371-3 and 2074-7) clustered within the group formed by IDT and Tornau, with the exceptions of isolate 2074-7 and Den_11 (indicated by grey arrows), which were located at the base of the IDT/Tornau group and at the base of the POVNADL2 group, respectively. In general, the tree indicated a lack of resolution due to low bootstrap values, mainly within the group formed by POVNADL2 (M38367) and ZJ (EU790642). We also noticed the presence of reference strains with long branches (grey diamonds). Interestingly, trees constructed without these sequences usually had higher statistical support, but support within the POVNADL2 clade was continuously low. This weak phylogenetic signal coincides with the reduced genetic diversity of sequences within the POVNADL2 group. On the other hand, the long branch of the Nanjing strain was caused by the high frequency of A/T—G mutations, whereas in the POVG strain, the long branch was due to excessive single site deletions found between nt 1712 and 1973 of the Ns1/Vp1Vp2 region.

Network pattern of the PPV isolates

We used a network method to explore the ancestral relationships of the PPV isolates. Evolutionary history is usually represented by phylogenetic trees, in which two strains share a common ancestor. However, there are some biological events, such as genome recombination, that are better represented as a reticulated network. Thus, networks can be used to identify strains in a dataset that have conflicting ancestral relationships and display probable recombination. Fig. 2 shows two networks of the Ns1/Vp1Vp2 genomic region that was inferred with the PPV isolates. In both networks (panel a and panel b), there are two distinct clusters (delineated by dashed circles) of sequences: one composed of reference sequences IDT (AY684872) and Tornau (AY684869) and another composed of POVNADL2 (M38367) and ZJ (EU790642). Panel (a) of Fig. 2 shows a network in which isolate 2074-7 (arrow in Fig. 2) forms reticulated branches with other PPV isolates. On the other hand, the network shown in panel (b) has a more resolved (tree-like) branching pattern, and it is clear that the reticulated network pattern observed in panel (a) is due to the presence of isolate 2074-7. We also analysed the Vp1Vp2 gene region separately because the Ns1 region was absent in some reference strains (i.e. 27a_AY684871, 106b_AY684870, 225b_AY684864, 21a_AY684866, 143a_AY684867 and 15a_AY684865). The Vp1Vp2 network indicated that isolate 225b (AY684864) has an equal number of recombination events as other PPV isolates (Fig. S1, available in JGV Online).
Pervasive levels of recombination in the PPV genome

Although the network method used above suggested that isolate 2074-7 and the reference strain 225b_AY684864 were likely recombinant strains, information regarding the pattern and the rate of recombination were not available. In addition, phylogenetic analysis indicated some recurrent discrepancies, such as long branches and clades with extremely low statistical support, even after excluding strains 2074-7 and 225b_AY684864 from the trees (data not shown). Based on that, we decided to use more powerful methods (see a brief description in Methods) to characterize recombination in the PPV genome. Distinct initial parameters, such as gap penalties, window sizes and $P$-value thresholds, were used and then optimized to make the analyses more reliable and to avoid false positives. The results of these analyses indicated the presence of breakpoints (recombination events) in some of the PPV strains. Although we have used stringent $P$-values and optimized the parameters for each method, reliable signals of recombination (i.e. detected by more than one method) were observed only in two strains: 2074-7 and 225b (AY684864). PPV genomes are highly conserved, and six sequences were considered as a source of bias: one isolate (371-3) because it presented a long deletion in the Vp2 region and five other sequences (Nanjing_FJ822038, POVG_D00623, ChinaNanning_HM989009 and VRI-1_AY390557) because they had extremely long branches in the phylogenetic trees. We repeated the recombination analyses without the above sequences. In addition, ‘manual’ analyses were performed, in
Fig. 2. Phylogenetic networks of the PPV isolates. (a) Network constructed using the Ns1/Vp1Vp2 genomic fragment of PPV. It shows isolates clustered in two distinct clades (dashed circles), with the exception of 2074-7 (grey arrow), which forms a network between the main clades. (b) Network constructed with the same set of sequences but without isolate 2074-7. This network similarly has two clades, but the shape has a resolved structure that resembles the topology of a tree (cladogram). Networks were constructed using the split decomposition method and the Kimura two-parameters model, implemented with the SplitsTree4 software (http://splitstree.org/).
which parameters were optimized and sequences were added and removed several times. All analyses confirmed the recombination signals in isolates 2074-7 and 225b. On the other hand, evidence of recombination was detected in the interval between nt 1694 and 2087 in the genome of isolate POVG (D00623); in the interval between nt 1574 and 3351 in isolate SR-1 (DQ675456) and in the 1779 site of the Ns1/Vp1Vp2 genome region in isolate ZJ (EU790642). Although isolates POVG (D00623), SR-1 (DQ675456) and ZJ (EU790642) had one or two breakpoints, recombination was detected by only one method. Because the criterion used here was to accept as recombinant only the isolates detected by at least two methods. Thus, the above isolates were regarded as putative recombinant viruses. Interestingly, the presence of the isolate POVG and/or 371-3 (both isolates have gaps in their genomes) caused misidentification of recombinants by the GenconV method. Therefore, alignments with deletions should be analysed carefully by this method.

Fig. 3 summarizes the results of the recombination analyses and shows the location of breakpoints mapped in the genome of strains 2074-7 and 225b (AY684864). The upper panel of Fig. 3 shows the location of breakpoints in the Ns1/Vp1Vp2 genome region of isolate 2074-7. Vertical coloured lines indicate positions where breakpoints were detected by distinct methods, and the dashed vertical line in the grey area delineates the Ns1 and Vp1Vp2 genes. It is interesting to note that the methods failed to converge in terms of the number and location of breakpoints detected. Nevertheless, all breakpoints detected were restricted to the interval between nt 1689 and 2111, with the exception of site 2700, which was detected by the Gard method (green bar in the Fig. 3). Similarly, the methods disagreed on the location and number of breakpoints in the Vp1Vp2 genome region of isolate 225b (AY684864). Likewise, all of the methods correctly identified the two regions of the Vp1Vp2 alignment (nt 374–561 and 1688–1755) that have the most concentrated breakpoints (lower panel of Fig. 3).

**Mosaic structure of the PPV strains**

We used a similarity plot method in an attempt to describe the mosaic structure of recombinant strains of PPV. Fig. 4 (upper panel) shows that the genome of strain 2074-7 is a chimera between the IDT clade and the Kresse clade (U44978). Specifically, the genomic interval between nt 1 and 2000 and the short fragment between nt 3001 and 3200 (green area in the upper panel of Fig. 4) were similar to the IDT strain; the remaining genome (blue areas in the upper panel of Fig. 4) was similar to the Kresse strain. Based on the above results, an ML tree was inferred using the Ns1 gene region, and another tree was inferred using the Vp1Vp2 gene region (Fig. 5). As expected, the Ns1 and Vp1Vp2 genes of isolate 2074-7 have conflicting phylogenetic histories. Whereas the Ns1 tree suggests that isolate 2074-7 falls within the IDT/Tornau clade (Fig. 5a), the Vp1Vp2 tree locates this isolate within the POVNADL2 clade (Fig. 5b). Therefore, these discordant trees suggest that isolate 2074-7 is an inter-clade recombinant strain. Similarly, the 225b (AY684864) isolate has two main transitions in the similarity plots (lower panel in Fig. 4). The regions between nt 1 to 480 and 1891 to 2294 (orange lines in the lower panel of Fig. 4) are similar to the 27a strain (AY684871), while the internal region between nt 481 and 1890 (green lines in the lower panel of Fig. 4) is similar to the IDT isolate. In addition, independent neighbour-joining (distance-based) trees were inferred using the fragments corresponding to recombination breakpoints in Fig. 4. First, one tree was constructed using the region between nt 1 and 480, then another tree was inferred using the region between nt 1891 and 2294 (orange areas in the lower panel of Fig. 4). Lastly, a third tree was constructed using the internal fragment (green area in the lower panel of Fig. 4) comprising nt 481 to 1890. The first and second trees showed strain 225b clustered within the POVNADL2 clade, and the third tree, constructed with nt 481 to 1890, showed isolate 225b within the IDT/Tornau clade (Fig. 6). The isolate 225b was previously reported as a recombinant virus (Shackelton et al., 2007); nevertheless, no information about the mosaic genome was available. This may be due to the difficulty of determining breakpoints in chimeric genomes formed by the recombination of low divergence parental sequences. As a whole, transitions in the similarity plots in both recombinant strains (i.e. 2074-7 and 225b_AY684864) coincided with the breakpoints detected previously. Finally, the phylogenetic trees constructed with non-recombinant fragments corroborate the chimeric pattern found in the genomes of strains 2074-7 and 225b.

**DISCUSSION**

Our phylogenetic analysis indicated some inconsistencies in trees constructed with the Ns1 gene, the Vp1Vp2 gene and those constructed with the larger Ns1/Vp1Vp2 genomic fragment. For this reason, we started analysing the PPV sequences in an attempt to identify signals of recombination. Based on a network method we identified the isolates 2074-7 and 225b (AY684864) as potential recombinant genomes. Given that gene conversion, horizontal gene transfer and hybridization can also produce phylogenetic networks and false signals of recombination. Then we used a variety of more sophisticated methods that formally incorporate recombination into their statistical framework.

In general, these methods showed complete disagreement in determining the number and location of breakpoints. Although exact recombination sites were difficult to pinpoint, we were able to use the intervals in which most breakpoints were detected as a proxy to indicate recombined regions. This lack of convergence can be partially explained by different assumptions about the evolutionary processes incorporated by each method. In addition, parameters such as recombination cost and window size, affect the accuracy and performance of most methods (Ané, 2011; Wiuf et al., 2001). It is known that methods
used to detect recombination perform poorly and fail to identify breakpoints in low diversity genomes and increasing levels of diversity usually give more power to the results (Bay & Bielawski, 2011; Posada, 2002; Schultz et al., 2006). However, when sequences are highly divergent they yield false-positive breakpoints. For instance, it has been observed that the Gard method overestimates the number of breakpoints, while the MaxChi and Chimaera techniques produce false-positive results in the presence of long-branch attraction, both in simulated and empirical
sequences (Bay & Bielawski, 2011). Even more conservative methods, such as the distance-based bootscanning, can produce false-positive results in divergent sequences, regardless of high bootstrapping support of the results (Posada, 2002). There are many features that influence the precision of methods to determine the location of breakpoints. Window size, mutation, rate heterogeneity and the presence of gaps in sequences have been observed to increase false results of most algorithms used to detect recombination (Ané, 2011; Husmeier & Mantzaris, 2008; Schultz et al., 2006; Westesson & Holmes, 2009). In our results, spurious breakpoints were assigned in sequences due to the presence of gaps (see Results).

Great efforts have been undertaken to understand the contribution of recombination to the genetic diversity of viruses. Parvoviruses have extremely high mutation rates (Pereira et al., 2007; Shackelton & Holmes, 2006); thus chimera viruses should be observed more frequently in natural parvovirus infections. Since parvoviruses rely on cellular polymerases to replicate (Cotmore & Tattersall, 1998), it is likely that recombination also occurs mostly by the activity of cellular machinery (Hogan & Faust, 1986). Although the mechanism of parvovirus recombination has not yet been revealed, some scenarios can be considered (Martin et al., 2011). The single-stranded genome of parvovirus usually has hairpin structures that may stall DNA polymerase; if this enzyme swaps to a distinct genome upon the restart, then the final molecule will be a chimera. Alternatively, parvovirus replication generates an intermediate double-stranded concatamer molecule that later will be cleaved by the viral Ns1 protein to form single-stranded viral genomes (Christensen & Tattersall, 2002; Cotmore & Tattersall, 1998). Double-breaks in this intermediate molecule, can activate the DNA repair enzymes that may connect distinct concatamers by the homologous recombination system. Both scenarios will generate mosaic viruses if cells are infected by distinct parvovirus strains.

Although mosaic isolates have been described in circovirus (Cadar et al., 2012), bocavirus (Cheng et al., 2011; Csághola et al., 2012; Kapoor et al., 2010), erythrovirus (Abe et al., 2007; Shackelton & Holmes, 2006) and in distinct parvoviruses (Lukashov & Goudsmit, 2001; Ohshima & Mochizuki, 2009; Shackelton et al., 2007), recombination

![Fig. 4. Mosaic pattern of isolates 2074-7 and 225b. Horizontal lines represent similarity along the genomes of the recombinant isolate 2074-7 and the reference strain 225b (AY684864). The y-axis indicates the percentage of similarity, while the x-axis shows nucleotide positions. The coloured bars above the lines indicate the intervals with higher similarity between the recombinant isolate and one of the parental strains. The upper panel shows the similarity of isolate 2074-7 to the parental strains Kresse (U44978) and IDT (AY684872). The vertical dashed line indicates the limit between the Ns1 and Vp1/Vp2 genes. The lower panel represents the mosaic structure of the Vp1/Vp2 genomic fragment of isolate 225b, and coloured bars indicate the regions of similarity with the parental strains 27a (AY684871) and IDT (AY684872).](image-url)
Fig. 5. ML trees constructed using the Ns1 and Vp1Vp2 fragments of PPV. (a) Shows the Ns1 tree, in which isolate 2074-7 is located within the IDT/Tornau clade. (b) Shows the Vp1Vp2 tree, indicating that isolate 2074-7 is within the POVNADL2 clade. The tree was constructed using the HKY model implemented with PhyML software. Values at the nodes of the tree are bootstrap support, and they were obtained with 500 replicates.
Fig. 6. Phylogenetic trees constructed with fragments of the Vp1Vp2 genomic region of PPV. Trees were constructed to show that strain 225b (indicated by a filled arrow) is clustered within the POVNADL2 clade in the trees constructed with fragments 0–480 and 1892–2294. The third tree, constructed with nt 481–1890, showed isolate 225b within the IDT/Tornau clade (see results for more details). Neighbour-joining trees were estimated assuming the Kimura two-parameters model with 500 bootstrap replicates; analysis was performed using the software MEGA4.
Recombination in porcine parvovirus

is quite rare in the family Paroviridae. In addition, mosaic isolates have been detected under circumstances that may favor recombination. For example, the only B19 mosaic strain so far reported (GenBank accession no. AB126266) was isolated from an individual that was immune-suppressed following liver transplantation (Abe et al., 2007). A newly described porcine circovirus mosaic strain was detected in a domestic pig from a region where wild boars and domestic pigs have close contacts (Cadar et al., 2012). Moreover, it has been reported that parvoviruses of rodents are prone to recombination (Lukashov & Goudsmitt, 2001; Shackelton et al., 2007) and this would be partially explained by the high-density populations of rodents in nature. All the above conditions would facilitate the occurrence of coinfections and therefore increase the changes of appearance chimera viruses.

It is plausible that recombination could afford distinct features to novel PPV strains. Notably, isolate 27a (AY684871) showed low homologous neutralizing antibody titres compared with the PPV vaccine and other field strains, thus suggesting a novel antigenic variant (Zeeuw et al., 2007). We found that the 27a isolate contributed (in addition to the IDT strain) to the formation of the chimeric genome of the recombinant strain 225b. Likewise, the recombinant isolate 2074-7 was obtained from an animal that, though vaccinated, presented signs of disease caused by parvovirus. Based on the above evidences, recombination potentially increases the adaptive fitness of chimeric strains within-hosts. However, if recombination is an adaptive process to parvovirus then increasing frequency of common circulating forms of mosaic virus should be observed in natural host populations. In fact, this is not the case because our analysis identified only two recombinant strains in the PPV sequences isolated from 1964 to 2011. Therefore, evidence favouring recombination as an important mechanism of increasing genetic diversity in parvovirus still remain elusive.

METHODS

Ethics statements. This study was approved by the Ethics Committee of the Harbin Veterinary Research Institute of CAAS.

PPV isolates. All strains were isolated in swine testicle (ST) cells, a well-known permissive continuous cell culture type for PPVs (McClurkin & Norman, 1966; Oraveerakul et al., 1992). ST cells (ATCC catalogue number CRL-1746) were kindly provided by Dr. K. J. Yoon (Iowa State University), and cells were maintained in Eagle’s minimal essential medium with 1% (w/v) sodium pyruvate and 0.25% (w/v) lactalbumin hydrolysate. Briefly, 500 μl of a suspension of pooled liver tissue was inoculated onto ST cells grown in minimal essential medium with Hanks’ salts and l-glutamine as described elsewhere (Zimmermann et al., 2006). All samples were isolated in 2010 or 2011 from vaccinated animals in Denmark that were presenting classical symptoms of disease caused by parvovirus infection (for a detailed description of the disease see: Kresse et al., 1985; López-Bueno et al., 2006; Mengeling, 2006).

The following PPV sequences were used as references. They were obtained from GenBank and the accession numbers are:

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Js1/Vp1V2.</td>
<td>Nanjing200801 (FJ822038), ZJ (EU790642), BQ (EU790641), China_10 (IN822448), Challenge (AY684866), POVNADL2 (M38367), China_318 (AY583318), POVG (D00623), ChinaNanning (HM989009), SR-1 (DQ675456), Kresse (U44978), VRI-1 (AY390557), Tornau (AY684869) and IDT (AY684872).</td>
</tr>
<tr>
<td>Vp1Vp2.</td>
<td>27a (AY684871), 106b (AY684870), 225b (AY684864), 21a (AY684868), 143a (AY684867) and 15a (AY684865).</td>
</tr>
<tr>
<td>Vp1V2.</td>
<td>25 (AY684871), 106b (AY684867), 225b (AY684864), 21a (AY684868), 143a (AY684867) and 15a (AY684865).</td>
</tr>
</tbody>
</table>

PPV genomes. All strains were isolated in swine testicle (ST) cells, a well-known permissive continuous cell culture type for PPVs (McClurkin & Norman, 1966; Oraveerakul et al., 1992). ST cells (ATCC catalogue number CRL-1746) were kindly provided by Dr. K. J. Yoon (Iowa State University), and cells were maintained in Eagle’s minimal essential medium with 1% (w/v) sodium pyruvate and 0.25% (w/v) lactalbumin hydrolysate. Briefly, 500 μl of a suspension of pooled liver tissue was inoculated onto ST cells grown in minimal essential medium with Hanks’ salts and l-glutamine as described elsewhere (Zimmermann et al., 2006). All samples were isolated in 2010 or 2011 from vaccinated animals in Denmark that were presenting classical symptoms of disease caused by parvovirus. Based on the above evidences, recombination potentially increases the adaptive fitness of chimeric strains within-hosts. However, if recombination is an adaptive process to parvovirus then increasing frequency of common circulating forms of mosaic virus should be observed in natural host populations. In fact, this is not the case because our analysis identified only two recombinant strains in the PPV sequences isolated from 1964 to 2011. Therefore, evidence favoring recombination as an important mechanism of increasing genetic diversity in parvovirus still remain elusive.

Extraction and amplification of PPV genomic DNA. PPV strains were subjected to three passages in ST cells. Genomic DNA was extracted from PPV-infected ST cells at 48 h post-inoculation and then resuspended in 400 μl TE and stored at −20 °C. Using the entire sequence of the NADL2 (M38367) strain (GenBank accession no. NC_001718), seven primer pairs were designed to amplify the swine PPV genome; a detailed description of these primers and amplification strategy can be found elsewhere (Shangjin et al., 2009). Primers were synthesized by TaKaRa Biotech Co. The PCR contained 1 μl extracted DNA, the corresponding primer pair, 3 mM dNTPs (TaKaRa Biotech Co.), 0.2 mM MgCl₂, and 0.5 μ L LA Taq polymerase (TaKaRa Biotech Co.). Amplification was initiated by a pre-denaturation phase at 95 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 15 s, annealing at temperatures ranging from 51 to 57 °C for 30 s and extension for 30 to 90 s (depending on the primer pair used) at 72 °C; and a final extension for 7 min at 72 °C.

DNA sequencing. The PCR products were sent to TaKaRa Biotech Co. for sequencing, and the results were aligned to obtain complete PPV genome sequences using the server DNASTAR. Multiple sequence alignment was completed with the whole genome sequence isolates using the DNASTAR MegAlign program.

Sequence alignment and phylogenetic inference. Sequences were aligned using CLUSTAL_X (Larkin et al., 2007). After the editing process, the sequences were manually aligned using the SE-Al program, version 2.0 (Department of Zoology, Oxford University; http://evolve.zoo.ox.ac.uk/software/). To construct ML trees, we used the HKY model (Hasegawa et al., 1985) as implemented in PhyML (Guindon & Gascuel, 2003). The HKY model was selected according to the likelihood ratio test in jModeltest (Posada, 2008). Topali v. 2.5 was also used, mainly to edit and select unique sequences (Milne et al., 2004). The trees were visualized and edited using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). Genetic distances and neighbour-joining trees were estimated assuming the Kimura two-parameters method because it tends to infer more resolved graphs than other methods (Huson & Bryant, 2006). By this method, networks are reliable and the results are visual graphs that can be easily interpreted. Another advantage of the networks method is that there is no need to assume a query sequence (putative recombinant) or even parental sequences (those that contributed to form the chimeric genome of the recombinant) during the analysis. For our analysis, we used the split decomposition method because it tends to infer more resolved graphs than other methods (Huson & Bryant, 2006). Networks were generated using the SplitsTree software (http://splitstree.org/). Next, we used Gard-Genetic algorithms for recombination detection (Kosakovsky Pond et al., 2006). We used default parameters with a GTR+gamma rate variation model, and the analysis was performed using the software Mega4 (Tamura et al., 2007).

Recombination analyses. To determine the extent of recombination in the PPV sequences, we used distinct methods as described below. Initially, we used a phylogenetic network method that aims to describe non-tree-like evolution among related taxa, resulting from the presence of reticulate events such as hybridization, gene duplication, horizontal gene transfer and recombination (Huson & Bryant, 2006). By this method, networks are reliable and the results are visual graphs that can be easily interpreted. Another advantage of the networks method is that there is no need to assume a query sequence (putative recombinant) or even parental sequences (those that contributed to form the chimeric genome of the recombinant) during the analysis. For our analysis, we used the split decomposition method because it tends to infer more resolved graphs than other methods (Huson & Bryant, 2006). Networks were generated using the SplitsTree software (http://splits Tree.org/). Next, we used Gard-Genetic algorithms for recombination detection (Kosakovsky Pond et al., 2006). We used default parameters with a GTR+gamma rate variation model, and the analysis was performed using the Data Monkey web server (http://www.datamonkey.org/).
In addition, we used Recco software that uses a cost/optimization method to distinguish mutation from recombination (Maydt & Lengauer, 2006).

Finally, we used the software RDP v.4 (Martin et al., 2005), which utilizes a collection of methods. Below is a brief description of these methods; an excellent and detailed explanation of each method implemented in the RDP program can be found in the user’s manual (http://darwin.uvigo.es/rdp/rdp.html). 3Seq performs an exact non-parametric test to detect recombination in triplets of sequences (Boni et al., 2007).

Maximum $\chi^2$ (MaxChi) is a method implemented by Maynard-Smith (Smith, 1992), and it uses variable/invariable sites to detect recombination in pairs of sequences. This method generates random sequence pairs; the significance level is evaluated by the proportion of simulated sequence pairs with maximum $\chi^2$ values higher than the real data.

Maximum match $\chi^2$ (Chimaera) is a modification of Smith’s method. It uses variable sites to calculate maximum $\chi^2$ match statistics (Posada & Crandall, 2001).

Geneconv detects gene conversions (recombination) by evaluating conserved substitutions in fragments between two sequences (Sawyer, 1989). Although evolutionary methods are not explicitly implemented in Geneconv, it is robust and has low levels of false-positive detection of recombination, even those due to rate heterogeneity and natural selection (Bay & Bielawski, 2011).

Initially, we used default parameters, and then these were optimized to avoid detection of false-positive recombination. In addition, window sizes of 20–50 were used and Bonferroni correction was utilized.

Although the above methods are robust and extremely popular, they also suffer from false-positive detection of recombination (Bay & Bielawski, 2011). In the case of the PV isolates, genome conservation and purifying selection are the main scenarios that could increase the chance of false positives.

To describe the mosaic pattern of isolates identified as recombinants, we used a similarity plot method, which was implemented using the RAT software (Etherington et al., 2005).

ACKNOWLEDGEMENTS

We would like to thank Professor Joaquim Segales, Jens Nielsen, Lars Erik Larsen, Anette Botner, Mette Sif Hansen and Louise Lohse, who provided DNA, virus samples or technology support. This work was supported by the National Natural Science Foundation of China (nos 31001069, 31172349 and 31172341). E.L. was supported by the Fundação de Amparo e Desenvolvimento da Pesquisa (FADESP) and Pro-Reitoria de Pesquisa e Pos-Graduação (PROPESP) (grant no. 11/2012)

REFERENCES


