Isolation and cloning of the raccoon (Procyon lotor) papillomavirus type 1 by using degenerate papillomavirus-specific primers

Annabel Rector,1 Koenraad Van Doorslaer,1 Mads Bertelsen,2,3† Ian K. Barker,2 Rolf-Arne Olberg,2,3† Philippe Lemey,1 John P. Sundberg4 and Marc Van Ranst1

Correspondence Marc Van Ranst marc.vanranst@uz.kuleuven.ac.be

1Laboratory of Clinical and Epidemiological Virology, Rega Institute for Medical Research, University of Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium
2Department of Pathobiology, Ontario Veterinary College, University of Guelph, Ontario N1G 2W1, Canada
3Toronto Zoo, Ontario M1B 5K7, Canada
4The Jackson Laboratory, Bar Harbor, ME 04609-1500, USA

Received 3 January 2005
Accepted 18 March 2005

Partial sequences of a novel papillomavirus were amplified from a cutaneous lesion biopsy of a raccoon (Procyon lotor), by using PCR with degenerate papillomavirus-specific primers. The Procyon lotor papillomavirus type 1 (PlPV-1) DNA was amplified with long template PCR in two overlapping fragments, together encompassing the entire genome, and the complete PlPV-1 genomic sequence was determined. The PlPV-1 genome consists of 8170 bp, and contains the typical papillomaviral open reading frames, encoding five early proteins and two late capsid proteins. Besides the classical non-coding region (NCR1) between the end of L1 and the start of E6, PlPV-1 contains an additional non-coding region (NCR2) of 1065 bp between the early and late protein region, which has previously also been described for the canine oral papillomavirus (COPV) and the Felis domesticus papillomavirus (FdPV-1). Phylogenetic analysis places PlPV-1 together with COPV and FdPV-1 in a monophyletic branch which encompasses the Lambda papillomavirus genus.

The Papillomaviridae are a large family of species-specific viruses that cause proliferations of the stratified squamous epithelium of the skin or the mucosa in a wide variety of host species (Sundberg et al., 2001). Over 90 human papillomavirus (HPV) types have been completely sequenced (de Villiers et al., 2004), and these have a wide spectrum of genotype-specific lesions (Van Ranst et al., 1992). Only a limited number of non-human PV genotypes have been fully genetically characterized so far (Sundberg et al., 1997, 2001; de Villiers et al., 2004), but these cover a broad range of host species, and it is likely that most mammalian and avian species carry their own set of species-specific PV types.

Papillomatosis has been documented in a number of carnivores, mainly Canidae and Felidae (Sundberg, 1987; Sundberg et al., 2000). In 1994, the genomic sequence of the canine oral PV (COPV), associated with oropharyngeal papillomatosis in dogs, coyotes and wolves, was characterized (Delius et al., 1994). COPV is the largest of all known PV genomes, and contains a unique second non-coding region (NCR2) between the early and late protein region. Recently, a second carnivore PV genome, containing a similar NCR2, was isolated from a cutaneous lesion of a Persian cat (Felis domesticus PV type 1, FdPV-1) (Tachezy et al., 2002a). COPV and FdPV-1 share a high degree of sequence similarity, and belong to the genus Lambda PV.

Papillomavirus infection has been reported in raccoons (Procyon lotor), small carnivores that are widely distributed throughout North America. In one study, 2 of 53 wild trapped raccoons showed proliferative skin lesions that stained positive for PV group-specific antigens (Hamir et al., 1995). Since dogs and raccoons both belong to the sub-order Caniformia of the Carnivora, investigation of the relationship between the canine COPV and the raccoon PV would be interesting to test the hypothesis that PVs have co-evolved with their host species.

We report here the complete genomic sequence and

†Present address: Copenhagen Zoo, Copenhagen, Denmark.
‡Present address: Kristiansand Zoo, Kristiansand, Norway.

The GenBank/EMBL/DDBJ accession number of the sequence reported in this paper is AY763115.

Supplementary material is available in JGV Online.

The Papillomaviridae are a large family of species-specific viruses that cause proliferations of the stratified squamous epithelium of the skin or the mucosa in a wide variety of host species (Sundberg et al., 2001). Over 90 human papillomavirus (HPV) types have been completely sequenced (de Villiers et al., 2004), and these have a wide spectrum of genotype-specific lesions (Van Ranst et al., 1992). Only a limited number of non-human PV genotypes have been fully genetically characterized so far (Sundberg et al., 1997, 2001; de Villiers et al., 2004), but these cover a broad range of host species, and it is likely that most mammalian and avian species carry their own set of species-specific PV types.

Papillomatosis has been documented in a number of carnivores, mainly Canidae and Felidae (Sundberg, 1987; Sundberg et al., 2000). In 1994, the genomic sequence of the canine oral PV (COPV), associated with oropharyngeal papillomatosis in dogs, coyotes and wolves, was characterized (Delius et al., 1994). COPV is the largest of all known PV genomes, and contains a unique second non-coding region (NCR2) between the early and late protein region. Recently, a second carnivore PV genome, containing a similar NCR2, was isolated from a cutaneous lesion of a Persian cat (Felis domesticus PV type 1, FdPV-1) (Tachezy et al., 2002a). COPV and FdPV-1 share a high degree of sequence similarity, and belong to the genus Lambda PV.

Papillomavirus infection has been reported in raccoons (Procyon lotor), small carnivores that are widely distributed throughout North America. In one study, 2 of 53 wild trapped raccoons showed proliferative skin lesions that stained positive for PV group-specific antigens (Hamir et al., 1995). Since dogs and raccoons both belong to the sub-order Caniformia of the Carnivora, investigation of the relationship between the canine COPV and the raccoon PV would be interesting to test the hypothesis that PVs have co-evolved with their host species.

We report here the complete genomic sequence and
The bacteria were selectively grown on Luria–Broth agar competent cells (TOPO XL PCR Cloning Kit; Invitrogen). The fragments were ligated into a pCR-XL-TOPO purification columns (TOPO XL PCR Cloning kit; Invitrogen). Violet staining, and isolated from the gel by using SNAP purification columns. The fragments on both strands. Sequencing was performed on an ABI Prism 3100 Genetic Analyser (Perkin-Elmer Applied Biosystems), chromatogram sequencing files were inspected with Chromas 2.2 (Technelysium), and contigs were prepared using SeqMan II (DNASTAR). The complete nucleotide sequence of the PIPV-1 genome consists of 8170 bp, and was deposited in GenBank under the accession no. AY763115.

PIPV-1 has the fourth largest PV genome, after COPV (8607 bp; GenBank accession no. NC_001619), deer DPV (8374 bp; NC_001523) and FdPV-1 (8300 bp; AF480454), all of which contain two non-coding regions (Delius et al., 1994; Groff & Lancaster, 1985; Tachezy et al., 2002a). PIPV-1 contains the seven classical PV major ORFs, encoding five early (E) proteins E1, E2, E4, E6 and E7, and two late (L) capsid proteins L1 and L2. The exact location of the PIPV-1 ORFs and the molecular mass of the predicted proteins are indicated in Fig. 1. The position of the first nucleotide of the PIPV-1 genome was fixed corresponding to the start of the first major ORF in the early region.

Pairwise nucleotide and amino acid sequence alignments of the different ORFs and their proteins were performed with the GAP-program on the Sequence Analysis Server at Michigan Technological University (http://genome.cs.mtu.edu/align/align.html), to determine the degree of similarity of PIPV-1 to the feline FdPV-1, the canine COPV, the prototype benign cutaneous HPV-1a (GenBank accession no. NC_001356), the mucosal high-risk HPV-16 (NC_001526) and the bovine fibropapillomavirus BPV-1 (X02346) (Table 1). The PIPV-1 sequence is most similar to the other PVs from carnivores, COPV and FdPV-1. The L1 ORF is the most conserved region among PV types, and PVs that share between 60 and 70 % nucleotide identity in this region are defined as different species of the same genus. Since PIPV-1 shares 68 % nucleotide identity with COPV

![Fig. 1. Linear representation of the ORFs of the PIPV-1 genome (with the molecular mass in kDa of the predicted proteins shown in parentheses). Numbers show the nucleotide positions of the start and stop codons. URR, Upstream regulatory region or non-coding region (NCR1), NCR2, Second non-coding region.](image-url)
Table 1. Percentage nucleotide (amino acid) similarity of the different PlPV-1 ORFs with the ORFs of COPV (GenBank accession no. NC_001356), FdPV-1 (AF480454), HPV-1a (NC_001356), HPV-16 (NC_001526) and BPV-1 (X02346)

<table>
<thead>
<tr>
<th>PlPV-1 ORF</th>
<th>COPV</th>
<th>FdPV-1</th>
<th>HPV-1a</th>
<th>HPV-16</th>
<th>BPV-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>E6</td>
<td>45 (39)</td>
<td>53 (46)</td>
<td>44 (37)</td>
<td>27 (22)</td>
<td>26 (25)</td>
</tr>
<tr>
<td>E7</td>
<td>59 (53)</td>
<td>61 (49)</td>
<td>42 (35)</td>
<td>37 (32)</td>
<td>24 (20)</td>
</tr>
<tr>
<td>E1</td>
<td>65 (60)</td>
<td>65 (63)</td>
<td>58 (53)</td>
<td>49 (44)</td>
<td>46 (39)</td>
</tr>
<tr>
<td>E2</td>
<td>53 (45)</td>
<td>54 (51)</td>
<td>46 (40)</td>
<td>42 (34)</td>
<td>31 (24)</td>
</tr>
<tr>
<td>E4</td>
<td>36 (25)</td>
<td>33 (21)</td>
<td>37 (14)</td>
<td>NA (NA)</td>
<td>NA (NA)</td>
</tr>
<tr>
<td>L2</td>
<td>57 (54)</td>
<td>58 (57)</td>
<td>49 (44)</td>
<td>39 (32)</td>
<td>38 (24)</td>
</tr>
<tr>
<td>L1</td>
<td>68 (71)</td>
<td>63 (67)</td>
<td>58 (57)</td>
<td>52 (49)</td>
<td>55 (47)</td>
</tr>
</tbody>
</table>

NA, Insufficient similarity between the two sequences to allow unambiguous alignment.

and 63% with FdPV-1, it can be classified as the third species of the genus Lambda.

The putative PlPV-1 E6 protein contains four C-X-C motifs, involved in zinc-binding, whereas the E7 contains two such motifs. The PlPV-1 E7 also contains the conserved retinoblastoma tumour suppressor binding domain (DLYCDEHPMSDEEE). The E1 encodes the largest PlPV-1 protein (600 aa), and contains the conserved ATP-binding site for the ATP-dependent helicase (GPPTNGKS) in its carboxy-terminal part. In the E2, a slightly modified leucine zipper domain (F-X6-L-X6-L-X6-L) is present, in which the first leucine residue is replaced by a phenylalanine. Since these are both neutral and hydrophobic amino acids, this modification will probably only result in limited structural changes. The E4 ORF is completely contained within E2, and in PlPV-1 an E4 start codon was identified, which is not the case in most other PVs. A high proline content, which is typical for the E4 ORF, was noted (19 proline residues of 117 aa). The late region contains the major (L1) and minor (L2) capsid protein genes. Both L1 and L2 contain a series of arginine and lysine residues at their carboxy terminus, likely to function as a nuclear localization signal.

The classic non-coding region (NCR1) between the stop codon of L1 and the start codon of E6 consists of 494 bp in PlPV-1, from nt 7797 to 120. PVs usually contain an E1 recognition site (E1BS) flanked by two E2-binding sites, for binding of an E1/E2 complex in order to activate the origin of replication. In the PlPV-1 NCR1, an E1BS (TTATT-GTGTGTTAAACAAT) is present at nt 10–26. Three typical E2-binding sites (E2BSs) with the consensus sequence ACC-N6-GGT are present at nt 7972–7983, 8093–8104 and 55–66. Three additional modified E2-binding sites (E2BS*), one with the sequence AAC-N6-GGT at nt 73–84, and two with the sequence ACC-N6-GCT, at nt 8147–8158 and 8011–8022, could be identified through comparison of the NCR1 regions of PlPV-1, COPV and FdPV-1. Since the putative E2BS* at position 8147–8158 and the E2BS at 55–66 are equidistant to the E1BS, this E2BS* might be functionally important, although it is possible that the modifications could result in lower affinity binding. In its 5’ end, the NCR1 also contains a polyadenylation site (AATAAA, nt 7889–7894), 20 bp upstream of a CA dinucleotide, and the G/T cluster, necessary for the processing of the L1 and L2 capsid mRNA transcript. In the 3’ end, NCR1 contains a TATA box of the E6 promoter present at nt 43.

The PlPV-1 genome contains a second non-coding region (NCR2) of 1065 bp (nt 3666–4730) between the early and the late protein region. A similar NCR2 has previously only been characterized in COPV (Delius et al., 1994) and FdPV-1 (Tachezy et al., 2002a), both also belonging to the genus Lambda, and DPV, which belongs to the genus Delta (Groff & Lancaster, 1985). It is tempting to speculate that the presence of an NCR2 is a common feature of all Lambda PVs, and that the NCR2 of all three Lambda PVs could originate from a DNA sequence (of unknown function and origin) that was incorporated in their common ancestor. Since this is a non-protein-coding region, numerous point mutations, insertions and deletions have accumulated in this sequence during evolution, and a possible common origin is no longer recognizable above this background. A BLAST search with the PlPV-1 NCR2 failed to detect similarity with the NCR2 of FdPV-1 and COPV, or any other known sequences in GenBank. In contrast to the NCR1, the NCR2 contains no recognizable E1BS and E2BS. There is no classical polyadenylation site for processing of the early viral mRNA transcripts present in this NCR2, but several degenerate polyadenylation sites, followed at an appropriate distance of about 20 bp by a CA dinucleotide, were detected.

A neighbour-joining phyllogenetic tree was constructed, based on a concatenated E1/E2/L2/L1 nucleotide sequence alignment of PlPV-1 and 45 type species of the different PV genera and species. Multiple nucleotide sequence alignments were performed at the amino acid level by using the CLUSTALW program (Thompson et al., 1994) in the DAMBE software package version 4.2.7 (Xia & Xie, 2001), after which the nucleotide sequences were aligned according to the aligned amino acid sequences. This was done separately for the different ORFs, and the regions of the genome that contained unambiguously alignable homologous positions (four regions in E1, three in E2, four in L2 and seven in L1) were pasted together in one compiled alignment of 2658 nt. The resulting neighbour-joining phyllogenetic tree (Fig. 2), which was constructed in MEGA version 2.1 (Kumar et al., 2001), clusters the PVs in the different genera described in the new PV classification (de Villiers et al., 2004), and the additional Rho and Sigma genera (Rector et al., 2004; Rector et al., 2005). In this tree, PlPV-1 clusters with COPV and FdPV-1 in the genus Lambda.

With a mutation rate of 0.73–1.2 x 10^-8 nucleotide substitutions per base per year (Van Ranst et al., 1995; Tachezy et al., 2002a), PVs are very stable viruses that evolve through slow accumulation of point mutations, and recombination.
between different PV types has never been documented. The observed global distribution of a broad genetic diversity of PVs, together with the viral species specificity, the stability of their genomes, and the requirement for close contact for PV transmission, has led to the hypothesis that PVs are ancient viruses that have co-evolved and co-speciated with their host species during vertebrate evolution (Van Ranst et al., 1995). In order for this hypothesis of co-phylogenetic descent to hold, PVs of closely related host species should be closely related themselves, with dating of PV divergence largely coinciding with the host species divergence. This has previously been confirmed for the parrot PePV and the chaffinch FcPV (Tachezy et al., 2002b), for PVs from Artiodactyla (even-toed ungulates), and for the pygmy chimpanzee CCVP-1, and human HPV-13 (Van Ranst et al., 1995). The raccoon PlPV-1 is most closely related to COPV and FdPV-1, the only other PVs isolated from carnivores. The second requirement, that the PVs evolved and speciated in synchrony with their hosts (Fahrenholz’s rule), was examined by comparing the phylogeny of PlPV-1 with COPV and FdPV-1 with the phylogeny of their carnivore host species. Comparative genomics of the Carnivora indicate monophyly of the order, with an early split into two monophyletic suborders, the Feliformia and the Caniformia. Within the Caniformia, there exists a well-supported dichotomy between the canids and arctoids (including the mustelids, ursids and procyonids) (Flynn & Nedbal, 1998; Bininda-Emonds et al., 1999). Based on the co-evolution hypothesis, stating that the PV evolutionary tree should mirror the host species phylogeny, it was expected that PlPV-1 would cluster with COPV in a monophyletic branch. A maximum-likelihood phylogenetic analysis of the Lambda PV dataset, with HPV-1a as the outgroup, did not indicate this scenario as the best tree, but it could not be statistically rejected (data not shown). A more expanded dataset, including novel Lambda PV sequences, could prove to be useful to fully resolve the Lambda PV phylogeny. The recent understanding of parasite–host co-evolution, however, also argues that a less strict interpretation of Fahrenholz’s rule is often desirable, since most host and parasite phylogenies are only imperfect mirrors. These imperfections could be caused by host switching between lineages, but possible causes also include the interplay between co-speciation and other co-phylogenetic events, such as sorting (extinction, when parasites are entirely or apparently removed from host species), duplication (intrahost speciation), and inertia (lack of parasite speciation) (Paterson & Banks, 2001). Also in the case of carnivores and their PVs, these effects could be of influence.

Figure 2. Neighbour-joining phylogenetic tree, based on a concatenated E1/E2/L2/L1 nucleotide sequence alignment of PlPV-1 and 45 other PVs. The PV genera are indicated with their Greek symbols. The numbers at the internal nodes represent the number of bootstrap probabilities, as determined for 1000 iterations by the neighbour-joining method. Only bootstrap values greater than 85% are shown. The scale bar indicates the genetic distance (nucleotide substitutions per site).

Acknowledgements

We would like to thank the colleagues of the laboratory of Clinical & Epidemiological Virology, Department of Microbiology & Immunology, Rega Institute for Medical Research, University of Leuven, Belgium, for helpful comments and discussion. This work was.
supported by a fellowship of the University of Leuven (Leuven, Belgium) to A.R. and by the Flemish Fonds voor Wetenschappelijk Onderzoek (FWO-grant G.0288.01).

References


