Isolation and cloning of two variant papillomaviruses from domestic pigs: *Sus scrofa* papillomaviruses type 1 variants a and b

Hans Stevens, Annabel Rector, Kees Van Der Kroght and Marc Van Ranst

Laboratory of Clinical and Epidemiological Virology, Rega Institute for Medical Research, University of Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium

The healthy skin of two female domestic pigs (*Sus scrofa domestica*) was sampled with cotton-tipped swabs. Total genomic DNA was extracted from the samples and subjected to PCR with degenerate papillomavirus (PV)-specific primers. Similarity searches performed with BLASTN showed that partial E1 and L1 sequences of two novel PVs were amplified. Subsequently, the complete genomes of these *Sus scrofa* papillomaviruses (SsPVs) were amplified by long-template PCR, cloned and sequenced using a transposon insertion method. They contained the typical PV open reading frames (ORFs) E1, E2, E4, E6, L1 and L2, but the E7 ORF was absent in both viruses. Pairwise nucleotide sequence alignment of the L1 ORFs of the SsPVs showed 98.5% similarity, classifying these viruses as SsPV type 1 ‘variants’ (SsPV-1a and -1b). Based on a concatenated alignment of the E1, E2, L1 and L2 ORFs of SsPV-1 variants a and b, and 81 other human and animal PV type species, a neighbour-joining phylogenetic tree was constructed. This phylogenetic analysis showed that the SsPV-1a and -1b variants did not cluster with the other PVs of artiodactyls (cloven-hoofed) host species, but clustered on the edge of the genus *Alphapapillomavirus*, very near to the root of this genus.

INTRODUCTION

The *Papillomaviridae* are a large family of small, non-enveloped, double-stranded DNA viruses that cause proliferations of the stratified squamous epithelium of the skin or the mucosa in a wide variety of higher vertebrates. A large number of different papillomavirus (PV) types have been genetically characterized and most of these genotypes are highly species specific (Sundberg, 1987; Sundberg *et al.*, 1997). In humans, the most extensively studied host, nearly 100 human PV (HPV) types have been described, and these cause a wide spectrum of genotype-specific lesions (Van Ranst *et al.*, 1992). The limited number of non-human PVs that have been fully characterized so far mostly includes PVs from domestic mammals, but also from a number of wild and exotic mammals and two avian species (de Villiers *et al.*, 2004; Sundberg, 1987; Sundberg *et al.*, 1997, 2001; Tachezy *et al.*, 2002). Taken together with the numerous partial sequences of putative novel non-human PVs that have been reported (Antonsson & Hansson, 2002; Chan *et al.*, 1997; Ogawa *et al.*, 2004), and a large number of case reports of suspected PV infections in non-human vertebrates (Sundberg, 1987), this suggests that every vertebrate species could carry its own set of species-specific PVs.

A major obstacle in the characterization of novel PVs is the absence of a conventional cell culture system for virus propagation. In the past, identification of PVs could therefore only be accomplished when the virus was present in large quantities. Recent methods for the identification of novel PVs have generally been PCR-based. They make use of consensus or degenerate primers, which are developed based on sequence information from previously characterized PV types. These primers are typically located in the L1 and/or E1 genes, the most conserved regions of the PV genome. By means of this approach, it is possible to amplify and characterize a part of an unknown virus, even when there are only very small quantities of this virus. Given the circular structure of the PV genome, the complete genomic sequence can then be amplified by inverse or overlapping long-template PCR (Rector *et al.*, 2005b; Terai & Burk, 2002).

Although the first report of pig papillomatosis dates back to 1961 (Parish, 1961), PV DNA was not detected in papillomatous lesions on the skin of pigs. We combined a non-invasive sampling technique and the use of degenerate primers to amplify novel PVs from the healthy skin of two pigs. Here, we report the complete genomic sequence of...
two variants of a novel PV type: *Sus scrofa* PV type 1 variants a and b (SsPV1a and -1b). These two viruses were found to belong to the genus *Alphapapillomavirus*, but clustered very close to the root of this genus.

**METHODS**

**Origin of samples.** Samples were collected with pre-wetted (0.9 % NaCl solution) cotton-tipped swabs that were drawn back and forth over the healthy skin of two female domestic pigs (pigs A and B) and then suspended in 1 ml 0.9 % NaCl solution. The two pigs were swabbed at four different anatomical locations: the top of the head, the nose, the left front leg and the back. After removal of the swabs, the samples were stored at ~80 °C until further analysis. A volume of 400 µl of the samples was used to extract DNA using a QIAamp DNA Blood Mini kit (Qiagen) and following the manufacturer’s protocol.

**Degenerate-primer PCR.** PCRs with degenerate PV-specific primers were performed on the isolated DNA, using the degenerate primer pairs AR-E1F2/AR-E1R3 (371 bp) and AR-L1F1/AR-L1R3 (600 bp) (Rector et al., 2004); the expected length of the amplification product, based on the sequence of HPV-1a, is given in parentheses. PCR was carried out in a total volume of 50 µl, containing 200 µM each dNTP, 0.75 µM forward and reverse primer, 1 U Taq DNA polymerase and 2.5 mM MgCl2 (pH 8.5), with 1 µl of the extracted DNA as template. PCR conditions comprised denaturation for 10 min at 94 °C, followed by 45 cycles of 1 min at 94 °C, 1 min at 50 °C and 1 min at 72 °C. Amplons suggestive of PV-specific amplification were generated with these primer pairs in one sample that was taken from the back of pig A. The PCR products were sequenced with the same primers as used for PCR. Similarity searches, performed with the National Center for Biotechnology Information (NCBI) BLAST server on GenBank DNA database release 155.0 (Altschul et al., 1990) showed that partial sequences of the L1 and E1 ORFs of a novel PV were amplified.

**SsPV-specific PCR.** After sequencing the SsPV-1a genome, pig PV-specific primer pairs SsPV-L1F (5’-CACGCCAGAATGTATCGACC-3’) and SsPV-L1R (5’-GTCCTGATGTCATGATCC-3’), and SsPV-E1F (5’-GTACAAAGAAGCTGATGC-3’) and SsPV-E1R (5’-GATACTGGGTTCACCTTTGGC-3’) were designed for the L1 and E1 region, respectively. A PCR with these primer sets was performed on the isolated DNA of the pig A and pig B samples under the following conditions: 5 min at 94 °C, followed by 45 cycles of 1 min at 94 °C, 1 min at 50 °C and 1 min at 72 °C. Amplons suggestive of SsPV-specific amplification were generated with these primer pairs in one sample that was taken from the back of pig A. The PCR products were sequenced with the same primers as used for PCR. Similarity searches, performed with the National Center for Biotechnology Information (NCBI) BLAST server on GenBank DNA database release 155.0 (Altschul et al., 1990) showed that partial sequences of the L1 and E1 ORFs of a novel PV were amplified.

**Long-template PCR.** Primers for long template PCR were chosen in the partial E1 and L1 sequences of the SsPV isolates, in order to amplify the complete genomes of the different pig SsPVs in two overlapping long PCR fragments. For SsPV-1a and -1b, respectively, the following fragments were amplified: SsPV1a_long1 of approximately 3.9 kb, amplified with forward and reverse primers SsPV1a_long1F (5’-AGAATTCTGGGTTTGAGTGGCC-3’) and SsPV1a_long1R (5’-CCCACTTCCCGCCGTTTGC-3’), and fragment SsPV1a_long2 of approximately 3.7 kb, amplified with primers SsPV1a_long2F (5’-CTGCAATTCCTGGGTTTCG-3’) and SsPV1a_long2R (5’-CTGCAATTCCTGGGTTTCG-3’); fragment SsPV1b_long1 of approximately 4.0 kb, amplified with primers SsPV1b_long1F (5’-ACTTCCAGGCTTCTCTGTAGGGCATC-3’) and SsPV1b_long1R (5’-CAGATACGGTGTTAGTTAGGCA-3’), and fragment SsPV1b_long2 of approximately 3.9 kb, using primers SsPV1b_long2F (5’-GATTCTACTCCTTGCAGCCTTGGC-3’) and SsPV1b_long2R (5’-GTCACAAATTTCTGCTGAGCCACGTCC-3’). The long-template PCR was performed with the Expand Long Template PCR System (Roche Diagnostics). PCR products were run on a 0.8 % agarose gel, and isolated and purified from the gel by using SNAP purification columns (TOPO XL PCR Cloning kit; Invitrogen).

**DNA transformation and cloning.** The long-template PCR fragments were ligated into vector pCRII-TOPO, followed by transformation into One Shot TOP10 competent cells (TOPO XL PCR Cloning kit; Invitrogen). The bacteria were grown selectively on Luria–Bertani agar plates containing 50 µg kanamycin ml⁻¹. For SsPV-1a, one clone containing the 3.9 kb SsPV1a_long1 PCR fragment and one containing the 3.7 kb SsPV1a_long2 PCR fragment were selected. For SsPV-1b, one clone containing the 4.0 kb SsPV1b_long1 PCR fragment and one containing the 3.9 kb SsPV1b_long2 PCR fragment were selected.

**DNA sequencing.** The EZ::TN<–TET-1> insertion kit (Epitope) was used to retrieve the complete viral sequences. This kit uses the Tn5 transposase to randomly insert primer-binding sites and a tetracycline resistance selection marker into target DNA in vitro. The reaction was performed according to the manufacturer’s protocol. The transposon insertion reaction product was used to transform One Shot MAX Efficiency DH5α-T1R competent cells (Invitrogen). For each fragment, 12 colonies were selected and the provided primers were used to sequence the insertion clones bidirectionally from primer-binding sites at the 5’ and 3’ ends of the inserted transposon. The remaining gaps in the sequences were determined by primer walking on the PCR clones. Sequencing was performed on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems). The chromatogram sequencing files were inspected with Chromas 2.3 (Technelysium). The complete nucleotide sequences of the SsPV-1a and -1b genomes were 7260 and 7252 bp, respectively.

**DNA and protein sequence analysis.** Open reading frame (ORF) analysis was performed with the ORF Finder tool on the NCBI server of the National Institutes of Health (http://www.ncbi.nlm.nih.gov/orf/orf.html). The molecular mass of the putative proteins was calculated using the ExPaSy Compute pI/Mw tool (http://www.expasy.org/tools/pi_tool.html). Pairwise sequence alignments of the different ORFs were calculated using the ClustalW program. Multiple nucleotide sequence alignments were constructed in DAMBE version 4.2.13 (Xia & Xie, 2001). The sequences of SsPV-1a and -1b and 81 other animal and human PV types were imported and aligned at the amino acid level using CLUSTAL-W (Thompson et al., 1994), after which the nucleotide sequences were aligned according to the aligned amino acid sequences. This was carried out separately for the different ORFs, and the unambiguously alignable parts of the E1, E2, L2 and L1 ORFs were pasted together in one concatenated alignment of 2667 nt. Concatenated alignments for the early ORFs E1 and E2 and the late ORFs L2 and L1 were also constructed, and were 1278 and 1389 nt, respectively. Nucleotide positions that were included in the alignments were nt 1384–1512, 1543–1581, 1600–1815, 1852–2238 and 2239–2402 in E1, nt 2588–2827 and 2849–2960 in E2, nt 3826–4014, 4618–4635 and 4744–4851 in L2, and nt 5243–5362, 5393–5602, 5636–5719, 5768–6019, 6092–6253, 6296–6424, 6458–6502 and 6544–6625 in L1, relative to the SsPV-1a sequence. Based on these alignments, phylogenetic trees were constructed using the neighbour-joining method in MEGA version 3.1 (Kumar et al., 2004). Trees were constructed using all three codon positions, as well as using only the
first and second codon position. Bootstrap support values were obtained for 10000 replicates.

RESULTS

Complete SsPV-1a and -1b genomic sequences

The complete nucleotide sequences of SsPV-1a and -1b were 7260 and 7252 bp, respectively, and both had a G+C content of 53.8 mol%. They contained seven classical PV major ORFs. The exact locations of the ORFs and the calculated molecular masses of the putative proteins are shown in Fig. 1.

SsPV-1a and -1b early genes

The putative SsPV-1a and -1b E6 proteins contained two conserved zinc-binding domains (C-X-X-C-X29-C-X-X-C), separated by 36 aa. An E7 ORF was not present in either SsPV-1a or -1b. The E1 ORFs encoded the largest protein, 597 aa in both SsPVs, and the ATP-binding site of the ATP-dependent helicase (GPPNTGKS) was conserved in the C-terminal part of both E1 proteins. In the E2 proteins, a leucine zipper domain was present, consisting of a periodic repetition of leucine residues at every seventh position over a distance covering eight helical turns (L-X6-L-X6-L-X6-L). The E4 ORFs of the two SsPVs were completely contained within the E2 genes and, as in most other PVs, did not contain a start codon. E4 proteins usually have a high proline content (mean of 15–20 %) and this was certainly the case for the SsPVs: 50 and 54 out of 164 aa for SsPV-1a and -1b, respectively. A putative E5 ORF was found in both SsPVs. The E5 protein is often a short protein (less than 100 aa) and has a high percentage of hydrophobic, aliphatic amino acids [isoleucine (I), leucine (L) and valine (V)]. The putative E5 protein of SsPV-1a was 47 aa with 42.6 % I+L+V, whilst the SsPV-1b putative E5 protein contained 42 aa with 42.9 % I+L+V.

SsPV-1a and -1b late genes

The late regions of the genomes contained the major (L1) and minor (L2) capsid protein genes. In both viruses, the L1 and L2 proteins contained a series of arginine and lysine residues at their 3′ end, which can function as a nuclear localization signal.

Non-coding region (NCR)

The classic NCR between the stop codon of L1 and the first ATG of E6 was 542 bp in both SsPVs (nt 6730–12 in SsPV-1a and nt 6722–12 in SsPV-1b). PVs usually contain an E1-binding site (E1BS) flanked by two E2-binding sites (E2BS), for binding of an E1/E2 complex in order to activate the origin of replication. In SsPV-1a and -1b, an E1BS (TGATTGTTAGTAGCAAT) was present at nt 7166–7182 and 7158–7174, respectively. The E1BS was flanked by two E2BS with the consensus sequence ACCN₆GGT located equidistant to the E1BS at nt 7127 and 7208 for SsPV-1a, and at nt 7119 and 7200 for SsPV-1b. In each virus, there was a third E2BS at nt 6982 for SsPV-1a and at nt 6974 for SsPV-1b. The NCR of SsPV-1a also contained two putative nuclear factor 1 (NF-1)-binding sites (TGGCC) at nt 7041 and 7083, a putative SP1 transcription factor-binding site (GCGGG) at nt 6962 and a putative AP-1 transcription factor-binding site (TGANTCA) at nt 6884. The SsPV-1b NCR contained the same putative binding sites at nt 7033 and 7075 for NF-1, at nt 6954 for SP1 and at nt 6876 for AP-1. At its 5′ end, the NCR of both

---

**Fig. 1.** Linear representation of the ORFs of the SsPV-1a and -1b genomes (with the molecular mass of the predicted proteins in kDa shown in parentheses). Numbers show the nucleotide positions from the start of the ORF (first nucleotide after the last stop codon in the same reading frame) to the stop codon.
SsPV-1a and -1b also contained a polyadenylation site (AATAAA at nt 6823 and 6815, respectively), upstream of a CA dinucleotide (nt 6857 and 6849, respectively), and a G/T cluster, necessary for the processing of the L1 and L2 capsid mRNA transcript. The TATAA box of the E6 promoter was present in both viruses at the 3' end of the NCR (nt 7074 for SsPV-1a and nt 7066 for SsPV-1b).

**Phylogenetic analysis**

To make an optimal sequence alignment of 81 PV types (type species of the PV genera and species), separate nucleotide sequence alignments were constructed for the different ORFs, based on the corresponding amino acid alignments. Only the PV core ORFs E1, E2, L1 and L2 were included, as only these ORFs are present in all characterized PVs. Unambiguously aligned regions were compiled in one concatenated alignment of 2667 nt. The resulting neighbour-joining phylogenetic tree using only the first and second codon position clustered the PVs in 18 different PV genera from alpha- to sigmapapillomaviruses that have been defined to date (de Villiers et al., 2004; Rector et al., 2004, 2005a) (Fig. 2). In this tree, SsPV-1a and -1b did not group with the other
artiodactyl PVs in the genus *Deltapapillomavirus*, but clustered on the edge of the genus *Alphapapillomavirus*, very near to the root of the genus, with a bootstrap support value of 96%. In the neighbour-joining phylogenetic tree using all three codon positions, the clustering of the PVs in the different PV genera remained unchanged. In this tree, the SsPVs also clustered on the edge of the genus *Alphapapillomavirus*, close to the root of the genus, but with a bootstrap support value of only 57% (data not shown). Concatenated alignments were also constructed for the early ORFs E1/E2 and the late ORFs L2/L1. The clustering of the SsPVs did not differ in the L2/L1 tree, but in the E1/E2 tree, the SsPV variants did not cluster within the genus *Alphapapillomavirus* (data not shown).

**Sequence similarity to other PVs**

Table 1 shows the sequence similarity between SsPV-1a, SsPV-1b and deer PV (DPV), ovine PV type 1 (OvPV-1) and bovine PV type 1 (BPV-1), as representatives of the major artiodactyl PVs in the genus *Deltapapillomavirus*, and HPV-16 and HPV-94, as representatives of the genus *Alphapapillomavirus*. The L1 ORFs of SsPV-1a and HPV-94 showed 61% nucleotide similarity. Therefore, according to the current definition of a novel papillomavirus genus, SsPV-1a and -1b cannot be placed in a novel genus, but have to be placed within the genus *Alphapapillomavirus* (de Villiers et al., 2004). A pairwise nucleotide alignment of the L1 ORFs of SsPV-1a and -1b showed that they had a similarity of 98.6%, which makes these two viruses variants.

**Table 1.** Percentage nucleotide and amino acid similarities of the different SsPV-1a ORFs with the ORFs of SsPV-1b, DPV (NC_001523), OvPV-1 (NC_001789), HPV-16 (NC_001526), HPV-94 (NC_005352) and BPV-1 (NC_001522)

<table>
<thead>
<tr>
<th>SsPV-1a</th>
<th>SsPV-1b</th>
<th>DPV</th>
<th>OvPV-1</th>
<th>HPV-16</th>
<th>HPV-94</th>
<th>BPV-1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6</td>
<td>97</td>
<td>97</td>
<td>29</td>
<td>14</td>
<td>NA</td>
<td>16</td>
</tr>
<tr>
<td>E1</td>
<td>97</td>
<td>97</td>
<td>44</td>
<td>38</td>
<td>47</td>
<td>39</td>
</tr>
<tr>
<td>E2</td>
<td>97</td>
<td>97</td>
<td>40</td>
<td>30</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>E4</td>
<td>97</td>
<td>97</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>L2</td>
<td>98</td>
<td>98</td>
<td>36</td>
<td>27</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>L1</td>
<td>98</td>
<td>99</td>
<td>48</td>
<td>44</td>
<td>50</td>
<td>47</td>
</tr>
</tbody>
</table>

NA, Insufficient similarity between the two sequences to allow an unambiguous alignment.
according to the current classification criteria (de Villiers et al., 2004).

**DISCUSSION**

All PVs have a conserved genomic organization, which provides strong evidence for a monophyletic origin. The general consensus on PV evolution is that different ancient PV lineages have co-evolved and co-speciated with their vertebrate host species during vertebrate evolution. PVs have a species-specific nature, with interspecies transmission being a very rare event that only occurs between closely related species, such as infection of horses with BPV-1 (Ammann et al., 1980). In order for this hypothesis of co-phylogenetic descent of PVs and their host species to hold, PVs and their hosts have to comply with Fahrenholz’s rule (Hafner & Nadler, 1988). This rule states that PVs of closely related host species should be closely related themselves and cluster together in the PV phylogenetic tree, with dating of PV divergence largely coinciding with host species divergence. This has previously been confirmed for *Psittacus erithacus tinmeh* papillomavirus type 1, isolated from an African grey parrot, and chaffinch *Fringilla coelebs* papillomavirus type 1 (Tachezy et al., 2002), for PVs from Artiodactyla (Moreno-Lopez et al., 1987) and for the pygmy chimpanzee PV type 1, the common chimpanzee PV type 1 and human HPV-13 (Van Ranst et al., 1995).

The SsPVs, however, were most closely related to members of the genus *Alphapapillomavirus*, which are mostly oral and genital HPVs, and did not cluster with any of the artiodactyl PVs that have previously been genetically fully characterized. It is possible that the SsPVs and members of the genus *Alphapapillomavirus* have a common ancestor that was already present in the common ancestor of both artiodactyls and primates. A similar situation is found for *Erethizon dorsatum* PV type 1, isolated from a North American porcupine. Phylogenetic analysis has revealed that this PV does not cluster with the previously characterized rodent PVs, hamster oral PV and *Mastomys natalensis* PV, but is most closely related to HPV-41, a cutaneous HPV (Rector et al., 2005a).

The large difference in bootstrap support between the phylogenetic tree using all three codon positions and the phylogenetic tree using only the first and second codon positions is due to potential saturation of the third codon position. This saturation can become a problem when phylogenies are made for distantly related sequences (Zardoya & Meyer, 2001), as is the case for the SsPVs.

As the complete DNA sequences of more animal PVs become available for analysis, it will be possible to create a more detailed picture of PV evolution. Therefore, it is absolutely necessary to sample and study many more animal PVs in order to test the PV co-evolution hypothesis further.

Although SsPV-1a and -1b are, according to the current classification criteria, members of the genus *Alphapapillomavirus*, there are some arguments to place them in a separate genus. Firstly, the SsPVs differ from the other members of the genus *Alphapapillomavirus* in genomic organization: all known alphapapillomaviruses have a classic E7 ORF, whilst this ORF is missing in SsPV-1a and -1b. It seems that an E7 protein is not essential for PVs, because the lack of E7 ORFs has been reported previously for the polar bear *Ursus maritimus* PV type 1 (Stevens et al., 2007), *Phocoena spinipinnis* PV type 1 isolated from a Burmeister’s porpoise and *Tursiops truncatus* PV type 2 from a bottlenose dolphin (Rehtanz et al., 2006). Secondly, in a phylogenetic tree based on alignment of the E1 and E2 ORFs, SsPV-1a and -1b did not cluster within the genus *Alphapapillomavirus* (data not shown). Generally, the position of different genera in a phylogenetic tree depends on the ORF used in the analysis. Different genera cluster in a different way for different ORFs, whilst the clustering of PV types within the same genus is stable, independent of the analysed ORF. The fact that the SsPVs had a different clustering pattern in the phylogenetic analysis of the early ORFs may be an indication to question the current classification criteria.

**REFERENCES**


Sus scrofa papillomavirus variants a and b