High rate of viral evolution in the capsid protein of porcine parvovirus

André Felipe Streck,1 Sandro Luis Bonatto,2 Timo Homeier,1 Carine Kunzler Souza,3 Karla Rathje Gonçalves,3 Danielle Gava,3 Cláudio Wageck Canal3 and Uwe Truyen 1

Correspondence
Uwe Truyen
truyen@vetmed.uni-leipzig.de

1Institute for Animal Hygiene and Veterinary Public Health, Faculty of Veterinary Medicine, University of Leipzig, An den Tierkliniken 1, 04103 Leipzig, Germany
2Laboratório de Biologia Genómica e Molecular, Pontifícia Universidade Católica, Av. Ipiranga 6681, Prédio 12, bloco C, sala 172, 90619-900 Porto Alegre, Brazil
3Laboratório de Virologia, Faculdade de Veterinária, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9090, Prédio 42.602, CEP 91540-000, Porto Alegre, Brazil

In recent years, it has been shown that some paroviruses exhibit high substitution rates, close to those of RNA viruses. In order to monitor and determine new mutations in porcine parvovirus (PPV), recent PPV field isolates from Austria, Brazil, Germany and Switzerland were sequenced and analysed. These samples, together with sequences retrieved from GenBank, were included in three datasets, consisting of the complete NS1 and VP1 genes and a partial VP1 gene. For each dataset, the nucleotide substitution rate and the molecular clock were determined. Analysis of the PPV field isolates revealed that a recently described amino acid substitution, S436T, appeared to be common in the VP2 protein in the Austrian, Brazilian and German virus populations. Furthermore, new amino acid substitutions were identified, located mainly in the viral capsid loops. By inferring the evolutionary dynamics of the PPV sequences, nucleotide substitution rates of approximately $10^{-5}$ substitutions per site per year for the non-structural protein gene and $10^{-4}$ substitutions per site per year for the capsid protein gene (for both viral protein datasets) were found. The latter rate is similar to those commonly found in RNA viruses. An association of the phylogenetic tree with the molecular clock analysis revealed that the mutations on which the divergence for both capsid proteins was based occurred in the past 30 years. Based on these findings, it was concluded that PPV variants are continuously evolving and that vaccines, which are based mainly on strains isolated about 30 years ago, should perhaps be updated.

INTRODUCTION

Porcine parvovirus (PPV) was first isolated in Germany and the USA in 1965 and later found worldwide (Cartwright & Huck, 1967; Johnson, 1973; Joo et al., 1976; Siegl, 1976). PPV is considered to be a major cause of reproductive failure in swine. Clinical signs of PPV infection are characterized by the re-occurrence of oestrus, abortion and the delivery of mummified and stillborn fetuses (Mengeling et al., 2000), commonly described with the acronym SMEDI (stillbirth, mummification, embryonic death and infertility).

PPV is a small ssDNA virus. It has a genome of about 5000 nt encoding four proteins transcribed from two promoters, and the coding capacity is extended by alternative splicing. Two non-structural proteins, NS1 and NS2, are transcribed and translated from the 5’ ORF and are important for DNA replication. Additionally, two structural proteins are transcribed and translated from the 3’ ORF (VP1 and VP2). The smaller protein, VP2, is produced by splicing from the same RNA template as the larger protein (VP1). The whole VP2 sequence is therefore present within the VP1 sequence, which has a unique N terminus of ~120 aa. A third protein, VP3, is a post-translational modification product of VP2 (Simpson et al., 2002). Together, these three viral proteins assemble to form the icosahedral capsid (Chapman & Rossmann, 1993; Parrish, 2010). Various biotypes of PPV with different pathogenic properties are known. The genetic basis of the pathogenicity has not yet been defined, but the structural protein appears to play a major role. A genomic comparison between the non-pathogenic strain NADL-2 (GenBank accession no. L23427) and the virulent strain...
Kresse (U44978) revealed that the non-coding regions of the genome were nearly identical (Bergeron et al., 1996). All nucleotide substitutions found in the non-structural genes were silent, whilst six out of eight substitutions located in the structural genes (VP1/VP2) changed the transcribed amino acid. In detail, among the VP2 amino acids, five changes were consistent in comparison with field isolates (I215T, D378G, H383Q, S436P and R565K), and three of these changes (D378G, H383Q and S436P) were considered to be responsible for the different tissue tropisms (Bergeron et al., 1996). Recently, genetic variation and the possible emergence of a new antigenic type of PPV has been described (Zimmermann et al., 2006), although its importance in the field is still unclear.

As PPV replicates by using the host DNA replication machinery, it has generally been assumed that the virus has a low rate of nucleotide substitution, close to that found in its mammalian host. However, in recent years, it has been shown that canine parovirus (CPV) and human B19 erythrovirus, both autonomous paroviruses, have a nucleotide substitution rate of approximately $1 \times 10^{-4}$ substitutions per site per year (Shackelton & Holmes, 2006; Shackelton et al., 2005). This rate is similar to that known for RNA viruses.

The continuous use of an inactivated vaccine in swine herds over the past three decades, and the remaining occurrences and reports of reproductive failures caused by PPV, highlight the importance of continuous monitoring of PPV isolates. To address these questions, recent PPV isolates were analysed by determining nucleotide substitutions, phylogenetic analysis and an estimation of the molecular evolutionary rate of PPV.

## RESULTS

### Genetic analysis of the newly identified sequences

In order to identify new mutations, recent isolates from Austria, Brazil, Germany and Switzerland were analysed. Sequence analysis of the structural VP2 gene of these isolates ($n=9$) revealed nucleotide substitutions at 32 sites. Seventeen substitutions were synonymous and 15 were non-synonymous. In the region previously considered to be highly variable (located between nt 3889 and 4239 of the full PPV genome; Zimmermann et al., 2006), three synonymous substitutions and five non-synonymous substitutions were found. Two conserved regions (with no nucleotide substitutions) could be defined within nt 3507–3718 and 4251–4425.

Analysis of the non-structural NS1 gene of the new isolates ($n=9$) revealed nucleotide substitutions in 44 sites. Twenty-seven substitutions were synonymous and 17 were non-synonymous. Conserved regions (with no nucleotide substitutions) were identified within nt 568–781, 847–1022, 1072–1362 and 1504–1793.

At the protein level, conserved regions are located in the structural gene at aa 83–225 and 437–579. Two variable regions were found at aa 226–233 and 365–436. For the non-structural gene, conserved regions (with no amino acid substitutions) were located at aa 83–163, 165–357 and 368–561. A region with higher variability was observed at the end of the gene, between aa 562 and 658.

The amino acid substitutions in the VP2 structural gene for each isolate are shown in Table 1. In the sites that differed between the NADL-2 and Kresse strains, the recent isolates appeared to be nearly identical to the Kresse sequence. Unique amino acid substitutions could be identified at aa 436 and 565. In all of the new German and Austrian strains, substitutions were located at aa 228, 414 and 419. At aa 436, all German strains contained the amino acid threonine, which was also observed in the Brazilian strain S30. A higher number of amino acid substitutions in comparison with the other isolates was found in strain 15425, with seven unique amino acid changes.

### Substitution rates and selection pressures

Only the maximum $\chi^2$ method (of the five methods tested) estimated a recombination event in the samples 7a (this study), 143a (GenBank accession no. AY684867) and 3226005_1d (GQ884037) in the NS1 dataset. As no consensus was obtained with three or more methods, all samples were used for further analysis.

The root-to-tip analysis was constructed to examine whether the samples exhibited adequate temporal structure for a substitution rate analysis. The first plot of the genetic distance versus the isolation year of the sample displayed an outlier sequence (vaccine virus IDT; GenBank accession no. AY684872). This sequence was excluded from further analysis (according to the methodology adopted by Shackelton & Holmes, 2006) and a new maximum-likelihood (ML) tree and root-to-tip plot were calculated (Fig. 1). The values of $R^2$ obtained by linear regression for the VP1 complete gene, VP1 partial gene and NS1 complete gene were 0.3828, 0.4438 and 0.2984, respectively.

The mean evolutionary rate estimated for the three datasets using the BEAST approach ranged from $10^{-5}$ to $10^{-4}$ substitutions per site per year. According to the model best fitting the data, the dataset of the NS1 gene, VP1 complete gene and VP1 partial gene presented mean rates of $5.39 \times 10^{-5}$, $3.02 \times 10^{-4}$ and $4.04 \times 10^{-4}$ substitutions per site per year, respectively (Table 2). According to the methodology, the mean variation range was $4.00 \times 10^{-2}$–$7.73 \times 10^{-3}$ substitutions per site per year for the non-structural dataset, $1.74 \times 10^{4}$–$3.02 \times 10^{4}$ substitutions per site per year for the VP1 complete dataset and $3.93 \times 10^{4}$–$4.45 \times 10^{4}$ substitutions per site per year for the VP1 partial dataset.

The ratio of non-synonymous to synonymous substitution rates ($d_{NS}/d_S$) for the NS1 and VP1 datasets presented variable results according to the methodology used. A $\chi^2$
test comparison between models M7 and M8 revealed that both complete genes were under purifying selection (NS1, \( P = 0.3146 \); VP1, \( P = 0.0407 \)). Furthermore, no positive selection sites were found in the M8 model. In contrast, evidence of a positive selection for the VP1 partial dataset was found with the M7 and M8 models (\( P = 0.0006 \)) and in the MEGA analysis (\( d_N/d_S = 1.610 \)). For VP1 (complete and partial), no evidence could be identified in the branch-site model test (\( P = 0.0786 \) and 0.9972, respectively) for the case that different selective pressures would drive evolution of the distinct lineages/branches. In contrast, branches characterized by different selective pressures were present within the NS1 dataset (\( P = 0.0001 \)).

**Phylogenetic analysis**

As shown in Fig. 2a, the Bayesian maximum clade credibility tree generated for the VP1 complete gene revealed two different clades (here called A and B; posterior probability <0.90) and two clusters (C and D; posterior probability =1.0). The new German and Austrian strains were located in cluster C and the Swiss strain was located in cluster D.

In the tree based on the partial VP1 gene sequence (Fig. 2b), clades and clusters A–D were also present, but additionally two other clades (E and G; posterior probability <0.90) and one other cluster (F; posterior probability=1.0) could be identified. The new German and Austrian strains remained in the same cluster, whilst the new Brazilian and the Swiss strains were located in clade B and cluster D, respectively. In contrast, except for strain ZJ (GenBank accession no. EU790642) [with no clade (VP1 partial)\( \rightarrow \)clade B (VP1 complete)] and strain 225b (AY684864) [cluster C (VP1 partial)\( \rightarrow \)cluster D (VP1 complete)], all of the sequences were affiliated with the same clade/cluster in the VP1 complete and VP1 partial trees. In the NS1 gene tree (Fig. 2c), almost all sequences formed distinct clusters, and the only clusters in agreement with the VP1 trees were D and F.

Chronological analysis of the NS1 and VP1 datasets demonstrated two distinct temporal periods. In the structural VP1 gene, the main divergences occurred in the last 30 years for both datasets, and the events resulting in the outcome of new strains dated back to the last 20 years. In the non-structural NS1 dataset, the divergences were estimated to be formed in an earlier time period of 10–125 years ago, being mainly concentrated in the last 30–60 years.

**DISCUSSION**

The possible antigenic selection of PPV mutations as a consequence of the current vaccination schedules has been discussed in recent years (Soares et al., 2003; Zeeuw et al., 2007; Zimmermann et al., 2006). In the present paper, new virus sequences from Austria, Brazil, Germany and Switzerland, together with available PPV sequences from

### Table 1. Amino acid substitutions within PPV VP2 sequences

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Kresse</th>
<th>NADL-2</th>
<th>27a</th>
<th>15425</th>
<th>7a</th>
<th>8a</th>
<th>14a</th>
<th>693a</th>
<th>Embrapa</th>
<th>PA</th>
<th>PB</th>
<th>S30</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>A</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>45</td>
<td>S</td>
<td>T</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>82</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>K</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>215</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>226</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>E</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>228</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>233</td>
<td>S</td>
<td>S</td>
<td>T</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>304</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>P</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>320</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>T</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>366</td>
<td>D</td>
<td>D</td>
<td>N</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>378*</td>
<td>G</td>
<td>D</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>383*</td>
<td>Q</td>
<td>H</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>407</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
</tr>
<tr>
<td>414</td>
<td>A</td>
<td>A</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>419</td>
<td>E</td>
<td>E</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>Q</td>
<td>E</td>
<td>E</td>
<td>E</td>
<td>E</td>
</tr>
<tr>
<td>423</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>K</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>436*</td>
<td>P</td>
<td>S</td>
<td>T</td>
<td>P</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>565</td>
<td>K</td>
<td>R</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>K</td>
<td>R</td>
<td>R</td>
<td>R</td>
</tr>
</tbody>
</table>

*Sites considered responsible for the different tissue tropisms (Bergeron et al., 1996).
GenBank, were analysed and the evolutionary rate of this virus was calculated.

The analysis was performed using three different datasets. The VP1 partial dataset represented a small region of the VP1 gene; however, it enabled us to analyse a higher number of sequences over a broad range of years. The phylogenetic trees generated for the VP1 complete and partial sequences contained four (A–D) and seven (A–G) clades/clusters, respectively, and the majority of sequences located in clades/clusters A–D matched in both datasets (Fig. 2). This suggests that analyses of this short region of the VP1 gene (e.g. Soares et al., 2003) can be used to represent the full-length gene.

In a previous study that analysed the complete VP1 gene, the authors showed that the recent German strains 15a (GenBank accession no. AY684865), 21a (AY684868) and 27a (AY684871) formed a new phylogenetic cluster (Zimmermann et al., 2006). The new German and Austrian strains were located in this cluster (cluster C) present in both VP trees (complete and partial), and clustered together with European sequences isolated in 2005–2006. The Swiss strain was located in cluster D, together with German and other European strains. Both of the clusters containing mainly German sequences and a cluster with Chinese sequences were characterized by a high phylogenetic confidence, indicating that these clusters could have emerged from a well-established ancestor. The Brazilian strain S30 was located in clade B, together with two USA strains and one European strain. One other cluster (F) and two other clades (E and G) were seen in the partial dataset phylogeny. Cluster F was formed by Chinese strains, and clades E and G contained only Brazilian strains. The two clades formed mainly by Brazilian strains have been described in a previous study, where maximum-parsimony analysis was carried out on a similar dataset (Soares et al., 2003).

A comparison of both VP1 datasets with the NS1 dataset on the phylogenetic level was difficult. Only two of the VP clusters could be observed (D and F) in the tree generated for the NS1 gene. According to our temporal analyses, two distinct evolutionary rates between these genes were found. The phylogenetic incongruencies between the two genes could reflect the different evolutionary development of these genes.

At the protein level, amino acids differing in the VP2 protein, when compared with the NADL-2 and Kresse strains, were distributed equally between loops and β-strands. Here, we observed that the substitutions in the new strains were located mainly in the protein loops (Fig. 3). The only exception was the substitution of R→K at aa 82 in the Swiss strain 15425. Several substitutions in these loops have also been observed in CPV and are reported to be involved in cell specificity and haemagglutination activity (Chapman & Rossmann, 1993). The substitutions Q226E, Q228E, I320T, E419Q and N423K were located near the three-fold shoulder of the capsid subunit. This location has been considered to be a common antigenic surface region in distinct parvoviruses (Chapman & Rossmann, 1993). These sites were also further displaced from the centre of the PPV capsid in comparison with other capsid protein amino acids (Simpson et al., 2002), suggesting that these sites may interact with the host immune system.

The Swiss strain 15425 displayed several unique amino acid substitutions and appeared to be distant from the other samples. These unique substitutions were also found in PPV isolate 21620005_1h (GenBank accession no. 21620005_1h) (Fig. 1).
Table 2. Sequence information, clock model, demographic model, marginal likelihood, mean nucleotide substitution rate (in substitutions per site per year) and the corresponding 95% higher posterior density (HPD) lower and upper values for the three datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Sequence length (nt)</th>
<th>No. sequences</th>
<th>Clock model</th>
<th>Demographic model</th>
<th>Marginal likelihood</th>
<th>Mean rate</th>
<th>HPD</th>
</tr>
</thead>
<tbody>
<tr>
<td>VP1 complete</td>
<td>2187</td>
<td>31</td>
<td>UCED</td>
<td>Constant</td>
<td>−4022.3</td>
<td>3.02 × 10⁻⁴</td>
<td>4.86 × 10⁻⁴, 2.56 × 10⁻⁴</td>
</tr>
<tr>
<td>VP1 partial</td>
<td>739</td>
<td>65</td>
<td>UCED</td>
<td>BS</td>
<td>−1762.7</td>
<td>4.04 × 10⁻⁴</td>
<td>5.72 × 10⁻⁴, 1.37 × 10⁻⁴</td>
</tr>
<tr>
<td>NS1 complete</td>
<td>1989</td>
<td>31</td>
<td>UCED</td>
<td>BS</td>
<td>−3867.2</td>
<td>5.39 × 10⁻⁵</td>
<td>1.97 × 10⁻⁷, 1.37 × 10⁻⁴</td>
</tr>
</tbody>
</table>

UCED, Uncorrelated exponential distribution of rates; BS, Bayesian skyline.

GQ884035) from Europe. The relationship between these two samples could be observed by their proximity in the phylogenetic trees (Fig. 2).

In all new German and Austrian strains and in the Brazilian strain S30, the amino acid threonine was found at aa 436, an amino acid site located right in the three-fold spike centre of the capsid subunit. The higher incidence of threonine in the recent samples suggested that this amino acid is providing some adaptive advantage to the virus. According to the M8 model in the VP1 complete dataset, a higher selective pressure at this site in comparison with the whole virus genome was detected (0.998). The presence of this amino acid at this site has also been related to a possible decrease in the affinity of neutralizing antibodies (Zeeuw et al., 2007). To contribute to this discussion, future studies introducing mutations in these sites using the recent virulent strains need to be performed to determine whether these mutations are influencing/mediating the pathogenicity.

In contrast, a high identity between the β-strands within the VP sequences was observed. This was expected, as these regions have important functions in maintaining capsid integrity and in capsid–DNA interactions (Chapman & Rossmann, 1993).

In this study, a high evolutionary rate was found for PPV capsid genes (~3 × 10⁻⁴–4 × 10⁻⁴ substitutions per site per year) and only a moderate evolutionary rate was found for PPV non-structural genes (~5 × 10⁻⁵ substitutions per site per year). The HPD data reflect the uncertainty in the analysis. In the capsid genes, the lower and higher HPD were quite similar to the mean rate, providing support to the data. The root-to-tip analysis also supported the temporal structure for both VP1 datasets. In contrast, the NS1 dataset revealed that a lower R² and temporal analysis could only be suggested (Firth et al., 2010). The HPD for this dataset revealed a larger substitution rate range (10⁻⁴–10⁻⁷ substitutions per site per year), indicating that this parameter could not be estimated with confidence. A similar HPD range for the substitution rate could also be seen in the NS1 gene for feline panleukopenia parvovirus (FPLV) (Fig. 4), highlighting an uncertainty in the temporal analysis for the NS1 gene of the parvovirus (Shackelton et al., 2005).

It is expected that selective pressure and evolutionary rate should exhibit similar levels. In previous parvovirus studies, a strong correlation between selective pressure and the substitution rate for distinct viruses or genes was observed (Shackelton & Holmes, 2006; Shackelton et al., 2005). In the present study, the mean rate of dN/ds substitutions per site suggests that the NS1 and VP1 genes were under purifying selection, whilst the partial VP dataset was under positive pressure (PAML and MEGA analysis). According to all the models tested, a pressure order of VP1 partial>VP1 complete>NS1 complete was verified. The same relationship was demonstrated in the PPV evolutionary rate.

Recombination is one of the main diversity forces and allows genomic regions to have independent histories. Recombination events complicate the use of phylogenetic parameters such as timing events and selection pressures (Awadalla, 2003). The ubiquitous character of PPV in association with an immune pressure could result in recombination events, as has been suggested by Shackelton et al. (2007). In our dataset alignments, no consensus for recombination events was found by applying several methods. However, incongruence in the partial structural protein dataset tree could be observed in comparison with the complete structural gene tree, e.g. for strains 225b (GenBank accession no. AY684864) and ZJ (EU790642). Strain 225b was previously reported to be a possible recombinant strain originating from two distinct phylogenetic clusters (Shackelton et al., 2007). In order to avoid any false results in our timing analysis, the substitution rate was recalculated without the incongruent sequences. No significant differences were observed.

Traditionally, PPV is considered to be a virus with a more conservative genome than other parvoviruses and viruses with ssDNA (Duffy et al., 2008; López-Bueno et al., 2006; Lukashov & Goudsmidt, 2001). In a previous study, surface amino acid substitutions were found in German strains...
Fig. 2. Phylogenetic trees based on Bayesian inference analysis of PPV for the three datasets for the VP1 complete gene (a), VP1 partial gene (b) and NS1 complete gene (c). The scale axis indicates the distance in years. Posterior probabilities (≥0.90) are indicated above the branches. The identification name of each sequence is followed by the location and year of isolation. AUS, Austria; BR, Brazil; CH, China; EU, Europe (no precise location available); GE, Germany; SW, Switzerland. For both VP trees (a, b), the clades/clusters are delimited and indicated with the letters A–G; a solid line indicates a cluster and a dashed line a clade.
isolated from 2001 to 2002 (Zimmermann et al., 2006). Furthermore, it was demonstrated that antibodies raised against the vaccine strains (strains NADL-2 and IDT) have a lower neutralization activity against field strains (strains 27a and 143a) with these substitutions (Józwik et al., 2009; Zeeuw et al., 2007). In the present study, the modifications found in these two strains could also be observed for the current PPV isolates in the population, indicating that the current vaccines may no longer be fully protective. Additionally, new surface amino acid substitutions were also observed. These structural modifications found in strains collected in recent years may represent an escape mechanism from the inactivated PPV vaccine, often used in the past 30 years. Rapid evolution is already known among ssDNA viruses, including CPV and human B19 erythrovirus (Fig. 4) (Shackelton & Holmes, 2006; Shackelton et al., 2005). The reasons for the observed high substitution rates in parvoviruses and other ssDNA viruses remain unclear. Unlike RNA viruses, which use their own error-prone polymerases, PPV replicates using the cellular DNA polymerase of the swine host, implying that they should have the same replication fidelity as the host. Previous studies suggest that the proofreading or repair mechanism may not be efficient or accurate enough in these genomes or in cells with an active viral infection (Shackelton & Holmes, 2006; Shackelton et al., 2005). Alternatively, processes such as ssDNA deamination due to a lack of the double helix or non-functional host exonucleases (resulting from the absence of proper methylation patterns) could also lead to mutational vulnerability (Duffy & Holmes, 2008; Sanz et al., 1999).

In agreement with our report, previous studies could also define new virus profiles with new amino acid substitutions (Shangjin et al., 2009; Soares et al., 2003; Zimmermann et al., 2006). These substitutions were located mainly at the capsid surface, and the development of a surface profile distinct from the vaccine strains (mainly based on the NADL-2 strain) could be clearly noted. In addition, the phylogenetic tree associated with the molecular clock analysis revealed that the divergences between the main isolates were introduced for the non-structural gene dataset in the last 100 years and for both structural gene datasets in the last 10–30 years. Therefore, our data indicate that the currently used immunization schedules with inactivated whole-virus vaccines may need to be discussed. In the closely related canine and feline parvoviruses, inactivated whole-virus vaccines are no longer common and have been fully replaced by modified live-virus vaccines. Modified live-virus vaccines in these animals induce a long-lasting immune response that provides protection for at least several years and could also be an attractive alternative for PPV vaccination.

**METHODS**

**PPV samples.** In total, 11 recent isolates from Austria, Brazil, Germany and Switzerland were analysed. The Austrian (693a), German (7a, 8a and 14a), Swiss (15411, 15421 and 15425) and Brazilian (PA and PB) strains originated from clinical cases involving reproductive losses. The Brazilian strain S30 was obtained from 55-day-old (mean age) wasting piglets, and the strain Embrapa was isolated from an unknown source in Brazil.

**Amplification.** Total DNA was purified from the clinical samples and cell-culture supernatant using a QIAamp DNA Mini kit (Qiagen), in accordance with the manufacturer’s instructions. The DNA was stored at −22 °C until further analysis. NS1 and VP1/VP2 gene amplification was performed using the primer pairs listed in Table 1.
Supplementary Table S1 (available in JGV Online). PCR was performed using a ReddyMix PCR Master Mix (Abgene) in a final volume of 50 µl. Amplification by PCR was performed under the following conditions: one cycle at 94 °C for 2 min and 35 cycles of 94 °C for 30 s, 55 °C for 40 s and 72 °C for 1 min, with a final elongation at 72 °C for 5 min. PCR products were purified using a NucleoSpin purification kit (Macherey-Nagel), sequenced using a BigDye Terminator version 3.1 Cycle Sequencing kit (Applied Biosystems) and analysed on an ABI3100 Genetic Analyzer (Applied Biosystems) by Zentraler Funktionsbereich DNA-Sequenzierung (University of Leipzig, Germany). The DNA from the Brazilian strains was extracted using a silica-based protocol (Boom et al., 1990). Nested PCRs were performed as described previously (Soares et al., 2003) and custom sequenced as described above.

Assembly of the obtained sequences to a full-length sequence was performed using the SeqMan program of the Lasergene software (DNASTAR).

All nucleotide numbers used in the present study refer to the Kresse strain (GenBank accession no. U44978). Amino acid numbers are according to the VP2 protein of the same strain.

**PPV dataset.** All NS1 and VP gene sequences deposited in GenBank (up to September 2010), containing the isolation year, were retrieved from the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) using 'porcine parvovirus' as the key words. These sequences, together with the new sequences described here, totalled 31 complete sequences of the NS1 gene and 31 complete sequences of the VP2 gene, which were analysed separately. In order to increase the number of sequences in the analysis, a third dataset containing 65 partial sequences of the VP2 gene was also constructed. This last dataset consisted of sequences of 739 nt (nt 3701–4439). The year of isolation of the Brazilian samples PA, PB and Embrapa were unknown; therefore, these samples were excluded from further analyses. The datasets were aligned by applying CLUSTAL W using MEGA4 software (Tamura et al., 2007). Possible genetic recombination events between virus strains were examined using the Bootscan, Chimaera, GENECONV, maximum d² and SiScan methods with default parameters, implemented in the RDP3 software package (Martin et al., 2010). Details of these methods are given in Martin et al. (2010) and the program manuals. Recombination events were assumed when a consensus between three or more methods was found.

**Phylogenetic analysis and selection pressures.** To perform ML phylogenies, the DNA substitution model was obtained using the ModelTest version 3.7 software (Posada & Crandall, 1998). The analysis was performed with the software PAUP version 4.0b (Swofford, 2002). A heuristic search was performed using the tree bisection and reconnection method of branch swapping. Support was obtained using 100 bootstrap repetitions. The clock-like behaviour of each dataset was visualized using regression of the root-to-tip genetic distance inferred from the ML trees against the sampling time in the Path-O-Gen version 1 software (Drummond et al., 2003).

The rate of nucleotide substitution per site per year was estimated with a Bayesian Markov-chain Monte Carlo (MCMC) method, using BEAST version 1.5.4 (Drummond & Rambaut, 2007). These analyses were run using the Hasegawa–Kishino–Yano DNA substitution model with partitions into codon positions, performing 100 million generations through the MCMC and subsampling each 10 000 generations. The population dynamic models tested were: (i) a constant population size and (ii) a Bayesian skyline coalescent model. The molecular clock models tested were: (i) a strict molecular clock, (ii) a relaxed clock with an uncorrelated log-normal distribution of rates and (iii) a relaxed clock with an uncorrelated exponential distribution of rates. Multiple runs were performed for each dataset to find the best model fitting the data (i.e. with a higher marginal likelihood score; Suchard et al., 2001). The resulting data were analysed visually using TRACER (http://tree.bio.ed.ac.uk/software/tracer/) after removing a 10 % 'burn in' for each data. A ‘consensus’ tree for each dataset was generated with the TreeAnnotator version 1.5.4 software (in the BEAST package). Phylogenetic trees were visualized with FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/). The resulting trees were sorted to facilitate comparison with the previously described VP1 and VP2 gene clusters. Groupings with posterior probabilities ≥0.90 were considered to be clusters and those with posterior probabilities <0.90 were considered to be clades.

The value of the dS/dê ratio for each branch of the phylogenetic tree was estimated for the NS1, VP1 complete and VP1 partial datasets using the PAML version 4.2b package (Yang, 1997) following the
method of Shackleton et al. (2005) with a few adjustments and using the ML tree as described above. A comparison of the M7 and M8 models was used to analyse selection pressures at specific codons in the CODEML option (P<0.001). The branch-site model was used to detect positive selection acting in particular lineages/branches and compared with a null ω² fixed model (P<0.001). In addition, the modified method of Nei & Gojobori (1986), applied with the MEGA4 software, was also used to find the dS/dN ratio changes.

**ACKNOWLEDGEMENTS**

This research was supported by the Deutscher Akademischer Austausch Dienst (DAAD, Germany), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) and Coordenacao de Aperfeicoamento Pessoal de Nivel Superior (CAPES, Brazil). We thank Professor Ritzmann, Vienna, Professor Sidler, Zürich, Dr Ohlinger, Münster, and Dr Ciacci-Zanella, Concórdia, for providing parvovirus-positive samples and Professor Pfeffer for critically reading the manuscript.

**REFERENCES**


